Discover Applied Sciences

Research

Engineering synthetic and recombinant human lysosomal β-glucocerebrosidase for enzyme replacement therapy for Gaucher disease

Lílian L. Souza Figueiredo[1] · Wilson Lau Junior[1] · Victor Wendel da Silva Goncalves[1] · Ester Silveira Ramos[1] · Vania D'Almeida[2] · Lucas Eduardo Botelho de Souza[3] · Maristela Delgado Orellana[3] · Kuruvilla Joseph Abraham[4,5] · Flávio Lichtenstein[6] · Lucas Bleicher[7] · Vasco Azevedo[8,9] · Rigoberto Gadelha Chaves[10] · Giuliano Bonfá[11] · Velia Siciliano[11] · Ron Weiss[12] · Stanton Gerson[13] · Aparecida Maria Fontes[1]

Received: 23 June 2024 / Accepted: 23 September 2024
Published online: 04 October 2024

Abstract

Gaucher Disease (GD) is an autosomal recessive, lysosomal storage disease caused by pathogenic variants in the glucocerebrosidase gene, leading to the loss of β-glucocerebrosidase (GCase) enzymatic activity. Enzyme replacement therapy (ERT) with recombinant GCase is the standard of care in GD patients. Our study investigates the combined use of in silico molecular evolution, synthetic biology and gene therapy approaches to develop a new synthetic recombinant enzyme. We engineered four GCases containing missense mutations in the signal peptide (SP) from four selected mammalian species, and compared them with human GCase without missense mutations in the SP. We investigated transcriptional regulation with CMV and hEF1a promoters alongside a GFP control construct in 293-FT human cells. One hEF1a-driven mutant GCase shows a 5.2-fold higher level of transcription than control GCase. In addition, this mutant exhibits up to a sixfold higher activity compared with the mock-control, and the predicted tertiary structure of this mutant GCase aligns with human GCase. We also evaluated conserved and coevolved residues mapped to functionally important positions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42452-024-06227-z.

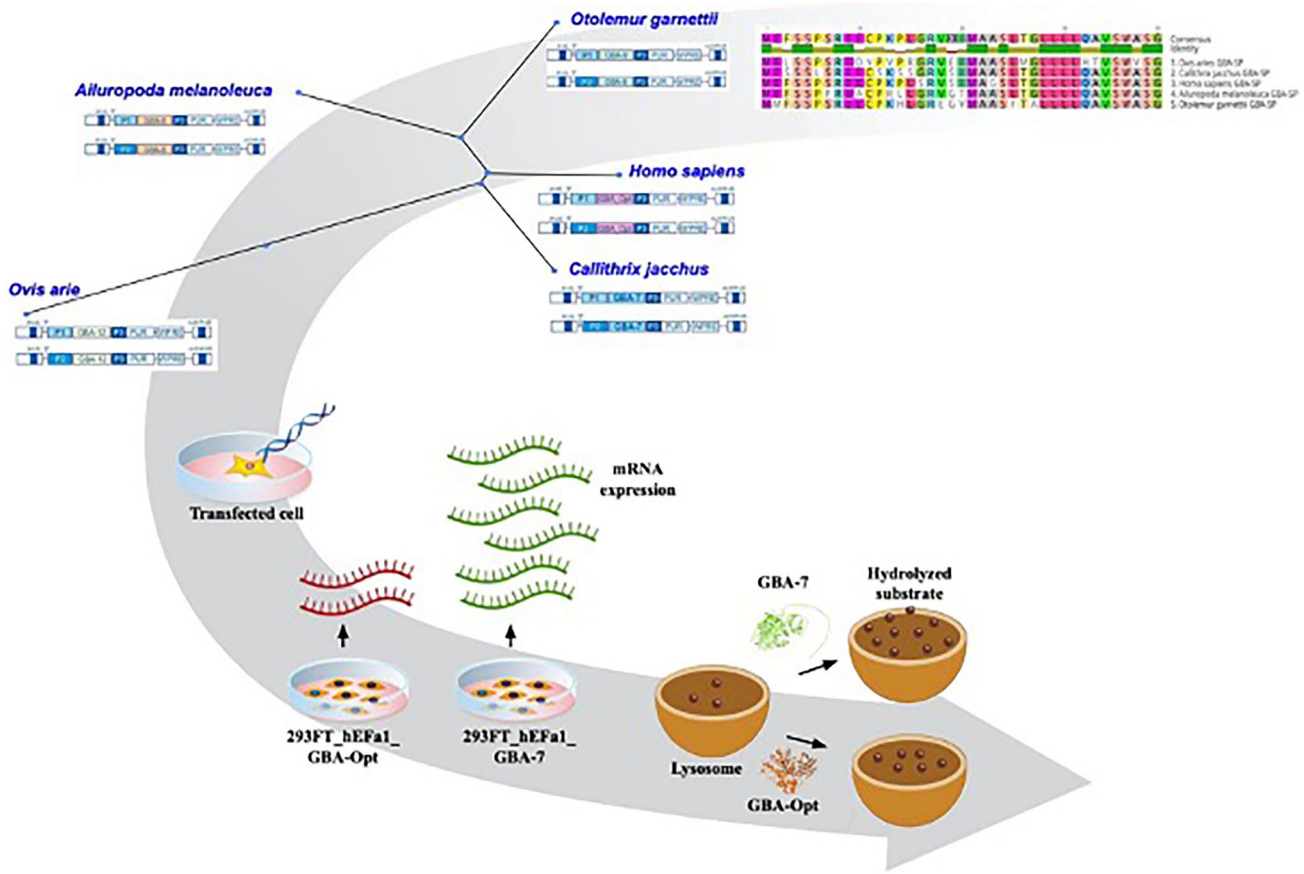
✉ Aparecida Maria Fontes, aparecidamfontes@usp.br | [1]Department of Genetics, Ribeirão Preto Medical School – University of São Paulo, Ribeirão Preto, SP 14049-900, Brazil. [2]Department of Psychobiology, Escola Paulista de Medicina, Federal University of São Paulo, São Paulo, SP 04024-002, Brazil. [3]Center for Cell-Based Therapy, Blood Bank of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP 14049-900, Brazil. [4]Department of Computer Science, Institute of Mathematics and Computer Science – University of São Paulo de São Paulo, São Carlos, SP 13566-590, Brazil. [5]Center for Artificial Intelligence, C4AI, University of São Paulo, São Paulo, SP 05508-010, Brazil. [6]Center of Excellence in New Target Discovery, Butantan Institute, São Paulo 0550390, Brazil. [7]Department of Biochemistry and Immunology, Federal University of Minas Gerais, Belo Horizonte, MG 31270-901, Brazil. [8]Department of Genetics, Ecology and Evolution, Federal University of Minas Gerais, Belo Horizonte, MG 31270-901, Brazil. [9]Bacterial Diseases Laboratory, Postgraduate Program in Animal Science in Tropics - Federal University of Bahia, Salvador, BA 40110-909, Brazil. [10]Hospital Distrital Maria José Barroso de Oliveira Parangaba, Fortaleza, CE 61923-075, Brazil. [11]Synthetic and Systems Biology Laboratory for Biomedicine, Istituto Italiano Di Tecnologia-IIT, Largo Barsanti E Matteucci, 80125 Naples, Italy. [12]Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 01239, USA. [13]School of Medicine, University Hospital Medical Center, Case Western Reserve University, Cleveland, OH 44106, USA.

Further studies are needed to assess its functionality in a GD animal model. Altogether, our findings provide in vitro evidence of the potential of this engineered enzyme for improved therapeutic effects for GD.

Article highlights.

Conservation and coevolution analysis of amino acid frequencies in GBA1 homologs of pathogenic variants associated with Gaucher disease, Parkinson' disease risk or Dementia with Lewy bodies risk.

Graphic Abstract



Article Highlights

(1) A novel synthetic recombinant GCase enzyme engineered with missense mutations in the signal peptide from selected mammalian species
(2) Correlation between transfection efficiency and catalytic activity indicates possible intrinsic sequence features influencing DNA uptake and gene expression

Keywords Gaucher disease · Glucocerebrosidase · Synthetic biology · In silico molecular evolution · Signal peptide · Enzyme coevolution · CMV promoter · hEF1a promoter · 293FT cells

1 Introduction

Lysosomal storage diseases (LSDs) are a group of over 70 rare inborn errors of metabolism that perturb lysosomal homeostasis [1, 2]. At the cellular level, they are characterized by the accumulation of proteins, polysaccharides and lipids in the lysosomes due to enzyme deficiencies, primarily lysosomal hydrolases [3]. As they are monogenic diseases, the reintroduction of a functional enzyme into the lysosome is in principle a viable treatment strategy for LSD [4]. Enzyme replacement therapy (ERT) has emerged as an effective treatment for nine LSDs, with ongoing clinical trials exploring its efficacy in five other LSDs [2].

Human β-glucocerebrosidase (GCase, EC 4.2.1.45), also known as β-glucosylceramidase or lysosomal acid β-glucosidase (LBG), is a well-characterized 60 kDa lysosomal enzyme and defective forms of GCase are responsible for Gaucher disease (GD) [5, 6]. The GCase enzyme is codified by the *GBA1 g*ene, a 7.6 kb gene localized on chromosome 1q.21 [7, 8]. Adjacent to *GBA1*, approximately 16 kb downstream, there is a highly homologous 5.7 kb pseudogene [7, 9]. The GBA mRNA contains two in-frame ATG translational start sites and both are translated, producing two peptides with different size signal peptides [10]. The protein derived from the first ATG exhibits a 39-residue leader, while the second ATG results in a 19-residue leader sequence. The mature glycoprotein, derived from the first ATG codon, possesses a longer 39-residue leader sequence that directs the nascent polypeptide to the endoplasmic reticulum (ER) membrane [10, 11]. This signal sequence interacts with the signal-recognition particle (SRP) and the SRP receptor in the ER membrane, initiating translocation into the ER lumen [12, 13]. Like other lysosomal enzymes, GCase is synthesized in the rough endoplasmic reticulum (RER) and transported through the Golgi apparatus to the lysosome by the lysosomal integral membrane protein type 2 (LIMP-2), independent of the mannose 6-phosphate [14]. The interaction between GCase and LIMP-2 occurs in the pH-neutral environment of the ER to form a complex that passes through the Golgi and eventually reaches the lysosome [15–17]. Additionally, it has been shown that LIMP-2 also has allosteric effects on GCase, which leads to activity increase in vitro, and works as a chaperone, maintaining the appropriate lysosomal response [18].

GD is classified into three clinical subtypes: type 1 GD (non-neuropathic), type 2 GD (acute neuropathic) and type 3 (chronic neuropathic) [19]. To date, more than 700 pathogenic variants are associated with Gaucher Disease [20], some of which are also linked to Parkinson's disease (PD) risk [21, 22] and Lewy body dementia risk [23, 24]. There are two common pathogenic GBA1 variants found in patients. The mutation p.N409S (N370S), found in patients with type 1 GD, and p.L483P (L444P) which is associated with GD type 2 and 3 [25]. Among patients with GD3, mutations p.N227S (N188S), p.V433L (V394L), p.G416S (G377S) and p.D448H (D409H) are frequently found [26, 27]. Moreover, polymorphisms in GBA1 that do not cause GD, such as p.E365K (E326K) and p.T408M (T369M) have been found in patients with PD [20, 26].

Recombinant human GCase (rGCase) has become the standard of care for patients with symptomatic Type 1 GD [28]. There are three commercial rGCase which are produced in Chinese hamster ovary (CHO) cell lines (Imiglucerase, Cerezyme; Sanofi Corporation, Cambridge, MA, USA; Food and Drug Administration-approval, 1994 [29–32]), human fibrosarcoma cell line (Velaglucerase Alfa; VPRIV; Takeda Pharmaceuticals USA, Inc., Lexington, MA, USA; FDA-approval, 2010 [33, 34]) and carrot cells (Taliglucerase Alfa; Elelyso; Pfizer, Inc., New York, NY, USA; FDA-approval, 2012 [35]).

In addition to three commercially available rGCase [28–36], several alternative strategies have been described for the development of recombinant GCase; among them, rGCase using retroviral vector and lentiviral vectors [37, 38], rGCase produced in CHO-DXB11 cells followed by methotrexate (MTX) amplification [39], a peptide-linked rGCase to the delivery to neural cells [40], a rGCase for production in *Nicotiana benthamiana* root culture [41, 42], a fused rGCase using an exosome-targeting transmembrane protein [43], a rGCase using adeno-associated virus 9 [44], and rGCase mutants developed with the use of computer-based algorithms to stabilize the enzyme [45].

In order to improve the catalytic activity of engineered enzymes, we have used directed evolution, as in the case of the O^6-alkylguanine-DNA alkyltransferase (AGT) repair protein for cancer gene therapy [46–48]. These engineered enzymes have also been used as a drug-resistance gene to optimize lentiviral vectors for HIV and sickle disease gene therapy [49, 50]. Additionally, engineered enzymes such as iduronate-2-sulfatase (IDS) have been developed for the treatment of Hunter syndrome-α lysosomal storage disease [51], and α-galactosidase (GLA) for the treatment of Fabry disease [52]. Furthermore, researchers have used synthetic biology technologies for programming, controlling transgene expression [53, 54] and for synthesizing codon-optimized DNA constructs for gene therapy for several genetic diseases [55–57].

Based on these considerations, in the present study, we evaluated the potential of integrating in silico molecular evolution, synthetic biology, and gene therapy approaches to develop a novel synthetic and recombinant GCase

for enzyme replacement therapy for GD. We engineered and assembled a total of 10 lentiviral vector plasmids, each designed to express a unique GCase variant under the control of either the hEF1a or CMV promoter in 293FT cells. Additionally, we constructed GFP-labeled lentiviral vector plasmids to evaluate transfection efficiencies. After transfection, we analyzed the mRNA expression levels and the catalytic activity of these GCase variants. This novel GCase variant was characterized through 3D structural analysis. We also evaluated conserved and coevolved patterns in the protein family to identify potential mutagenesis targets comparing them to known mutations at specific sites within the mature protein.

2 Materials and methods

2.1 Construction of two recombinant and synthetic lentiviral vector plasmids that encode optimized human GCase with R534H missense mutation under the control of two different promoters

First, we generated the optimized cDNA (GBA-Opt). This cDNA is transcribed to mRNA of GCase from *Homo sapiens* (NM_000157.3), which is 2,324 bp in length. We included the Kozak sequence after 5'UTR (GCCACC), therefore the final cDNA that encodes *Homo sapiens* GCase has 2330 bp and includes: 5'UTR (1–166 bp); Kozak sequence (167–172); coding sequence (173 – 1,783 bp); and 3'UTR (1784—2330). Also, at codon 534, we included the R534H mutation, Genzyme Patent: US 5,549,892A [58] (see Table S1). The cDNA sequence of the GCase enzyme was synthesized by Integrated DNA Technologies (IDT). All mRNA molecules were optimized using the online tool IDT Codon Optimization (https://www.idtdna.com/CodonOpt).

We obtained two synthetically derived double-stranded DNA fragments of ~ 1,000 bp and used a PCR-based overlap strategy, gBlock 1 (1–1078) and gBlock 2 (1047–2330), to generate the full-length *GBA1* cDNA (2,230 bp). The PCR product was extracted from agarose gel according to the manufacturer's instructions (QIAGEN). The gel-eluted PCR product (2330 bp) was analyzed by electrophoresis, and the sample concentration was estimated spectrophotometrically with NanoDrop 1,000 (Thermo Scientific, Inc). Gibson assembly was used to obtain a 5'entry vector (pENTRY) with human-optimized *GBA1* (GBA-Opt) cDNA. The pENTRY-GBA-Opt was sequenced bi-directionally using standard M13 forward and reverse primers. Next, two lentiviral vector plasmids were generated based on the strategy of a multisite gateway system as previously described [59]: (1) pLV-hEF1a-GBA-Opt_UbC-puromycin resistance gene (BP reaction with pENTRY-hEF1a + pENTRY_GBA-Opt in a pLV-Destination_UbC-puromycin) and (2) pLV-CMV-GBA-Opt_UbC-puromycin (BP reaction with pENTRY-CMV + pENTRY_GBA-Opt in a pLV-Destination_UbC-puromycin). Geneious Prime 2023.2.1 software was used to generate these synthetic lentiviral DNA constructs in silico.

2.2 Construction of eight recombinant and synthetic lentiviral vector plasmids that encode human GCase variants under the control of two different promoters

We implemented an evolutionary process to generate new GCase variants, and incorporated specific mutations at signal peptides from four selected mammalian species described in Fig. 1 (see Table S1). The sequences were retrieved from EMBL-EBI: (1) *Callithrix jacchus*—ENSCJAG00000008613, (2) *Otolemur garnetti*—ENSOGAG00000031511, (3) *Ailuropoda melanoleuca*—ENSAMEG00000000077 and (4) *Ovis aires*—ENSOARG00000003937. Geneious Prime 2023.2.1 was used to generate these synthetic lentiviral DNA constructs. We obtained four different gBlock 1 DNA fragments, overlapping with the same gBlock 2 DNA fragment described earlier. Each full-length *GBA1* variant obtained by PCR-based overlap strategy was purified by gel agarose. The gel-eluted PCR product (2230 bp) was analyzed by electrophoresis, and the sample concentration was estimated spectrophotometrically using NanoDrop 1000 (Thermo Scientific, Inc). Gibson assembly was used to obtain 5'entry vectors with the cDNA of each *GBA1* variant. The pENTRY-GBA-7 (*Callithrix jacchus* variant), pENTRY-GBA-8 (*Ailuropoda melanoleuca* variant), pENTRY-GBA-9 (*Otolemur garnetti* variant), and pENTRY-GBA-12 (*Ovis aires* variant) were sequenced bi-directionally using standard M13 forward and reverse primers. Next, in addition to the two lentiviral vector plasmids just described we generated eight lentiviral destination vectors based on the same multisite gateway system strategy: (3) pLV_hEF1a-GBA-7_UbC-puromycin, (4) pLV_CMV-GBA-7_UbC-puromycin, (5) pLV_hEF1a-GBA-8_UbC-puromycin, (6) pLV_CMV-GBA-8_UbC-puromycin, (7) pLV_hEF1a-GBA-9_UbC-puromycin, (8) pLV_CMV-GBA-9_UbC-puromycin, (9) pLV_hEF1a-GBA-12_UbC-puromycin, and (10) pLV_CMV-GBA-12_UbC-puromycin.

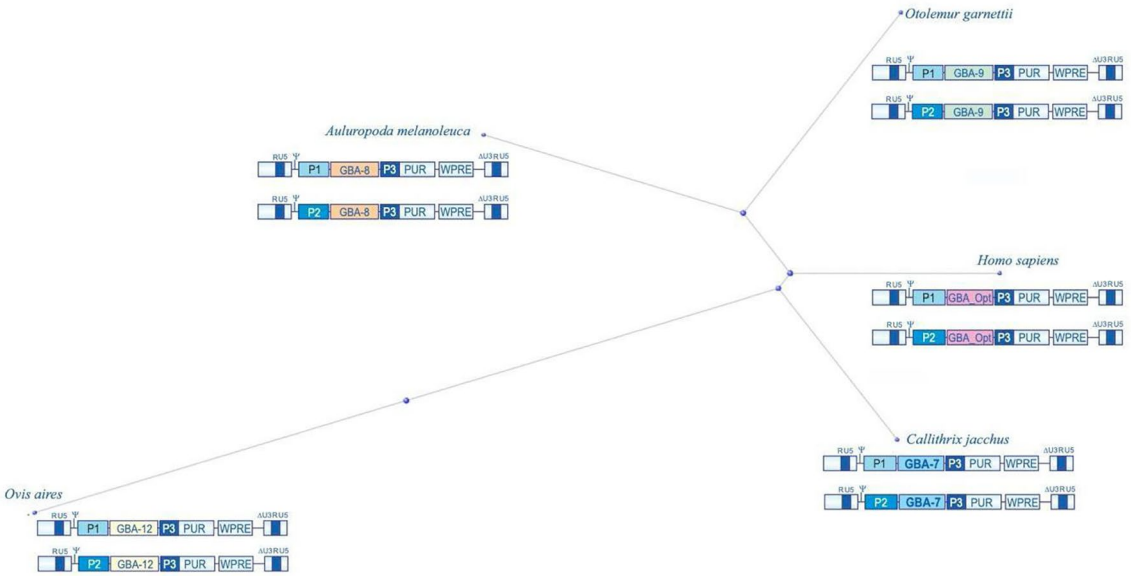


Fig. 1 The phylogeny tree of 5 mammalian GBA proteins is based on the signal peptide (SP) sequences. The phylogeny tree represents the evolutionary distances, measured as the number of substitutions per amino acid residue at the SP region of each GBA protein. The schematic drawing of each lentiviral expression vector that encodes GBA under the control of two different promoters (P1 = cytomegalovirus promoter [CMV]) and P2 (Human translation elongation Factor 1-alpha promoter [HeF1a]) is described. Each DNA plasmid also encodes the puromycin-resistant gene (PUR) driven by the Ubiquitin C (UBC) promoter (P3). The phylogeny tree was constructed using the COBALT multiple sequence alignment tool [60]

We also generated two synthetic lentiviral vector plasmids with the green fluorescent protein (GFP) gene under hEF1a and CMV promoters in order to investigate the cell transfection efficiencies in 293FT cells. These lentiviral vector plasmids are named: (11) pLV_hEF1a-GFP_UbC-puromycin and (12) pLV_CMV-GFP_UbC-puromycin.

2.3 Plasmid preparation and cell culture

DH5a Escherichia coli K12 strain was used to generate the recombinant bacteria with the respective lentiviral vector plasmid, and the Stbl-3 *Escherichia coli* HB101 strain was used to generate the recombinant bacteria with mock-DNA plasmid. Recombinant colonies were cultured at 37 °C in Luria–Bertani (LB) broth medium supplemented with 100–150 µg/mL of ampicillin. The DNA plasmids were isolated and purified using the QIAPrep Spin Mini kit (Qiagen, USA). The quality of DNA plasmids was assessed through a Nanodrop spectrophotometer based on the A260/A280 ratio and agarose gel electrophoresis.

The 293FT human cell line was purchased from Thermo, USA (R700-07). Cells were maintained in DMEM (Dulbecco modified Eagle's medium) supplemented with 10% fetal bovine serum (FBS) (HyClone, USA), 1% Penicillin/Streptomycin/L-glutamine (Sigma-Aldrich), and 1% non-essential amino acids (HyClone). Cell culture was maintained at 37 °C in a humidified incubator with a 5% CO_2 atmosphere. Cell growth was monitored in a monolayer every two days by microscopy (phase contrast or fluorescence), and the cell culture medium was refreshed. This study was approved by the Ethics Research Committee of the Clinical Hospital of Ribeirao Preto and Ribeirao Preto Medical School of the University of Sao Paulo under protocol number 1036/2022.

2.4 Folding of synthetic and recombinant mRNAs

The mRNA secondary structure prediction of synthetic and recombinant RNAs was obtained using ViennaRNA Web Services [61].

2.5 293FT cell transfection

For transfection experiments, 293FT cells were seeded at a density of 7×10^4 cells per well in a 24-well tissue culture plate, containing 1 mL of complete medium 24 h before the experiment. Transfections were carried out using a final concentration of 400 ng of DNA lentiviral vector plasmid per well and Lipofectamine 3,000 reagent (ThermoFisher Scientific, USA) in a final volume of 500 µL of complete medium, following the manufacturer's instructions. After 24 h post-transfection, an additional 500 µL of media was added to the cells. Cells were harvested for qRT-PCR and specific enzyme activity analysis after 48 h. For negative control experiments, both non-transfected 293FT cells and 293FT cells transfected with mock-DNA lentiviral vector plasmids were used for qRT-PCR and enzymatic assays, respectively. For each analysis, five technical replicates were performed.

In order to maintain the final amount of 400 ng of DNA plasmid in the experiments with double transfections (GFP and *GBA1* lentiviral vector plasmids), 200 ng of each DNA plasmid was used. For the negative control experiments, 293FT cells were transfected with GFP and mock-DNA lentiviral vector plasmids. For flow cytometry, four individual experiments were performed. Similarly transfected 293FT cells were analyzed 48 h post-transfection by fluorescence microscopy and flow cytometry. The EVOS Automated Cell Imaging System (ThermoScientific) was used to obtain the fluorescence microscopy images. For flow cytometry, cells were washed with PBS, trypsinized, centrifuged at 800 × g for 5 min at 4 °C, and resuspended in PBS with 2% FBS. The percentage of GFP-positive cells was assessed using a BD FACS Canto flow cytometer (BD Biosciences, San Jose, CA, USA) with FL1 channel (excitation laser: 488 nm; emission filter: 525 nm), and the data was analyzed with FlowJo software (FlowJo, LLC, USA).

2.6 Characterization of transgenic 293FT cells with transient expression of recombinant and synthetic GBA

The GBA mRNA expression levels in 293FT transgenic cell lines with transient production of GCase were investigated by qRT-PCR. Total RNA was isolated using the E.Z.N.A. Total RNA Kit I (Omega, Bio-Tek) according to the manufacturer's instructions. Based on the 260 and 280 nm absorbance ratios, the RNA purities were confirmed with a NanoDrop 2,000 spectrophotometer (Thermo Fisher, Germany). For cDNA synthesis, 500 ng of total RNA was utilized with the PrimeScript RT-PCR Kit (Takara Bio). Following the manufacturer's instructions, total RNA was incubated with gDNA Eraser for 2 min at 42 °C and 5 min at room temperature (RT). Subsequently, cDNA was synthesized by incubation for 15 min at 37 °C, and the reaction was stopped after 5 min at 85 °C. To quantify the relative abundance of GBA mRNA levels, qRT-PCR was performed with 1:10 diluted cDNA and specific primers obtained from Sigma, designed using the Primer-Blast from NCBI [62]. The primer sequences are GBA-F1297 5′-CTTCGCTTCTGAGGCTTGCG-3′ and GBA-R1376 5′-ACTGCATCCCCGATCCCTG-3′. Quantitative real-time PCR was carried out with SYBR green Master Mix (Thermo Fisher Scientific), and samples were loaded in a MicroAmp Fast Optical 96-Well Reaction Plate (0.1 ml) with the QuantStudio 3 Real-Time PCR Machine (Thermo Fisher Scientific). Briefly, the reaction mixture (10 µl) consisted of 5 µl SYBR Green Master Mix 2X, 2 µL ddH2O, 1 µl of each primer, and 1 µl of template. A control reaction without a template (blank) was set. The PCR reaction was started at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. The reactions were run in duplicate. Transcription levels of the GBA transcripts were normalized by 18S ribosomal RNA (rRNA) levels. The forward primer was 5′-TGTGCCGCTAGAGGTGAAATT-3′, and the reverse primer was 5′-TGGCAAATGCTTTGCTTTCGTTT-3′. A melting curve analysis confirmed the amplification of only one product in qRT-PCR. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [63] and compared with the wild-type 293FT cell line.

2.7 Lysosomal recombinant and synthetic GCase activity assay

Lysosomal glucocerebrosidase activities were assayed using the synthetic fluorescent substrate 4-Methylumbelliferyl-beta-D-glucopyranoside (4-MU-β-glc) in the presence of sodium taurocholate according to the method described by [64] and adapted from [65]. The enzymatic reaction was initiated by mixing 30 µl of cell lysate (40–60 µg total protein from transfected or control 293FT cells) with 50 µL of 20 mM 4-MU-β-glc (Sigma, M3633), and 20 µL of 1% sodium taurodeoxycholate hydrate (Sigma, T0875) in a buffer with 0.5 M citrate–phosphate at pH 5.0. The reactions were incubated for 2 h at 37 °C in an orbital shaker and terminated by adding 2.0 mL of 0.25 M glycine–NaOH buffer, pH 10.3, at room temperature, followed by incubation on ice for 5 min. The same protocol was used for the blank mixture. The relative fluorescence from 200 µl of the reaction product was measured in a black 96-well microplate. The excitation and emission

wavelengths of 360 and 450 nm were employed using the Cary Eclipse Fluorescence Spectrophotometer. The results were compared with a 4-MU standard curve. The enzymatic activity, as described in the literature [64], was expressed as nmol of hydrolyzed substrate per mg of protein per 1 h. As a reference, the measurement of leukocyte lysate was performed, and the values between 8.68 and 11.57 nmol/ mg/ h were used following the suggestions established in [64].

2.8 Cell transfection efficiencies versus GCase-specific activity

To understand the differences between the specific values of activity of GCase variants, we measured transfection efficiency based on the fluorescence of cells after double transfection with GFP plasmid plus GBA1 variant plasmid. Quantification of the GFP fluorescence after double transfection (GBA1 variant + GFP under the same promoter) was normalized by GFP fluorescence after control double transfection (mock plasmid + GFP), and plotted on the X-axis. Additionally, we calculated the specific activity for each GCase variant in transgenic 293FT cells after transfection with the GBA1 variant and obtained the ratio normalized by 293FT cells transfected with mock-DNA lentiviral vector plasmid. Then, we plotted the ratio of specific enzyme activity versus the transfection ratio for each GCase variant. For 293FT cells transfected with mock-DNA plasmid, the slope of the line was about 1.0. This slope indicates little difference compared with the mock plasmid. Conversely, a slope larger than 1.0 suggests an increased catalytic activity compared with the mock.

2.9 Conservation, coevolution, and structural analysis on GBA1 and its homologs

We sought to conduct in silico screening to evaluate conserved and coevolutionary patterns between residues in the GCase protein sequences. We obtained 1569 GBA1 homologous sequences from The Hierarchical Catalog of Orthologs, OrthoDB [66] (Group 3473901at2759 – Glucosylceramidase at Eukaryota level) and 242 sequences from EMBL-EBI, which were aligned using Muscle [67]. The multiple sequence alignment (MSA) was manually curated to remove fragments and poorly aligned sequences, and processed to remove redundancy (utilizing an 80% identity cutoff). This process resulted in a final set of 494 sequences, which were then used to assess conservation and coevolution using PFstats [68]. Because signal peptides are poorly alignable over large number of sequences in different taxa, the analysis focused on the mature protein sequences, which could suggest how conservation and coevolution patterns could give information on previously described mutations [5, 27, 69–71].

Coevolution analysis was based on the criteria that pairs of residues were considered coevolved if both were present in at least 20% of the sequences, as determined by the minimal sub-alignment determined using the Dima-Thirumalai test [72]. Additionally, a residue pair was considered coevolved if the presence of one residue increased the frequency of the other to 80%, with an associated probability for this frequency shift being smaller than 10^{-10} [72].

For visualization purposes, we utilized the crystal structure of human GCase (PDB: 2F61) to illustrate some residues throughout the protein. Three-dimensional structural analysis was performed using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

2.10 Synthetic and recombinant GCase three-dimensional structure prediction

We predicted and studied the molecular structures of GCase-Opt and GCase-7 enzymes using ChimeraX [73]. ChimeraX offers a convenient notebook feature that integrates with AlphaFold2, facilitating the prediction of the three-dimensional structures [74].

As a first step in prediction, we performed a BLAST search for GCase-Opt and identified the closest aligned protein, PDB: 2F61—the crystal structure of partially deglycosylated acid beta-glucosidase [75]. Next, we utilized ChimeraX in conjuction with AlphaFold2 to predict the three-dimensional structures of GCase-Opt and GCase-7 enzymes.

2.11 Statistical analysis

The statistical analyses were done using the GraphPad Prism version 10.0.0 (Dotmatics, LLC). All experimental values are presented as the mean ± standard error of the mean (SEM) and derived from 3 to 5 independent experiments. Significance values were indicated as follows: * for $P < 0.05$, ** for $P < 0.01$, and *** for $P < 0.001$ by two-tailed tests.

Table 1 Missense mutations from the synthetic and recombinant enzyme variants at the signal peptide and position 534

GCase variants	Missense mutations at signal peptide and position 534
GCase-Opt	Signal peptide as NP_000148 + **R534H**
GCase-7	F3S, P6L, P12S, P14S, L15S, **S16G**, **G23A** + **R534H**
GCase-8	S7P, E10A, K13R, P14L, **S16G**, S19G, I20T, **G23A** + **R534H**
GCase-9	E2M, P14H, **S16G**, V18E, S19G, I20V, **G23A,** L25F, G27A + **R534H**
GCase-12	F3L, E10D, C11Y, K13V, L15R, **S16G**, S19E, **G23A**, T26M, Q32H, A33T, A37V + **R534H**

The two missense mutations, S16G and G23A, indicated in bold, are shared among the four GCase variants with missense mutation at the signal peptide. For all synthetic and recombinant GBAs in this study, we included the R534H missense mutation present at the Sanofi Corporation patent

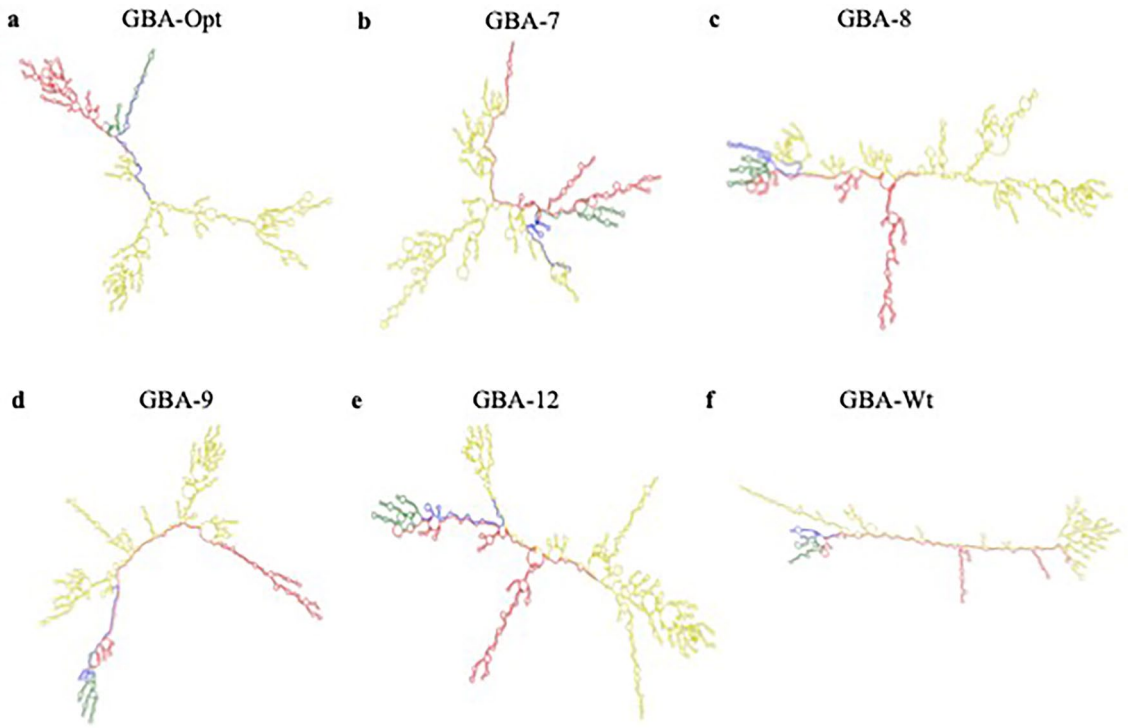


Fig. 2 mRNA secondary structures of human GBA variants. **a**–**f** Predicted lowest free energy secondary structures of mRNA from GBAs generated in this study with optimized codons (GBA-Opt, GBA-7, GBA-8, GBA-9, and GBA-12) (**a**–**e**) and mRNA from wild-type GBA (**f**). The nucleotides are colored according to their regions in the mRNA molecules: 5′UTR (green), signal peptide (blue), coding sequence (yellow), and 3′UTR (red)

3 Results

3.1 Secondary structure analysis of codon-optimized mRNA GBA1 variants

In this study, four lysosomal β-glucocerebrosidase (GCase) variants were generated, each with different amino acid substitutions in the signal peptide. The respective GBA1 names assigned to these variants denote the number of amino acid substitutions (Table 1). Specifically, GBA-Opt encodes the signal peptide from human GBA (NM_000157.3/ NP_000148.2) and contains the missense mutation R495H (p.R534H), as documented in a Sanofi Corporation patent (Table 1).

The predicted lowest free energy secondary structures of these mRNA GBA1 variants and the secondary structure of the mRNA that encodes the wild-type GBA1 (with natural codons), were determined (Fig. 2**a**–**f**). Different regions including the 5′UTR, coding sequence, signal peptide, and 3′UTR are highlighted with different colors, allowing the visualization of the stem-loops and hairpin structures formed.

Moreover, the ΔG values for the GBA-Opt, GBA-7, GBA-8, GBA-9, GBA-12, and GBA-wild-type mRNAs were −829.400, −840.300, −835.500, −830.900, −822.200, and −826.100 kcal/mol, respectively. These data suggest that the GBA1 variants (GBA-OPT, GBA-7, GBA-8 and GBA-9) exhibit lower Gibbs free energy values, indicating relatively greater structural stability. Interestingly, GBA-7 mRNA showed the lowest Gibbs free energy value among the mRNAs with optimized codons.

3.2 Expression profile of codon-optimized synthetic and recombinant GBA1 variants reveals promoter-dependent transcription levels in 293FT cells

Next, we investigated mRNA expression levels of GBA-Opt, GBA-7, GBA-8, GBA-9, and GBA-12 in the same backbone expression vector and cell line. 293FT cells were transiently transfected to produce these GBA1 variants. Quantitative real-time PCR analysis revealed significant differences in expression levels between the hEF1a and CMV promoters (Fig. 3). It should be noted that 293FT virgin cells do not express the specific GBA mRNA molecules we are studying because our GBA mRNA is synthetic and contains optimized codons. Four GBA1 transcripts (GBA-Opt, GBA-7, GBA-9, and GBA-12) under the control of hEF1a promoter showed higher expression levels in 293FT cells compared to those under the CMV promoter ($p < 0,05$). However, the CMV promoter yielded the highest expression for the GBA-8 mRNA molecule ($p = 0.0046$). Among the GBA1 variants, we observed that 293FT_hEF1a_GBA-7 and 293FT_hEF1a_GBA-9 cells express 5.2-fold and 3.8-fold higher transcription levels of mRNA compared to 293FT_hEF1a_GBA-Opt ($p = 0.0003$ and $p = 0.0004$, respectively) (Fig. 3). For 293FT_CMV_GBA-7 and 293FT_CMV_GBA-9 cells, these values are 3.7-fold and 2.4-fold higher than transcription levels of mRNAs compared to 293FT_CMV_GBA-Opt ($p = 0.0012$ and $p = 0.013$, respectively). Additionally, GBA-7 mRNA under the control of the hEF1a promoter showed higher expression compared to the other GBA1 variants with missense mutations at the signal peptide ($p < 0.05$). These data demonstrate that the GBA1 variant with seven missense mutations at the signal peptide under the hEF1a promoter exhibits the highest expression levels in 293FT cells. Our findings demonstrate that the CMV promoter drives significantly higher mRNA expression levels for the GBA-8 variant compared to the hEF1a promoter, contrasting with other GBA variants where the hEF1a promoter is stronger. This suggests that the unique sequence and structural features of the first half of the GBA-8 cDNA, including

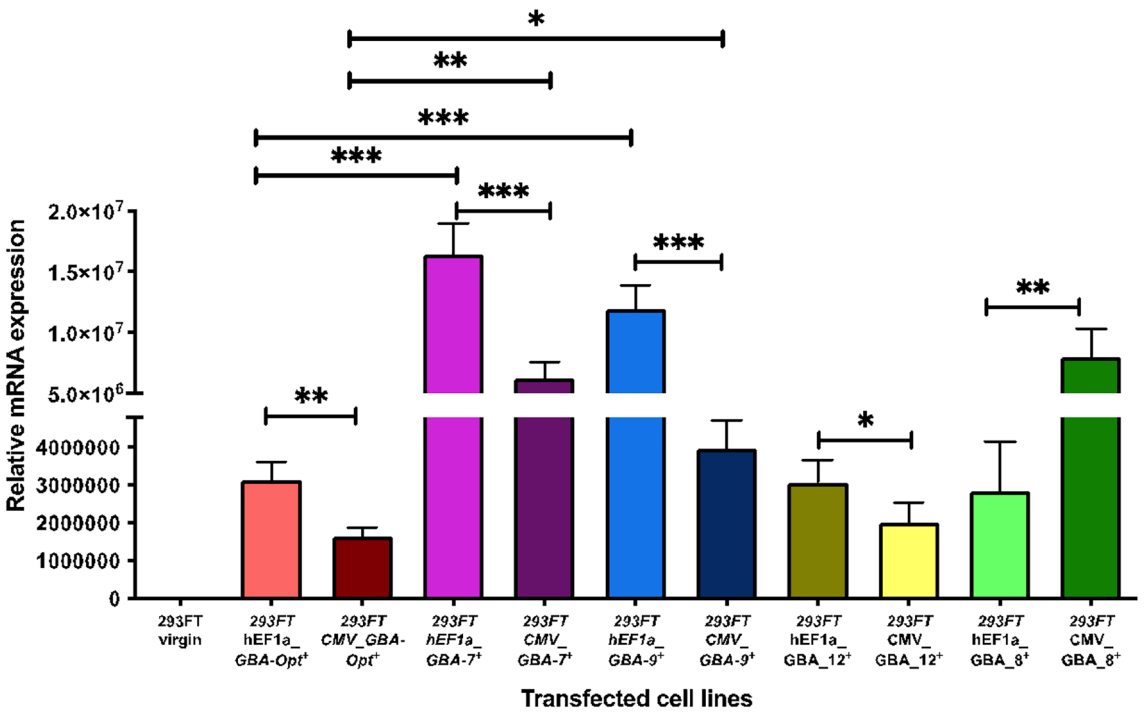


Fig. 3 mRNA expression levels of human GBA variants in transfected 293FT cell lines. Quantitative real-time PCR analysis of mRNA GBA variants in 293FT cells, 48 h post-transfection with the respective DNA lentiviral plasmid. Five cell cultures of each cell line transiently expressing the GBA variant were used (N = 5). The lines on the top of the bars represent a significant difference between hEF1a and CMV promoters and between GBA-Opt and GBA variants. The same color bars for transgenic cells as that in Fig. 6 are used. Error bars are the mean ± SD. Significant differences were determined by unpaired t-test with Welch's correction. *Asterisks* *, **, *** and ns (non-significant) represent $p \le 0.05$, $p \le 0.01$, $p \le 0.001$, and $p > 0.05$ respectively)

codon optimization and the signal peptide coding region, may influence the promoter's effectiveness. To further compare the interactions between unique sequence and structural features of GBA-8 cDNA and promoter activity, additional promoters should be tested.

3.3 Differences in transfection efficiency of GBA1 variants in 293FT cells

In order to investigate whether or not these differences in mRNA levels among GBA1 variants are associated with the GBA1 sequence itself rather than transient transfection efficiencies, we conducted co-transfections of 293FT cells with both GFP and GBA1 lentiviral vector plasmids. GFP-expressing 293FT cells were visualized using fluorescence microscopy 48 h post-transfection (Fig. 4). Flow cytometry analysis revealed that the transfection efficiencies ranging from 43.5% to 64.4% in 293FT cells (Fig. 5).

Interestingly, both GBA-Opt and GBA-7 constructs, independent of the hEF1a and CMV promoters, demonstrated similar transfection efficiencies ($p > 0.05$) (Fig. 5g–j). However, for GBA-8, GBA-9, and GBA-12 constructs under the hEF1a promoter, flow cytometry analysis indicated improved GFP delivery efficiencies into 293FT cells compared with CMV promoter ($p < 0.05$) (Fig. 5 k–p). These results suggest that the observed differences in mRNA levels among GBA1 variants.

3.4 Development of human GCase with higher activity and the impact of residues from the signal peptide of GCase in the catalytic activity

Next, we sought to assess the catalytic function of these GCase variants produced by 293FT cells 48 h post-transfection. Using 4-MU-β-glc as a substrate, we observed that 293FT_hEF1a_GCase-7 and 293FT_hEF1a_GCase-Opt cells produce 507.6 (± 38.17) and 426.6 (± 25.26) nmol hydrolyzed substrate/mg-protein/h, respectively ($p = 0.0056$). As a control, we analyzed 293FT transfected with mock DNA plasmid, and the biological activity of GCase was 82.92 (± 5.83) nmol substrate/mg/h (Fig. 6).

Furthermore, we found that the levels of GCase are promoter-dependent; the 293FT cells transfected with GBA-7 and GBA-Opt under CMV promoter produced similar levels of GCase, approximately 278.2 (± 17.86) and 277.5 (± 17.16) nmol substrate/mg/h, respectively ($p = 0.95$). For the other GCase variants produced by 293FT_hEF1a-GBA-8, 293FT_CMV-GBA-8, 293FT_hEF1a-GBA-9, 293FT_CMV-GBA-9, 293FT_hEF1a-GBA-12, and 293FT_CMV-GBA-12 cells, specific enzymatic activities are: 101.1 (± 12.78), and 76.53 (± 7.88), 86.9 (± 11.69), 89.27 (± 23.77), 75.28 (± 6.09) and 52.35 (± 5.96) nmol

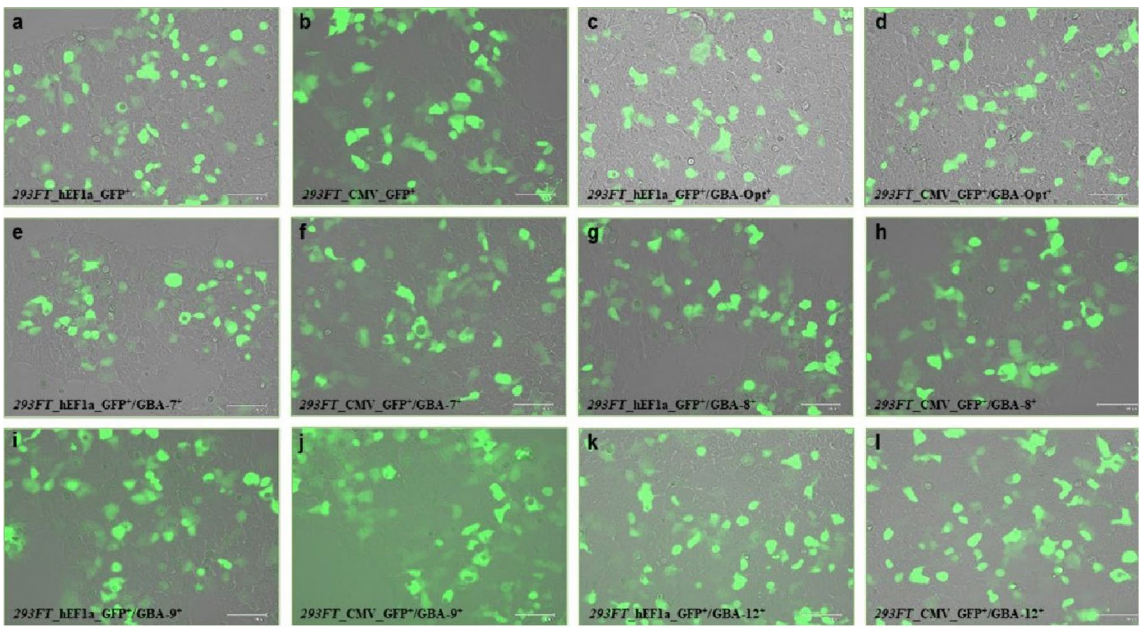


Fig. 4 Live-cell fluorescence of GFP protein in transgenic 293FT human cell lines transiently expressing GFP and GBA variants. A single representative image of 293FT_GFP.$^+$ cells 48 h post transfection with indicated constructs: GFP lentiviral vector plasmids (GFP under hEF1a or CMV promoter) (**a**, **b**) or two DNA lentiviral vector plasmids (GFP and GBA variant) (**c**–**l**) is shown. Four cell cultures of each transgenic 293FT cell lines were performed (N = 4). Untransfected 293FT cells or 293FT cells transfected with GFP + empty vector was used (not shown)

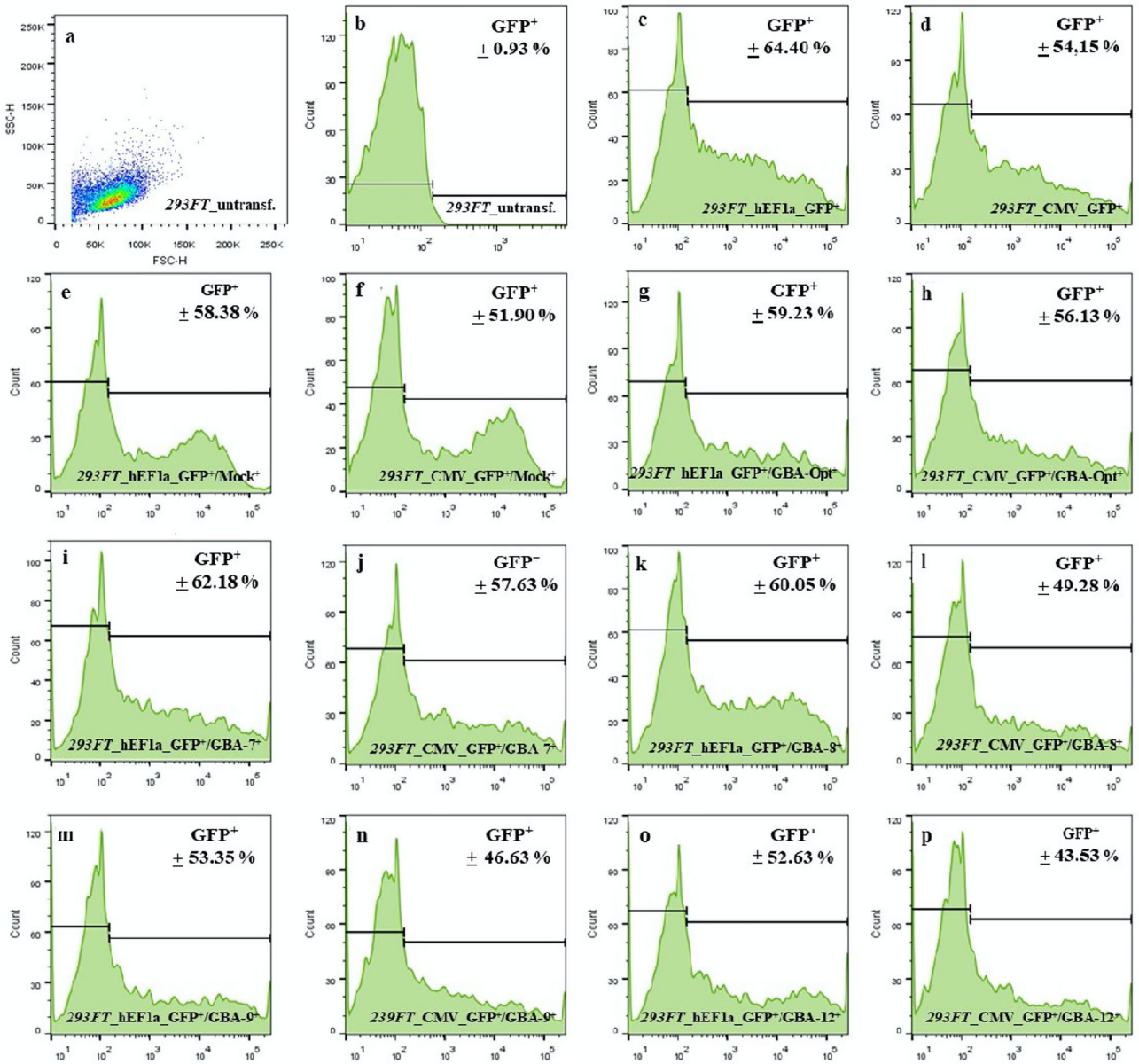


Fig. 5 Differences in transfection efficiencies among GBA lentiviral vector plasmids in 293FT cells 48 h post-transfection. **a** representative fluorescence-activated cell sorting dot blot light forward scatter/ side scatter profile of 293FT cells. **b** representative histogram of untransfected 293FT cells. **c** to p. representative histogram plots showing the percentage of GFP^+ cells 48 h post-transfection with indicated construct: GFP plasmid (under hEF1a or CMV promoter) (**c**, **d**) or two DNA lentiviral vector plasmids: (GFP and mock plasmid or GFP and GBA variant) (**e**–**p**). The mean of the flow cytometry analysis of four independent experiments is shown

substrate/mg/h, respectively. Except for 293FT_hEF1a-GBA-8 and 293FT_CMV-GBA-12 cells, in the other transfected cell lines, the levels of GCase variants are similar than those observed in 293FT cells transfected with mock-DNA plasmid, which produces 86.92 (± 5.83) (Fig. 6). Altogether, these data demonstrate that GCase-7 under the hEF1a promoter is produced at higher levels than GCase-Opt, and missense mutations in some amino acid residues of the signal peptide of GCase impair the catalytic activity.

Given that GCase-7 is the only GCase variant investigated here in which all proline residues from signal peptide (P6, P12, and P14) are replaced [P6L, P12S, and P14S] (Table 1), it may be worthwhile to investigate the biochemical characteristics of different missense mutations in GCase-7. Proline is known for its unique properties due to the fusion of its R-group to the α-nitrogen, which restricts mobility and causes proline-richer sequences to be more rigid [76]. Therefore, further studies are warranted to explore the effect of these mutations on chain conformation in this GCase variant.

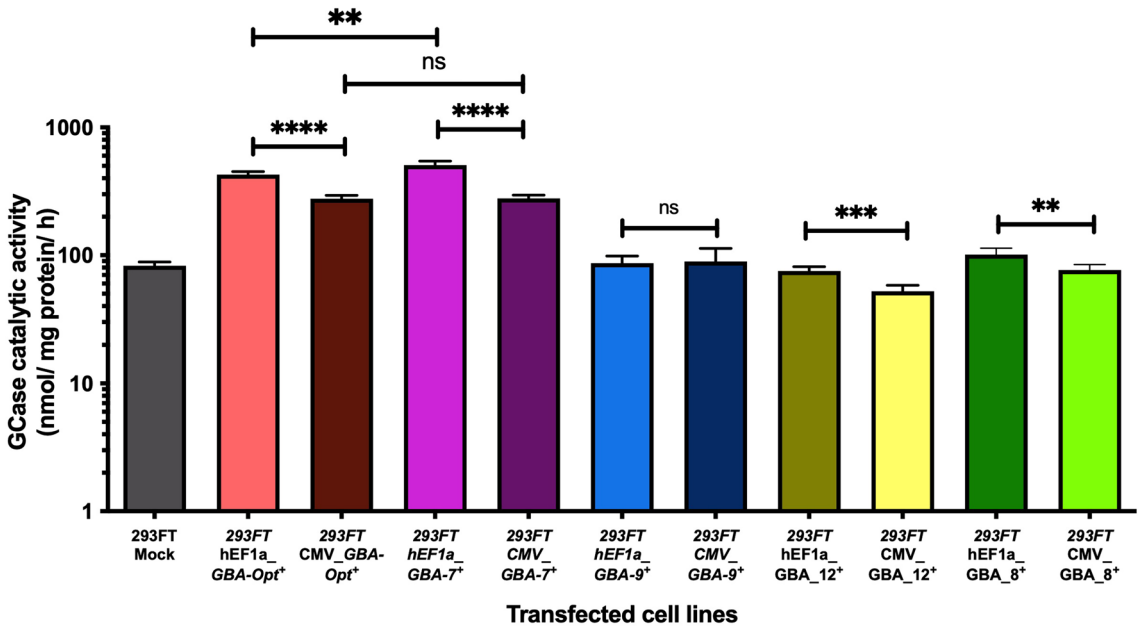


Fig. 6 Catalytic activity of synthetic GCase variants on 293FT cells. The fluorimetry assay determined the GCase catalytic activity using 4-MU-β-glc as substrate. Five cell cultures of each engineered cell line transiently expressing the GCase variant were used (N = 5). The lines on the top of the bars represent a significant difference between hEF1a and CMV promoters and between GCase-Opt and GCase-7 enzymes. Analyses of GCase activity on leukocytes were used as a positive control for activity assay (not shown) and as a negative control 293FT cells transfected with mock plasmid. The same color bars for transfected cells as that in Fig. 3 are used. Error bars are the mean ± SD. Significant differences were determined by unpaired t-test with Welch's correction. *Asterisks* *, **, *** and ns (non-significant) represent $p \le 0.05$, $p \le 0.01$, $p \le 0.001$, and $p > 0.05$ respectively

3.5 Analysis of transfection efficiency and catalytic activity correlation on 293FT cells among different GBA1 variants

To better understand the transfection efficiency and its correlation with catalytic activity, we generated a graph plotting the catalytic activity ratio *versus* the transfection ratio (Fig. 7). Considering that the transfection with mock-DNA plasmid serves as a reference value, and that the ratio values for hEF1a_GBA-7/DNA-mock, CMV_GBA-Opt/DNA-mock,

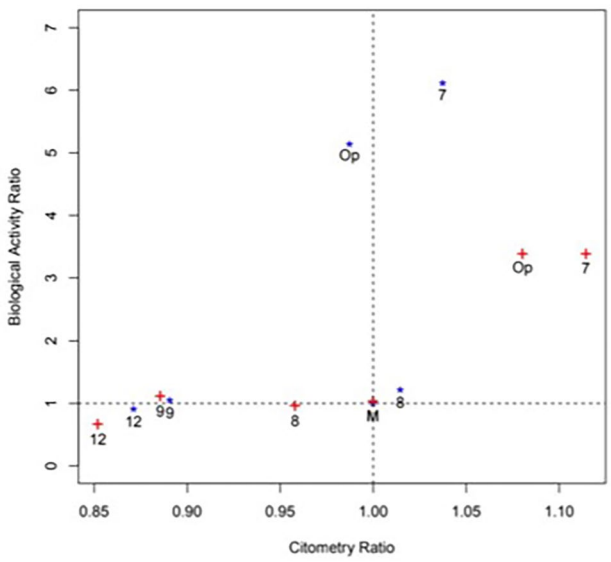


Fig. 7 Correlation of transfection efficiencies of GBA and GFP lentiviral plasmids and recombinant and synthetic GCase catalytic activities on 293FT cells. The scatterplot shows the transfection and catalytic activity ratios in transfected 293FT cells 48 h post-transfection with the indicated DNA construct. Blue and red indicate hEF1a and CMV promoters, respectively. The X-axis represents the ratio for transfection (GFP + GBA variant), and the Y-axis represents the ratio of catalytic activity for each GCase variant

and CMV_GBA-7/DNA-mock are higher than 1.0, these data indicate an increase in transfection capability associated with enhanced GCase variant catalytic activities for these variants.

Interestingly, although the hEF1a_GBA-Opt-DNA plasmid exhibited lower transfection efficiency, the GCase catalytic activity is higher than that of CMV_GBA-Opt and CMV-GBA-7. Based on this, we speculate that an intrinsic feature of the hEF_GBA-Opt construct sequence affected the DNA uptake andtransfection efficiency without affecting catalytic activity. Notably, among all GBA1 constructs, the GBA-7 sequence demonstrated improvements in both DNA release and catalytic activity for both promoters (Fig. 7). Therefore, the specific missense mutations present in the signal peptide might contribute to the higher catalytic activity of this synthetic and recombinant GCase variant.

3.6 Conservation and coevolution analysis of the GBA-1

We conducted a conservation analysis using 1569 genes from 793 species retrieved from The Hierarchical Catalog of Orthologs, OrthoDB plus sequences obtained from EMBL-EBI (*see Table S3*). After aligning and filtering the sequences, it was observed that the signal peptide exhibited poor alignment, while the regions corresponding to the mature protein could be subjected to further analysis.

Among 26 pathogenic variants either associated with Parkinson' disease risk or with Dementia with Lewy bodies risk, 50% showed high conservation (> 60%), while others ranged from 6 to 53% conservation (*see Table S3*). Specifically, p.R159W/Q (R120W/Q) occurs on a site with 99% conservation for arginines, while p.G416S (G377S) is situated within a site showing 92.1% conservation for glycines. Additionally, p.N227 (N188S) and p.K237 (K198) occur within sites with 72.9% conservation for asparagine and 73.7% for positively charged residues (K and R), respectively (see Table S3).

We further analyzed some residues whose mutations are the most common pathogenic variants, for example, p.N409S (N370S), and p.L483P (L444P), which occur on sites less conserved. The residue p.N409S (N370S) is located at a position that favors asparagine and aspartate, with conservation rates of 32.1% and 46.9%, respectively (see Table S3). Interestingly, it is not only the deviation from the pattern that results in a pathogenic variant. A mutation to aspartate in position 370, which conforms to the conservation pattern at this site, is predicted to enhance stability and activity [45]. The residue p.L483P (L444P) is located on a site with 39.4% conservation for leucine. However, when considering aliphatic residues (I, L and V), the site exhibits a high level of conservation (95,7%) (Fig. 8a) (see Table S3). The mutation to a proline not only contradicted the conservation pattern at this position, but is also likely to affect the structure since prolines are not favored in beta-sheets [76]. This mutation could disrupt the hydrogen bonds between neighboring sheets, particularly with p.L500 (L461), which is the last residue in the adjacent beta-sheet (Fig. 8a).

We also found evidence of coevolved amino acids in this enzyme family (see Tables S4 to S12). Of note, the amino acid at position p.D448H (D409H) (see Table S10) showed aspartates in half the sequences in the alignments, including human GBA1, but they were highly correlated with a serine (p.S136 [S97]) and formed a hydrogen bond (Fig. 8b). We observed that in sequences with a serine present in this position, the frequency of aspartates in position 448 increased from 53,54% to 81,4%, suggesting a conserved interaction pattern that would be lost upon the mutation p.D448H (D409H).

The residue p.E365K (E326K) is also less conserved (20.2%). Interestingly, p.E365K (E326K) has been identified as a risk variant for Parkinson's disease but does not cause GD [20]. The structure revealed that they are at the protein surface and form salt bridges (Fig. 8c). The p.E365K (E326K) is the central residue in the three-residue salt bridge cluster, including p.R368 (R329) and p.K332 (K293). Protein surface multi-residue salt bridge clusters may be related to protein stability [77], and mutating the central negatively charged residue to a positively charged one in mutation p.E365K (E326K) would completely disrupt this cluster.

The mutations p.R535H (R496H) and p.D513 (D474), with prevalences of 11,92% and 73,74%, respectively, interact with each other to form a salt bridge (Fig. 8d). Of note, p.R535H (R496H) appears as a compound heterozygous allele in patients with GD [20]. In principle, the mutation p.R535H (R496H) could still be possible, but that would depend on the protonation state of the histidine derived from the mutation Interestingly, as mentioned above, the preceding residue p.R534H (R495H) has been reported in enzyme replacement therapy, and was included in our recombinant and synthetic GCases.

3.7 GCase-Opt and GCase-7 synthetic and recombinant enzymes: 3D structure predictions

We further predicted and compared GCase-Opt and GCase-7 protein 3D structures against PDB: 2F61 chain A using the ChimeraX Matchmaker algorithm. We observed that both GCase-Opt and GCase-7 enzymes exhibit similar 3D structures compared with 2F61, with the exception of the signal peptides unique to GCase-Opt and GCase-7 (Fig. 9a-c). Furthermore, we observed a strong alignment between GCase-Opt and GCase-7 with 2F61, albeit with a slight deviation towards the

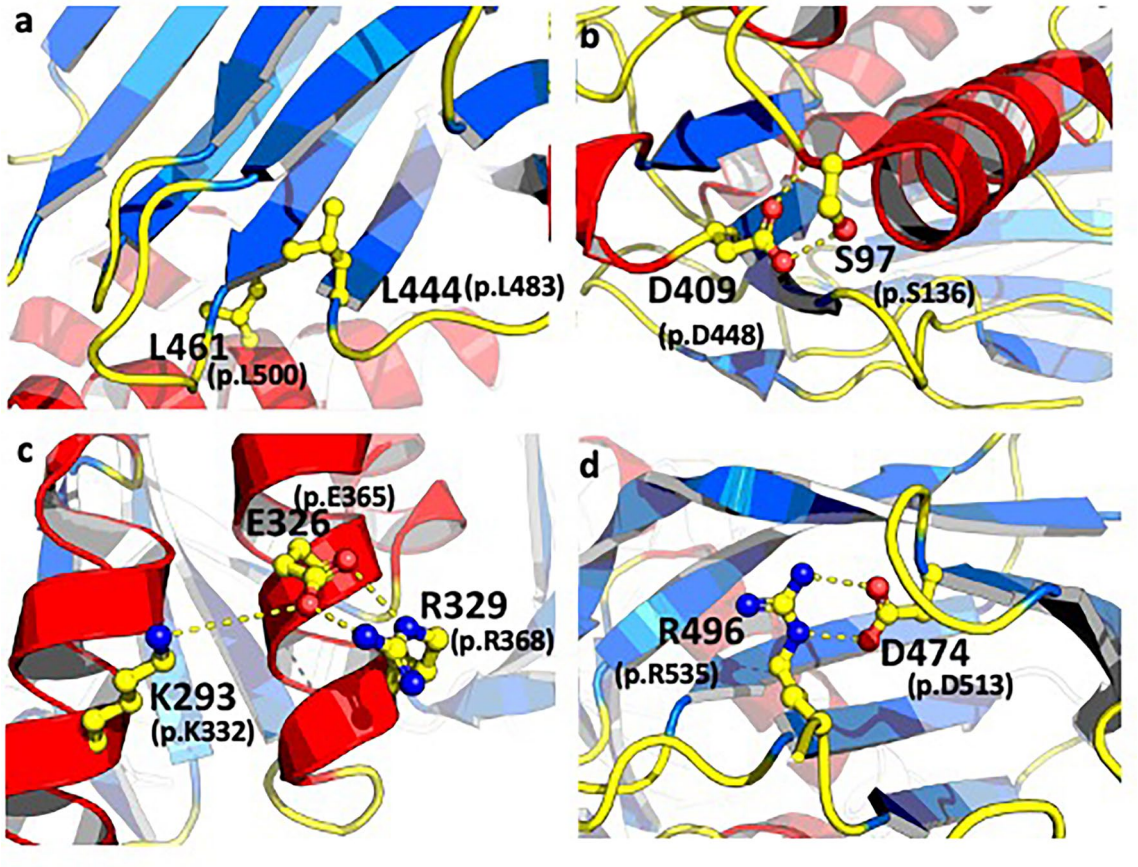


Fig. 8 Three-dimensional localization of GBA pathogenic variants in its crystal structure (PDB: 2D61). **a** three-dimensional configuration of residues p.L483P (L444P) and p.L500 (L461). **b** three-dimensional configuration of residues p.D448H (D409H) and p.S136 (S97) (hydrogen bonds shown as dashed lines). **c** A salt bridge network formed by residues p.E365K (E326K), p.R368 (R329) and p.K332 (K293) (salt bridges shown as dashed lines). **d** a salt bridge between residues p.R535H (R496H) and p.D513 (D474). The figure was prepared using The PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC)

signal peptide (root mean square distances [RMSD] of 0.528 and 0.485, respectively) (Fig. 9d-e). Moreover, the alignment between GCase-Opt x GCase-7 exhibited a lower RMSD value of 0.178, which can be influenced by the shared presence of the signal peptide in both GCase 3D structures (*see Table S13*). Importantly, minimal conformation changes were observed in the 3D structures of GCase-Opt and GCase-7, which may support their optimized catalytic activity (Fig. 9f).

In summary, through multiple lines of evidence obtained by in silico, in vitro, and culture cells experiments, we determine the structural and functional characteristics of synthetic and recombinant lysosomal β-glucocerebrosidase (GCase) variants, particularly GCase-Opt and GCase-7. We demonstrate that codon optimization enhances the structural stability of GBA1 variants, while promoter choice significantly influences their expression levels in 293FT cells. In particular, GCase-7 with seven missense mutations in the signal peptide under the hEF1a promoter, exhibits the highest expression levels and catalytic activity among the variants studied. Furthermore, our conservation and coevolution analysis define key residues and interactions relevant for GCase function and stability. Importantly, our 3D structural analyses reveal minimal conformation changes between GCase-OPT and GCase-7 suggesting that their optimized catalytic activities can be attributed to the missense mutations, particularly in the signal peptide.

4 Discussion

Enzyme replacement therapy (ERT) is a lifesaving therapy for symptomatic GD-type 1 patients. It is not a cure, and patients must be treated for life [2]. Although, ERT is currently the standard therapy of GD due to its widespread used and established efficacy, it presents challenges such as suboptimal response for some patients and high cost

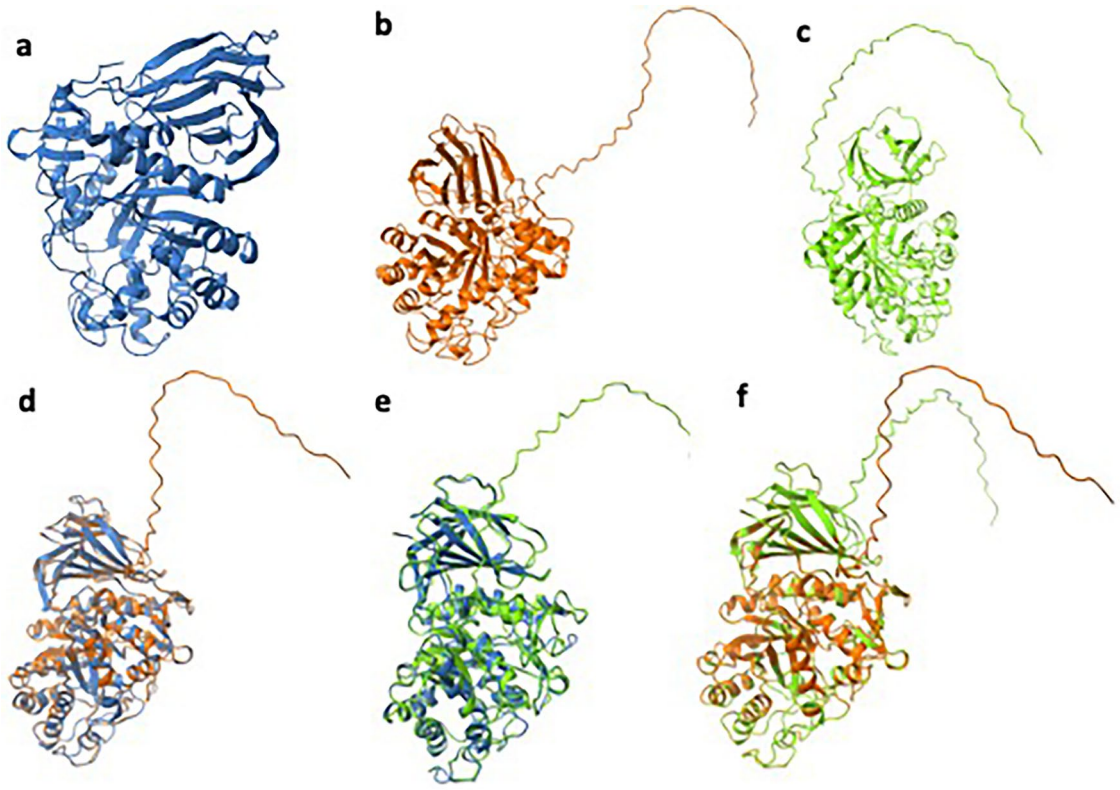


Fig. 9 Comparison of structures among the synthetic and recombinant GCase-Opt, GCase-7 and human GCase. **a**-**c** Three three-dimensional protein structures. **a** PDB: 2F61; **b** GCase-Opt and c. GCase-7. **d**-**e** Alignment of the three-dimensional protein structures with Matchmaker. **d** PDB: 2F61 x GCase-Opt; **e** PDB: 2F61 x GCase-7; **f** GCase-Opt x GCase-7

(e.g. $115,909 annually with imiglucerase and $80,941 with taliglucerase alfa per patient) [78]. Alternative treatment modalities such as substrate reduction therapy (SRT), hematopoietic stem cell transplantation (HSCT), and gene therapy (GT) have not proven to be as effective, since they often fail to address the full spectrum of disease manifestations [79–81]. Given these challenges, the development of novel functional synthetic and recombinant enzyme combining in silico molecular evolution, synthetic biology and gene therapy approaches provides a promising strategy to overcome the shortcomings of current therapeutic approaches for Gaucher disease.

To date, only one recombinant therapeutic enzyme produced in human cell line is commercially available [33]. In this study, we designed ten lentiviral vector plasmids based on a multisite gateway system for the efficient delivery of β-glucocerebrosidase to a human cell line for transient gene expression. These lentiviral vector plasmids were engineered to express GBA variants under the control of either the CMV or the hEF1a promoter. Specifically, we developed five lentiviral for each promoter. The GBA1 variants, which we denoted by GBA-7, GBA-8, GBA-9 and GBA-12 and the control GBA-Opt, were selected. Evolutionary criteria were utilized to introduce specific mutations into the signal peptides of these variants, derived from four selected mammalian species.

Our results show that the GBA-7 mRNAs transcribed from hEF1a promoter in 293FT cells are 2.62-fold higher compared to those from the CMV promoter ($p = 0.0003$). Furthermore, we observed stronger interactions of the transcriptional machinery with the hEF1a promoter for GBA-Opt, GBA-9, and GBA-12 mRNAs, which were expressed at 1.89 ($p = 0.001$), 3.0 ($p = 0.0004$) and 1.53 ($p = 0.017$) fold higher, respectively, compared to mRNAs from the CMV promoter. The hEF1a promoter, derived from the human EEF1A1 gene is known for its robust activity across various mammalian cell lines [82]. Several variants of the hEF1a promoters are available and exhibit differences in activity [83]. In our DNA constructs, the hEF1a promoter has the Sp1 binding site (G/TGGGCGGG/AG/AC/T) repeated five times and a single copy of the Ap-1 binding site (C/GTGACTC/AA) after the "TATA" box, as previously described [84]. Furthermore, the hEF1a promoter was also utilized in several clinical trials including X-linked severe combined immunodeficiency (SCID-X1) and adenosine deaminase severe combined immunodeficiency (ADA-SCID) as previously reported [85].

Based on the difference in GBA1 transcription levels, we proceeded to evaluate transfection efficiencies of our variants by quantifying the percentages of $EGFP^+$ cells 48 h post-delivery of both GFP and GBA1 DNA lentiviral plasmids. Interestingly, our results revealed that the hEF1a promoter significantly increased the expression levels of GBA-7 and GBA-Opt transcripts compared to the CMV promoter, with no significant difference observed in transfection efficiencies (see Table S2). On the other hand, we observed significant differences in transfection efficiencies for GBA-8, GBA-9, and GBA-12, with higher transfection efficiencies under the hEF1a promoter compared to the CMV promoter. These findings are consistent with the studies by Xu and colleagues [86] and Dou and colleagues [87], indicating the superior transfection efficiency of the hEF1a promoter across different cellular contexts.

Despite the high level of GBA1 expression from the different synthetic and recombinant GBA1 variants, we consistently obtained enzyme production only in four transgenic cells: 293FT_hEF1a_GCase-7, 293FT_CMV_GCase-7, 293FT_hEF1a_GCase-Opt and 293FT_CMV_GCase-Opt cells. Of note, our investigation showed that 293FT_hEF1a_GCase-7 cells exhibited the highest production level of the enzyme about 507.6 (± 38.17) nmol hydrolyzed substrate/mg-protein/h. In our expression system, we were able to produce levels of synthetic and recombinant GCase that were 6.1-fold higher than those produced by 293FT_Mock cells and 43.8 to 58.4-fold higher than those of leukocyte lysate [64]. This observed increase in enzymatic activity may have an impact for enzyme replacement therapy. Furthermore, structural models generated using the AlphaFold2 program revealed structural similarities between GCase-7 with 2F61 in the PDB, in agreement with the fact that it retains catalytic activity. For the other GCase variants, our results suggest that missense mutations in some amino acid residues of the signal peptide of GCase impair the catalytic activity. It is known that GCase differs from other lysosomal hydrolases in the mechanism underlying sorting and transport to lysosomes [88]. Following folding, the newly formed GCase molecules in the endoplasmic reticulum (ER) bind to the membrane protein LIMP2 (lysosomal membrane protein 2) [17]. This binding is mediated by hydrophobic helical interfaces on both proteins [17]. On the other hand, the full-dependence on LIMP-2 for trafficking of GCase to the lysosome was observed in fibroblast, but not in macrophages [20]. Subsequently, together with associated proteins, the GCase is transferred through the endosome and Golgi apparatus into the acidic environment of lysosome [24]. It is also known that in the lysosome, the activity of GCase depends on Saposin C (SAPC), a 9 kDa lysosomal membrane-interacting protein. The activation mechanism of GCase by SAPC has been characterized at atomistic-level, identifying the GCase residues that interact with SAPC [89]. Therefore, it remains to be determined, whether the missense mutations present at SP of GCase-8, GCas-9 and GCase-12 alter properties of these GCase variants that affect folding and trafficking to ER or lysosome.

Our investigation into the frequencies of common pathogenic variants identified a subset of mutations that were well conserved, while others showed lower conservation rates. Additionally, our analysis identified coevolved amino acids within the GCase enzyme family, highlighting important interaction patterns essential for protein stability and function. Collectively, these results suggest that not only mutations observed in Gaucher disease patients are usually related to the disruption of conservation, coevolution or structural patterns, but also that the modification of residues to those that actually follow these patterns can result in proteins with favorable characteristics, as was recently shown in a study by Do and colleagues [45]. For this reason, we have included a supplemental table that delineates amino acid frequencies (see Table S3), with a focus on different critical residues: pathogenic variants, mutations associated with Parkinson' disease risk and Dementia with Lewy bodies risk, catalytic site residues, as well as residues that bind to LIMP-2 and interact with Saposin C. Supplemental material with tables containing positional conservation and coevolution sets are also provided (see Tables S4 to S12).

Human cultured-cell based production is an attractive alternative manufacturing platform for biopharmaceuticals. The HEK293 cell line has been widely used for large-scale therapeutic protein production due to its high transfectability, growth to high cell densities, and shows glycosylation patterns compatible with human [90, 91]. Furthermore, since 2015, the FDA has approved therapeutic proteins produced in HEK293, and transient gene expression methods for producing recombinant proteins from HEK293 cells have been published [92–94]. With the advancement of synthetic biology, many synthetic toolboxes have been created to optimize cellular systems and, eventually, enhance the production of recombinant proteins [95–97]. In summary, in this study we combine in silico molecular evolution, synthetic biology and gene therapy approaches and engineered a functional synthetic and recombinant enzyme with seven missense mutations in the SP, named GCase-7. Further confirmation of GCase-7 activities using either patient-derived GBA mutant fibroblast cells and animal models are needed to demonstrate the potential of this combined strategy for improving therapeutic properties of lysosomal enzymes for enzyme replacement therapy for Gaucher disease.

5 Conclusion

To summarize, in our study, we provided a model to generate a new recombinant and synthetic human glucocerebrosidase. This model involved incorporating specific missense mutations at the signal peptide based on in silico evolution analysis. We choose signal peptide to reduce immunogenicity, as the signal peptide is cleaved during protein maturation. The resulting cDNAs were generated after codon optimization and tested for functionality in transient transfected 293FT cells. We demonstrated that GCase-7, which has 7 missense mutations at the signal peptide, exhibited higher specific activity under hEF1a promotor compared with GCase-Opt, which retained the wild-type signal peptide. This approach has not received much attention in the literature and our results suggest this approach is promising and warrants future studies. For example, animal models, should also be carried out to fully understand the possibilities and limitations of the approach that we suggest. A further limitation is that we did not assess the recombinant enzyme's functionality in fibroblasts derived from Gaucher patients, which would provide a more direct validation of its therapeutic potential in relevant cell models. This will be under consideration for future investigation. It is worth mentioning that our approach rests on the potential of the signal peptide to increase activity without inducing immunogenicity. This aspect has not been widely exploited in existing commercially available ERT products.

Acknowledgements We thank Patrícia Vianna Bonini Palma, Camila Menezes Bonaldo, Profa. Dra. Elza Tiemi Sakamoto Hojo, Daniela Perna, and members of our group Elisabete Maria S. Barreto-Beira, Rosana Márcia Silva Cruz and Dra. Claudia E. Vieira Wiezel for their constructive comments on this work. The graphical abstract was created with BioRender.com

Author contributions Conceptualization: L.L.S.F.; W.L.J.; V.W.S.G.; K.J.A.; F.L.; L.B.; G.B.; V.S.; R.W. and A.M.F. Data curation: L.L.S.F.; W.L.J.; K.J.A.; F.L.; L.B.; G.B. and A.M.F. Formal analysis: K.J.A.; F.L.; L.B. and A.M.F. Funding acquisition: V.S; R.W. and A.M.F. Investigation: L.L.S.F.; L.B.; G.B. and A.M.F. Methodology: L.L.S.F.; W.L.J.; M.D.O.; L.E.B.S.; V.D.A.; K.J.A.; F.L.; L.B.; G.B. and A.M.F. Project administration and Supervision: V.S.; R.W. and A.M.F. Validation: L.L.S.F.; G.B. and A.M.F.. Writing-original draft: L.L.S.F.; E.S.R.; L.B.; G.B.; K.J.A.; F.L. and A.M.F. Writing – review & editing: L.L.S.F.; W.L.J.; E.S.R.; V.D.A.; L.E.B.S.; M.D.O; V.A.; R.G.C.; V.S.; R.W.; S.G. and A.M.F.

Funding This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Agilent Technologies (Project no 2013/50450-2), FAPESP-MIT (Project no 2013/50764-7), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Apoio à Universidade de São Paulo (FUSP) - Project no 3541.

Data availability Data is provided as a supplementary material.

Declarations

Competing interests The authors declare no competing interests.

References

1. Platt FM, Boland B, van der Spoel AC. The cell biology of disease: lysosomal storage disorders: the cellular impact of lysosomal dysfunction. J Cell Biol. 2012;199:723–34. https://doi.org/10.1083/jcb.201208152.
2. Pastores GM. Lysosomal storage disorders: clinical and therapeutic aspects. Handb Clin Neurol. 2023;196:557–67. https://doi.org/10.1016/B978-0-323-98817-9.00006-5.
3. Futerman AH, van Meer G. The cell biology of lysosomal storage disorders. Nat Rev Mol Cell Biol. 2004;5:554–65. https://doi.org/10.1038/nrm1423.
4. Sugiman-Marangos SN, Beilhartz GL, Zhao X, Zhou D, Hua R, Kim PK, Rini JM, Minassian BA, Melnyk RA. Exploiting the diphtheria toxin internalization receptor enhances delivery of proteins to lysosomes for enzyme replacement therapy. Sci Adv. 2020. https://doi.org/10.1126/sciadv.abb0385.

5. Hruska KS, LaMarca ME, Scott CR, Sidransky E. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). Hum Mutat. 2008;29:567–83. https://doi.org/10.1002/humu.20676.
6. Brady RO, Kanfer JN, Bradley RM, Shapiro D. Demonstration of a deficiency of glucocerebroside-cleaving enzyme in Gaucher's disease. J Clin Invest. 1966;45:1112–5. https://doi.org/10.1172/JCI105417.
7. Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E. The human glucocerebrosidase gene and pseudogene: structure and evolution. Genomics. 1989;4:87–96. https://doi.org/10.1016/0888-7543(89)90319-4.
8. Ginns EI, Choudary PV, Tsuji S, Martin B, Stubblefield B, Sawyer J, Hozier J, Barranger JA. Gene mapping and leader polypeptide sequence of human glucocerebrosidase: implications for Gaucher disease. Proc Natl Acad Sci U S A. 1985;82:7101–5. https://doi.org/10.1073/pnas.82.20.7101.
9. Miyoshi K, Hagita H, Horiguchi T, Tanimura A, Noma T. Redefining GBA gene structure unveils the ability of cap-independent. IRES-Dependent Gene Regulation Commun Biol. 2022;5:639. https://doi.org/10.1038/s42003-022-03577-5.
10. Sorge JA, West C, Kuhl W, Treger L, Beutler E. The human glucocerebrosidase gene has two functional ATG initiator codons. Am J Hum Genet. 1987;41:1016–24.
11. Erickson AH, Ginns EI, Barranger JA. Biosynthesis of the lysosomal enzyme glucocerebrosidase. J Biol Chem. 1985;260:14319–24.
12. Lauffer L, Garcia PD, Harkins RN, Coussens L, Ullrich A, Walter P. Topology of signal recognition particle receptor in endoplasmic reticulum membrane. Nature. 1985;318:334–8. https://doi.org/10.1038/318334a0.
13. Lutcke H. Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. Eur J Biochem. 1995;228:531–50. https://doi.org/10.1111/j.1432-1033.1995.tb20293.x.
14. Reczek D, Schwake M, Schroder J, Hughes H, Blanz J, Jin X, Brondyk W, Van Patten S, Edmunds T, Saftig P. LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of beta-glucocerebrosidase. Cell. 2007;131:770–83. https://doi.org/10.1016/j.cell.2007.10.018.
15. Zachos C, Blanz J, Saftig P, Schwake M. A critical histidine residue within LIMP-2 mediates pH sensitive binding to its ligand beta-glucocerebrosidase. Traffic. 2012;13:1113–23. https://doi.org/10.1111/j.1600-0854.2012.01372.x.
16. Pavan E, Peruzzo P, Cattarossi S, Bergamin N, Bordugo A, Sechi A, Scarpa M, Biasizzo J, Colucci F, Dardis A. Deficiency of Glucocerebrosidase activity beyond Gaucher Disease: PSAP and LIMP-2 dysfunctions. Int J Mol Sci. 2024. https://doi.org/10.3390/ijms25126615.
17. Zunke F, Andresen L, Wesseler S, Groth J, Arnold P, Rothaug M, Mazzulli JR, Krainc D, Blanz J, Saftig P, Schwake M. Characterization of the complex formed by beta-glucocerebrosidase and the lysosomal integral membrane protein type-2. Proc Natl Acad Sci U S A. 2016;113:3791–6. https://doi.org/10.1073/pnas.1514005113.
18. Dobert JP, Bub S, Machtel R, Januliene D, Steger L, Regensburger M, Wilfling S, Chen JX, Dejung M, Plotz S, Hehr U, Moeller A, Arnold P, Zunke F. Activation and purification of ss-Glucocerebrosidase by exploiting its transporter limp-2 - implications for novel treatment strategies in Gaucher's and Parkinson's Disease. Adv Sci (Weinh). 2024;11: e2401641. https://doi.org/10.1002/advs.202401641.
19. Grabowski GA, Goldblatt J, Dinur T, Kruse J, Svennerholm L, Gatt S, Desnick RJ. Genetic heterogeneity in Gaucher disease: physicokinetic and immunologic studies of the residual enzyme in cultured fibroblasts from non-neuronopathic and neuronopathic patients. Am J Med Genet. 1985;21:529–49. https://doi.org/10.1002/ajmg.1320210316.
20. Horowitz M, Braunstein H, Zimran A, Revel-Vilk S, Goker-Alpan O. Lysosomal functions and dysfunctions: molecular and cellular mechanisms underlying Gaucher disease and its association with Parkinson disease. Adv Drug Deliv Rev. 2022;187: 114402. https://doi.org/10.1016/j.addr.2022.114402.
21. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, Bar-Shira A, Berg D, Bras J, Brice A, Chen CM, Clark LN, Condroyer C, De Marco EV, Durr A, Eblan MJ, Fahn S, Farrer MJ, Fung HC, Gan-Or Z, Gasser T, Gershoni-Baruch R, Giladi N, Griffith A, Gurevich T, Januario C, Kropp P, Lang AE, Lee-Chen GJ, Lesage S, Marder K, Mata IF, Mirelman A, Mitsui J, Mizuta I, Nicoletti G, Oliveira C, Ottman R, Orr-Urtreger A, Pereira LV, Quattrone A, Rogaeva E, Rolfs A, Rosenbaum H, Rozenberg R, Samii A, Samaddar T, Schulte C, Sharma M, Singleton A, Spitz M, Tan EK, Tayebi N, Toda T, Troiano AR, Tsuji S, Wittstock M, Wolfsberg TG, Wu YR, Zabetian CP, Zhao Y, Ziegler SG. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med. 2009;361:1651–61. https://doi.org/10.1056/NEJMoa0901281.
22. Ran C, Brodin L, Forsgren L, Westerlund M, Ramezani M, Gellhaar S, Xiang F, Fardell C, Nissbrandt H, Soderkvist P, Puschmann A, Ygland E, Olson L, Willows T, Johansson A, Sydow O, Wirdefeldt K, Galter D, Svenningsson P, Belin AC. Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. Neurobiol Aging. 2016. https://doi.org/10.1016/j.neurobiolaging.2016.04.022.
23. Gaubert S, Hourregue C, Mouton-Liger F, Millot P, Franco M, Amar-Bouaziz E, Aarsland D, Hugon J, Paquet C. Exploring the link between GBA1 mutations and Dementia with Lewy bodies. A mini-review Neurosci Biobehav Rev. 2022;141: 104856. https://doi.org/10.1016/j.neubiorev.2022.104856.
24. Granek Z, Barczuk J, Siwecka N, Rozpedek-Kaminska W, Kucharska E, Majsterek I. GBA1 Gene Mutations in alpha-synucleinopathies-molecular mechanisms underlying pathology and their clinical significance. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms24032044.
25. Daykin EC, Ryan E, Sidransky E. Diagnosing neuronopathic Gaucher disease: New considerations and challenges in assigning Gaucher phenotypes. Mol Genet Metab. 2021;132:49–58. https://doi.org/10.1016/j.ymgme.2021.01.002.
26. Gary SE, Ryan E, Steward AM, Sidransky E. Recent advances in the diagnosis and management of Gaucher disease. Expert Rev Endocrinol Metab. 2018;13:107–18. https://doi.org/10.1080/17446651.2018.1445524.
27. Chaves RG, Pereira Lda V, de Araujo FT, Rozenberg R, Carvalho MD, Coelho JC, Michelin-Tirelli K, Chaves Mde F, Cavalcanti GB Jr. Consanguinity and founder effect for Gaucher disease mutation G377S in a population from Tabuleiro do Norte, Northeastern Brazil. Clin Genet. 2015;88:391–5. https://doi.org/10.1111/cge.12515.
28. Grabowski GA, Mistry PK. Therapies for lysosomal storage diseases: Principles, practice, and prospects for refinements based on evolving science. Mol Genet Metab. 2022;137:81–91. https://doi.org/10.1016/j.ymgme.2022.07.014.
29. Zimran A, Elstein D, Levy-Lahad E, Zevin S, Hadas-Halpern I, Bar-Ziv Y, Foldes J, Schwartz AJ, Abrahamov A. Replacement therapy with imiglucerase for type 1 Gaucher's disease. Lancet. 1995;345:1479–80. https://doi.org/10.1016/s0140-6736(95)91038-7.
30. Cappellini MD, Carubbi F, Di Rocco M, Giona F, Giuffrida G. Long-term bone outcomes in Italian patients with Gaucher disease type 1 or type 3 treated with imiglucerase: a sub-study from the international collaborative Gaucher group (ICGG) Gaucher registry. Blood Cells Mol Dis. 2023;98: 102705. https://doi.org/10.1016/j.bcmd.2022.102705.

31. El-Beshlawy A, Tylki-Szymanska A, Vellodi A, Belmatoug N, Grabowski GA, Kolodny EH, Batista JL, Cox GF, Mistry PK. Long-term hematological, visceral, and growth outcomes in children with Gaucher disease type 3 treated with imiglucerase in the International collaborative Gaucher Group Gaucher Registry. Mol Genet Metab. 2017;120:47–56. https://doi.org/10.1016/j.ymgme.2016.12.001.
32. Weinreb NJ, Camelo JS Jr, Charrow J, McClain MR, Mistry P, Belmatoug N. Gaucher disease type 1 patients from the ICGG Gaucher registry sustain initial clinical improvements during twenty years of imiglucerase treatment. Mol Genet Metab. 2021;132:100–11. https://doi.org/10.1016/j.ymgme.2020.12.295.
33. Zimran A, Altarescu G, Philips M, Attias D, Jmoudiak M, Deeb M, Wang N, Bhirangi K, Cohn GM, Elstein D. Phase 1/2 and extension study of velaglucerase alfa replacement therapy in adults with type 1 Gaucher disease: 48-month experience. Blood. 2010;115:4651–6. https://doi.org/10.1182/blood-2010-02-268649.
34. Becker-Cohen M, Revel-Vilk S, Frydman D, Dinur T, Tiomkin M, Istaiti M, Arbel N, Bauer P, Cozma C, Rolfs A, Szer J, Zimran A. Rapid home therapy infusion of velaglucerase alfa in naive patients with Gaucher disease. Intern Med J. 2023. https://doi.org/10.1111/imj.16179.
35. Aviezer D, Brill-Almon E, Shaaltiel Y, Hashmueli S, Bartfeld D, Mizrachi S, Liberman Y, Freeman A, Zimran A, Galun E. A plant-derived recombinant human glucocerebrosidase enzyme–a preclinical and phase I investigation. PLoS ONE. 2009;4: e4792. https://doi.org/10.1371/journal.pone.0004792.
36. Velmishi V, Troja E, Tanka M, Bali D, Dervishi E, Tako A, Kollcaku L, Cullufi P. Skeletal manifestations, bone pain, and BMD changes in albanian type 1 Gaucher patients treated with Taliglucerase Alfa. J Osteoporos. 2023;2023:3254533. https://doi.org/10.1155/2023/3254533.
37. Kim EY, Hong YB, Lai Z, Kim HJ, Cho YH, Brady RO, Jung SC. Expression and secretion of human glucocerebrosidase mediated by recombinant lentivirus vectors in vitro and in vivo: implications for gene therapy of Gaucher disease. Biochem Biophys Res Commun. 2004;318:381–90. https://doi.org/10.1016/j.bbrc.2004.04.040.
38. Liu C, Bahnson AB, Dunigan JT, Watkins SC, Barranger JA. Long-term expression and secretion of human glucocerebrosidase by primary murine and human myoblasts and differentiated myotubes. J Mol Med (Berl). 1998;76:773–81. https://doi.org/10.1007/s001090050279.
39. Novo JB, Morganti L, Moro AM, Paes Leme AF, Serrano SM, Raw I, Ho PL. Generation of a Chinese hamster ovary cell line producing recombinant human glucocerebrosidase. J Biomed Biotechnol. 2012;2012: 875383. https://doi.org/10.1155/2012/875383.
40. Gramlich PA, Westbroek W, Feldman RA, Awad O, Mello N, Remington MP, Sun Y, Zhang W, Sidransky E, Betenbaugh MJ, Fishman PS. A peptide-linked recombinant glucocerebrosidase for targeted neuronal delivery: design, production, and assessment. J Biotechnol. 2016;221:1–12. https://doi.org/10.1016/j.jbiotec.2016.01.015.
41. Naphatsamon U, Ohashi T, Misaki R, Fujiyama K. The production of human beta-glucocerebrosidase in nicotiana benthamiana root culture. Int J Mol Sci. 2018. https://doi.org/10.3390/ijms19071972.
42. Uthailak N, Kajiura H, Misaki R, Fujiyama K. Production of recombinant beta-glucocerebrosidase in wild-type and glycoengineered transgenic Nicotiana benthamiana root cultures with different N-glycan profiles. J Biosci Bioeng. 2022;133:481–8. https://doi.org/10.1016/j.jbiosc.2022.01.002.
43. Do MA, Levy D, Brown A, Marriott G, Lu B. Targeted delivery of lysosomal enzymes to the endocytic compartment in human cells using engineered extracellular vesicles. Sci Rep. 2019;9:17274. https://doi.org/10.1038/s41598-019-53844-5.
44. Du S, Ou H, Cui R, Jiang N, Zhang M, Li X, Ma J, Zhang J, Ma D. Delivery of Glucosylceramidase beta gene using AAV9 vector therapy as a treatment strategy in mouse models of gaucher disease. Hum Gene Ther. 2019;30:155–67. https://doi.org/10.1089/hum.2018.072.
45. Pokorna S, Khersonsky O, Lipsh-Sokolik R, Goldenzweig A, Nielsen R, Ashani Y, Peleg Y, Unger T, Albeck S, Dym O, Tirosh A, Tarayra R, Hocquemiller M, Laufer R, Ben-Dor S, Silman I, Sussman JL, Fleishman SJ, Futerman AH. Design of a stable human acid-beta-glucosidase: towards improved Gaucher disease therapy and mutation classification. FEBS J. 2023;290:3383–99. https://doi.org/10.1111/febs.16758.
46. Davis BM, Encell LP, Zielske SP, Christians FC, Liu L, Friebert SE, Loeb LA, Gerson SL. Applied molecular evolution of O6-benzylguanine-resistant DNA alkyltransferases in human hematopoietic cells. Proc Natl Acad Sci U S A. 2001;98:4950–4. https://doi.org/10.1073/pnas.091601198.
47. Fontes AM, Davis BM, Encell LP, Lingas K, Covas DT, Zago MA, Loeb LA, Pegg AE, Gerson SL. Differential competitive resistance to methylating versus chloroethylating agents among five O6-alkylguanine DNA alkyltransferases in human hematopoietic cells. Mol Cancer Ther. 2006;5:121–8. https://doi.org/10.1158/1535-7163.MCT-05-0236.
48. Gerson SL. Clinical relevance of MGMT in the treatment of cancer. J Clin Oncol. 2002;20:2388–99. https://doi.org/10.1200/JCO.2002.06.110.
49. Chung J, Scherer LJ, Gu A, Gardner AM, Torres-Coronado M, Epps EW, Digiusto DL, Rossi JJ. Optimized lentiviral vectors for HIV gene therapy: multiplexed expression of small RNAs and inclusion of MGMT(P140K) drug resistance gene. Mol Ther. 2014;22:952–63. https://doi.org/10.1038/mt.2014.32.
50. Li C, Georgakopoulou A, Newby GA, Chen PJ, Everette KA, Paschoudi K, Vlachaki E, Gil S, Anderson AK, Koob T, Huang L, Wang H, Kiem HP, Liu DR, Yannaki E, Lieber A. In vivo HSC prime editing rescues sickle cell disease in a mouse model. Blood. 2023;141:2085–99. https://doi.org/10.1182/blood.2022018252.
51. Hendrikse NM, Sandegren A, Andersson T, Blomqvist J, Makower A, Possner D, Su C, Thalen N, Tjernberg A, Westermark U, Rockberg J, Svensson Gelius S, Syren PO, Nordling E. Ancestral lysosomal enzymes with increased activity harbor therapeutic potential for treatment of Hunter syndrome. iScience. 2021;24:102154. https://doi.org/10.1016/j.isci.2021.102154.
52. Hallows WC, Skvorak K, Agard N, Kruse N, Zhang X, Zhu Y, Botham RC, Chng C, Shukla C, Lao J, Miller M, Sero A, Viduya J, Ismaili MHA, McCluskie K, Schiffmann R, Silverman AP, Shen JS, Huisman GW. Optimizing human alpha-galactosidase for treatment of Fabry disease. Sci Rep. 2023;13:4748. https://doi.org/10.1038/s41598-023-31777-4.
53. Kitada T, DiAndreth B, Teague B, Weiss R. Programming gene and engineered-cell therapies with synthetic biology. Science. 2018. https://doi.org/10.1126/science.aad1067.
54. Ilia K, Shakiba N, Bingham T, Jones RD, Kaminski MM, Aravera E, Bruno S, Palacios S, Weiss R, Collins JJ, Del Vecchio D, Schlaeger TM. Synthetic genetic circuits to uncover the OCT4 trajectories of successful reprogramming of human fibroblasts. Sci Adv. 2023. https://doi.org/10.1126/sciadv.adg8495.
55. Birch SM, Lawlor MW, Conlon TJ, Guo LJ, Crudele JM, Hawkins EC, Nghiem PP, Ahn M, Meng H, Beatka MJ, Fickau BA, Prieto JC, Styner MA, Struharik MJ, Shanks C, Brown KJ, Golebiowski D, Bettis AK, Balog-Alvarez CJ, Clement N, Coleman KE, Corti M, Pan X, Hauschka SD, Gonzalez JP, Morris CA, Schneider JS, Duan D, Chamberlain JS, Byrne BJ, Kornegay JN. Assessment of systemic AAV-microdystrophin

gene therapy in the GRMD model of duchenne muscular dystrophy. Sci Transl Med. 2023. https://doi.org/10.1126/scitranslmed.abo1815.
56. Sekayan T, Simmons DH, von Drygalski A. Etranacogene dezaparvovec-drlb gene therapy for patients with hemophilia B (congenital factor IX deficiency). Expert Opin Biol Ther. 2023;23:1173–84. https://doi.org/10.1080/14712598.2023.2282138.
57. Liang Q, Catalano F, Vlaar EC, Pijnenburg JM, Stok M, van Helsdingen Y, Vulto AG, van der Ploeg AT, van Til NP, Pijnappel W. IGF2-tagging of GAA promotes full correction of murine Pompe disease at a clinically relevant dosage of lentiviral gene therapy. Mol Ther Methods Clin Dev. 2022;27:109–30. https://doi.org/10.1016/j.omtm.2022.09.010.
58. BethAnn F, Hayes M. Enhanced in vivo uptake of glucocerebrosidase. United States Patent. 1996;5(349):1–10.
59. Sasaki Y, Sone T, Yoshida S, Yahata K, Hotta J, Chesnut JD, Honda T, Imamoto F. Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the multisite gateway system. J Biotechnol. 2004;107:233–43. https://doi.org/10.1016/j.jbiotec.2003.10.001.
60. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics. 2007;23:1073–9. https://doi.org/10.1093/bioinformatics/btm076.
61. Lorenz R, Bernhart SH, Honer ZU, Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA package 2.0. Algorithms Mol Biol. 2011;6:26. https://doi.org/10.1186/1748-7188-6-26.
62. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134. https://doi.org/10.1186/1471-2105-13-134.
63. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29: e45. https://doi.org/10.1093/nar/29.9.e45.
64. Peters SP, Coyle P, Glew RH. Differentiation of beta-glucocerebrosidase from beta-glucosidase in human tissues using sodium taurocholate. Arch Biochem Biophys. 1976;175:569–82. https://doi.org/10.1016/0003-9861(76)90547-6.
65. Muller KB, Rodrigues MD, Pereira VG, Martins AM, D'Almeida V. Reference values for lysosomal enzymes activities using dried blood spots samples - a Brazilian experience. Diagn Pathol. 2010;5:65. https://doi.org/10.1186/1746-1596-5-65.
66. Zdobnov EM, Kuznetsov D, Tegenfeldt F, Manni M, Berkeley M, Kriventseva EV. OrthoDB in 2020: evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2021;49:D389–93. https://doi.org/10.1093/nar/gkaa1009.
67. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. https://doi.org/10.1093/nar/gkh340.
68. Fonseca-Junior NJ, Afonso MQL, Oliveira LC, Bleicher L. PFstats: a network-based open tool for protein family analysis. J Comput Biol. 2018;25:480–6. https://doi.org/10.1089/cmb.2017.0181.
69. Sheth J, Bhavsar R, Mistri M, Pancholi D, Bavdekar A, Dalal A, Ranganath P, Girisha KM, Shukla A, Phadke S, Puri R, Panigrahi I, Kaur A, Muranjan M, Goyal M, Ramadevi R, Shah R, Nampoothiri S, Danda S, Datar C, Kapoor S, Bhatwadekar S, Sheth F. Gaucher disease: single gene molecular characterization of one-hundred Indian patients reveals novel variants and the most prevalent mutation. BMC Med Genet. 2019;20:31. https://doi.org/10.1186/s12881-019-0759-1.
70. Horowitz M, Zimran A. Mutations causing Gaucher disease. Hum Mutat. 1994;3:1–11. https://doi.org/10.1002/humu.1380030102.
71. Rozenberg R, Araujo FT, Fox DC, Aranda P, Nonino A, Micheletti C, Martins AM, Cravo R, Sobreira E, Pereira LV. High frequency of mutation G377S in Brazilian type 3 Gaucher disease patients. Braz J Med Biol Res. 2006;39:1171–9. https://doi.org/10.1590/s0100-879x2006000900004.
72. Dima RI, Thirumalai D. Determination of network of residues that regulate allostery in protein families using sequence analysis. Protein Sci. 2006;15:258–68. https://doi.org/10.1110/ps.051767306.
73. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. UCSF ChimeraX: tools for structure building and analysis. Protein Sci. 2023;32: e4792. https://doi.org/10.1002/pro.4792.
74. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.
75. Liou B, Kazimierczuk A, Zhang M, Scott CR, Hegde RS, Grabowski GA. Analyses of variant acid beta-glucosidases: effects of Gaucher disease mutations. J Biol Chem. 2006;281:4242–53. https://doi.org/10.1074/jbc.M511110200.
76. MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. J Mol Biol. 1991;218:397–412. https://doi.org/10.1016/0022-2836(91)90721-h.
77. Bleicher L, Prates ET, Gomes TC, Silveira RL, Nascimento AS, Rojas AL, Golubev A, Martinez L, Skaf MS, Polikarpov I. Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from Rhodothermus marinus and molecular dynamics simulations. J Phys Chem B. 2011;115:7940–9. https://doi.org/10.1021/jp200330z.
78. Farahbakhshian S, Inocencio TJ, Poorman G, Wright E, Pathak RR, Bullano M. The budget impact of enzyme replacement therapy in type 1 Gaucher disease in the United States. J Med Econ. 2022;25:755–61. https://doi.org/10.1080/13696998.2022.2082200.
79. Cox TM, Charrow J, Lukina E, Mistry PK, Foster MC, Peterschmitt MJ. Long-term effects of eliglustat on skeletal manifestations in clinical trials of patients with Gaucher disease type 1. Genet Med. 2023;25: 100329. https://doi.org/10.1016/j.gim.2022.10.011.
80. Aboobacker FN, Kulkarni UP, Korula A, Devasia AJ, Selvarajan S, Lionel S, Sindhuvi E, Srivastava A, George B, Abraham A. Hematopoietic Stem Cell Transplantation is a cost-effective alternative to enzyme replacement therapy in Gaucher Disease. Blood Cell Ther. 2022;5:69–74. https://doi.org/10.31547/bct-2021-020.
81. Shaimardanova AA, Solovyeva VV, Issa SS, Rizvanov AA. Gene therapy of sphingolipid metabolic disorders. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms24043627.
82. Cabrera A, Edelstein HI, Glykofrydis F, Love KS, Palacios S, Tycko J, Zhang M, Lensch S, Shields CE, Livingston M, Weiss R, Zhao H, Haynes KA, Morsut L, Chen YY, Khalil AS, Wong WW, Collins JJ, Rosser SJ, Polizzi K, Elowitz MB, Fussenegger M, Hilton IB, Leonard JN, Bintu L, Galloway KE, Deans TL. The sound of silence: Transgene silencing in mammalian cell engineering. Cell Syst. 2022;13:950–73. https://doi.org/10.1016/j.cels.2022.11.005.

83. Zheng C, Baum BJ. All human EF1alpha promoters are not equal: markedly affect gene expression in constructs from different sources. Int J Med Sci. 2014;11:404–8. https://doi.org/10.7150/ijms.8033.
84. Uetsuki T, Naito A, Nagata S, Kaziro Y. Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha. J Biol Chem. 1989;264:5791–8.
85. Dahl M, Smith EMK, Warsi S, Rothe M, Ferraz MJ, Aerts J, Golipour A, Harper C, Pfeifer R, Pizzurro D, Schambach A, Mason C, Karlsson S. Correction of pathology in mice displaying Gaucher disease type 1 by a clinically-applicable lentiviral vector. Mol Ther Methods Clin Dev. 2021;20:312–23. https://doi.org/10.1016/j.omtm.2020.11.018.
86. Xu ZJ, Jia YL, Wang M, Yi DD, Zhang WL, Wang XY, Zhang JH. Effect of promoter, promoter mutation and enhancer on transgene expression mediated by episomal vectors in transfected HEK293, Chang liver and primary cells. Bioengineered. 2019;10:548–60. https://doi.org/10.1080/21655979.2019.1684863.
87. Dou Y, Lin Y, Wang TY, Wang XY, Jia YL, Zhao CP. The CAG promoter maintains high-level transgene expression in HEK293 cells. FEBS Open Bio. 2021;11:95–104. https://doi.org/10.1002/2211-5463.13029.
88. Boer DEC, van Smeden J, Bouwstra JA, Aerts J. Glucocerebrosidase: functions in and beyond the lysosome. J Clin Med. 2020. https://doi.org/10.3390/jcm9030736.
89. Romero R, Ramanathan A, Yuen T, Bhowmik D, Mathew M, Munshi LB, Javaid S, Bloch M, Lizneva D, Rahimova A, Khan A, Taneja C, Kim SM, Sun L, New MI, Haider S, Zaidi M. Mechanism of glucocerebrosidase activation and dysfunction in Gaucher disease unraveled by molecular dynamics and deep learning. Proc Natl Acad Sci U S A. 2019;116:5086–95. https://doi.org/10.1073/pnas.1818411116.
90. Gutierrez-Granados S, Cervera L, Kamen AA, Godia F. Advancements in mammalian cell transient gene expression (TGE) technology for accelerated production of biologics. Crit Rev Biotechnol. 2018;38:918–40. https://doi.org/10.1080/07388551.2017.1419459.
91. Hu J, Han J, Li H, Zhang X, Liu LL, Chen F, Zeng B. Human embryonic kidney 293 cells: a vehicle for biopharmaceutical manufacturing, structural biology, and electrophysiology. Cells Tissues Organs. 2018;205:1–8. https://doi.org/10.1159/000485501.
92. Tan E, Chin CSH, Lim ZFS, Ng SK. HEK293 cell line as a platform to produce recombinant proteins and viral vectors. Front Bioeng Biotechnol. 2021;9: 796991. https://doi.org/10.3389/fbioe.2021.796991.
93. Cervera L, Kamen AA. Large-scale transient transfection of suspension mammalian cells for VLP production. Methods Mol Biol. 2018;1674:117–27. https://doi.org/10.1007/978-1-4939-7312-5_10.
94. Hacker DL, Ordonez-Moran P. Large-scale production of recombinant noggin and r-spondin1 proteins required for the maintenance of stem cells in intestinal organoid cultures. Methods Mol Biol. 2020;2171:171–84. https://doi.org/10.1007/978-1-0716-0747-3_10.
95. Eisenhut P, Marx N, Borsi G, Papez M, Ruggeri C, Baumann M, Borth N. Manipulating gene expression levels in mammalian cell factories: An outline of synthetic molecular toolboxes to achieve multiplexed control. N Biotechnol. 2023;79:1–19. https://doi.org/10.1016/j.nbt.2023.11.003.
96. Di Blasi R, Pisani M, Tedeschi F, Marbiah MM, Polizzi K, Furini S, Siciliano V, Ceroni F. Resource-aware construct design in mammalian cells. Nat Commun. 2023;14:3576. https://doi.org/10.1038/s41467-023-39252-4.
97. Chen WCW, Gaidukov L, Lai Y, Wu MR, Cao J, Gutbrod MJ, Choi GCG, Utomo RP, Chen YC, Wroblewska L, Kellis M, Zhang L, Weiss R, Lu TK. A synthetic transcription platform for programmable gene expression in mammalian cells. Nat Commun. 2022;13:6167. https://doi.org/10.1038/s41467-022-33287-9.