Non-coding RNA Research 4 (2019) 1–14

Contents lists available at ScienceDirect

Non-coding RNA Research

journal homepage: http://www.keaipublishing.com/ncrna

Modifying inter-cistronic sequence significantly enhances IRES dependent second gene expression in bicistronic vector: Construction of optimised cassette for gene therapy of familial hypercholesterolemia

Faisal A. Al-Allaf[a,b,c,d,e,*,1], Zainularifeen Abduljaleel[a,b,**,1], Mohammad Athar[a,b], Mohiuddin M. Taher[a,b], Wajahatullah Khan[f], Huseyin Mehmet[e], Mukaddes Colakogullari[e], Sophia Apostolidou[e], Brian Bigger[d], Simon Waddington[d], Charles Coutelle[d], Michael Themis[d], Mohammed N. Al-Ahdal[g], Futwan A. Al-Mohanna[h], Zuhair N. Al-Hassnan[i], Abdellatif Bouazzaoui[a,b]

[a] Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P.O. Box 715, Makkah, 21955, Saudi Arabia
[b] Science and Technology Unit, Umm Al-Qura University, P.O. Box 715, Makkah, 21955, Saudi Arabia
[c] Molecular Diagnostics Unit, Department of Laboratory and Blood Bank, King Abdullah Medical City, Makkah, 21955, Saudi Arabia
[d] Gene Therapy Research Group, Department of Molecular and Cell Medicine, Faculty of Medicine, Imperial College London, South Kensington, London, SW7 2AZ, UK
[e] Institute of Reproductive and Developmental Biology, Division of Clinical Sciences, Faculty of Medicine, Imperial College London, London, W12 0NN, UK
[f] Department of Basic Sciences, College of Science and Health Professions, King Saud Bin Abdulaziz University for Health Sciences, PO Box 3124, Riyadh, 11426, Saudi Arabia
[g] Department of Infection and Immunity, King Faisal Specialist Hospital & Research Center, Riyadh, 11211, Saudi Arabia
[h] Department of Cell Biology, King Faisal Specialist Hospital and Research Center, Riyadh, 11211, Saudi Arabia
[i] Department of Medical Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, 11211, Saudi Arabia

ARTICLE INFO

Keywords:
Internal ribosome entry site
IRES
Gene therapy
Inter-cistronic sequences
MD simulation
Protein structure modeling
Familial hypercholesterolemia

ABSTRACT

Internal ribosome entry site (IRES) sequences have become a valuable tool in the construction of gene transfer and therapeutic vectors for multi-cistronic gene expression from a single mRNA transcript. The optimal conditions for effective use of this sequence to construct a functional expression vector are not precisely defined but it is generally assumed that the internal ribosome entry site dependent expression of the second gene in such as cassette is less efficient than the cap-dependent expression of the first gene. Mainly tailoring inter-cistronic sequence significantly enhances IRES dependent second gene expression in bicistronic vector further in construction of optimised cassette for gene therapy of familial hypercholesterolemia. We tailored the size of the inter-cistronic spacer sequence at the 5′ region of the internal ribosome entry site sequence using sequential deletions and demonstrated that the expression of the 3′ gene can be significantly increased to similar levels as the cap-dependent expression of the 5' gene. Maximum expression efficiency of the downstream gene was obtained when the spacer is composed of 18–141 base pairs. In this case a single mRNA transcriptional unit containing both the first and the second Cistron was detected. Whilst constructs with spacer sequences of 216 bp or longer generate a single transcriptional unit containing only the first Cistron. This suggests that long spacers may affect transcription termination. When the spacer is 188 bp, both transcripts were produced simultaneously in most transfected cells, while a fraction of them expressed only the first but not the second gene. Expression analyses of vectors containing optimised cassettes clearly confirm that efficiency of gene transfer and biological activity of the expressed transgenic proteins in the transduced cells can be achieved. Furthermore, Computational analysis was carried out by molecular dynamics (MD) simulation to determine the most emerges as viable containing specific binding site and bridging of 5′ and 3′ ends involving direct RNA-RNA contacts and RNA-protein interactions. These results provide a mechanistic basis for translation stimulation and RNA resembling for the synergistic stimulation of cap-dependent translation.

* Corresponding author. Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P. O. Box 18802, Makkah 21955, Saudi Arabia.
** Corresponding author. Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P. O. Box 18802, Makkah 21955, Saudi Arabia.
E-mail addresses: faallaf@uqu.edu.sa (F.A. Al-Allaf), zainulbio@gmail.com, zaabduljaleel@uqu.edu.sa (Z. Abduljaleel).
[1] Authors Contributed Equally to this Manuscript.

https://doi.org/10.1016/j.ncrna.2018.11.005
Received 26 September 2018; Received in revised form 21 November 2018; Accepted 21 November 2018
Available online 22 November 2018

1. Introduction

Simultaneous expression of multiple heterologous genes from a single gene transfer vector is an important requirement in a range of gene transfer and gene therapy protocols. There are generally four common strategies in which two or more genes can be co-expressed from a single vector. Firstly, two genes can be fused together in frame to produce a chimeric sequence ensuring simultaneous expression of both genes as a fusion protein [1–3]. However, this strategy may not work for all fusion proteins as some may be misfolded or mistargeted and, particularly important to gene therapy, the fusions might be immunogenic to the host. Secondly, separate promoters can be used to drive expression of different genes in the same vector. The major disadvantage of constructing such a dual promoter vector is possible transcriptional interference and/or dissociated gene expression, with a fraction of the transfected or transduced cells expressing only one gene such as the selectable marker and not the gene of interest and vice versa [4–6]. The third strategy relies on the utilisation of natural splicing signals of viruses by which multiple RNAs are produced from a single transcript [7]. However, this is not a frequently used strategy because of sub-optimal splicing of the viral vector. To overcome the above shortcomings, the fourth strategy involves the construction of a bicistronic

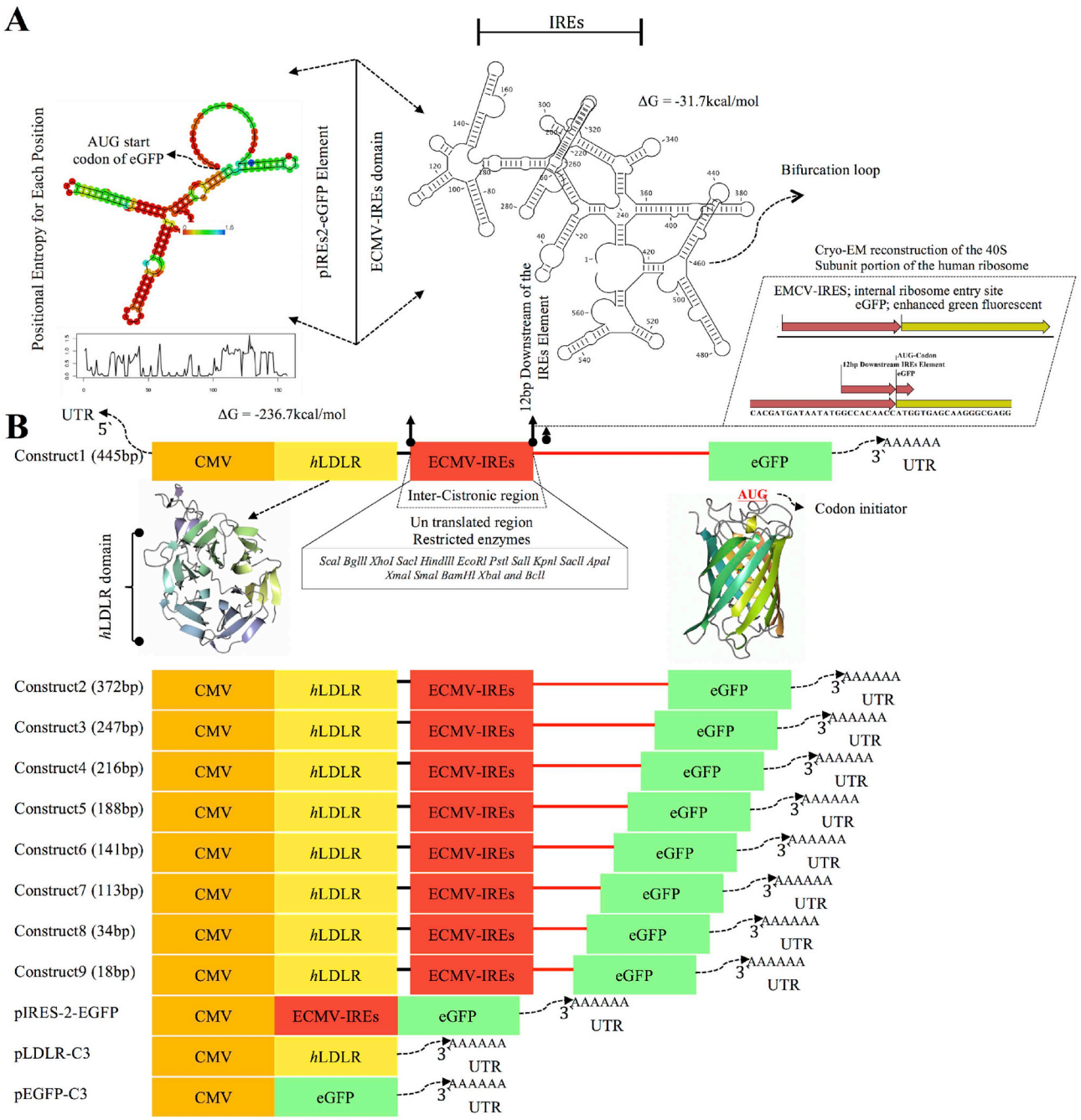


Fig. 1. Structures of the bicistronic cassettes. The sequence and minimum energy structure of the IRES (Emcv) highlight motifs of interest for bicistronic protein expression. RNA sequences near the IRES AUG differed for commercial and derived plasmids. Uppercase denotes native Emcv bases. The protein coding ORFs and all plasmids encode A6 in the JK bifurcation loop except pCITE-1, pIRES, and pF/R-att. A) Schematic diagram of the constructed 5′-LDLR-IRES-eGFP-3′ expression cassettes with sequential deletions at the 5′ of IRES element. Sequence annotation of spacer sequence in between IRES (Emcv)-eGFP. B) Inter-cistronic spacer sequence with different construct vectors.

cassette, in which the two heterologous genes are separated by an element known as the internal ribosome entry site (IRES) sequence. Transcription of both genes in the bicistronic cassette is driven from a common upstream promoter, thus eliminating promoter interference. As a result, a single messenger unit including the bicistronic transcript of both genes separated by IRES is produced to allow uncoupling of the two Cistrons during translation. Translation initiation of the first Cistron 5′ from the IRES is typically mediated by cap-dependent translation initiation [8], however translation of the second Cistron 3' from the IRES is mediated independently of capping by using the IRES element as an internal ribosome-binding site to initiate translation.

Since both the first and the second genes in a bicistronic cassette are under the control of the 5′ upstream promoter, detection of the protein encoded by the second Cistron is theoretically assured by the successful expression of the first Cistron. However, many researchers using IRES elements for construction of bicistronic gene transfer vectors have reported difficulties in achieving significant expression of the second Cistron. Studies conducted to determine the parameters that influence IRES-dependent translation initiation in bicistronic constructs have shown that the length and the secondary structure of non-coding inter-Cistronic sequences (ICS) that separate the 3′ end of the IRES element from the downstream Cistron have a strong effect on the efficiency of IRES function [9–11]. However, little is known about the ideal length of non-coding ICS that separates the 5′ end of the IRES element from the upstream Cistron and provides optimal expression both 5' and 3′ of the genes.

In the present study, we constructed nine expression vectors containing bicistronic cassettes under control of the CMV promoter. These bicistronic cassettes contain sequentially deleted non-coding ICS sequences, between the stop codon of the human low-density lipoprotein cDNA (hLDLR cDNA) and the IRES-eGFP cassette sequences. We then examined the efficiency of these vectors with the aim to screen and select for an optimal 5'-hLDLR-IRES-eGFP-3′ cassette, which can be used for preclinical gene therapy studies of familial hypercholesterolemia.

IRES direct recruitment of the initiation complex without requiring a 5-cap [12]. Generally present in the 5-UTR. IRES-dependent initiation may occur by direct binding of the 40S subunit to the mRNA, or may be mediated by a subset of the canonical initiation factors, and in some cases necessitates specific cellular proteins called IRES trans-acting factors (ITAFs) [13]. Furthermore, *in silico* analysis of RNA structure prediction, comparative sequence analysis and minimization of thermodynamic free energy equilibrium structure were applied with our IRES search system.

To better understand the structures and functions of IRES RNAs. We used comparative sequence analysis (CSA) to analysis and identity covering base pair. This approach was proven to be effective in the construction of reliable secondary structure models of ribosomal, signal recognition particle and transfer messenger RNAs [14]. Our studies yielded a revised secondary structure of IRES RNA that was supported by both covering base pairs and available biochemical data. The model was used to investigate IRES RNA in three dimensions in free form as well as when bound to IRES associated proteins and the 40s ribosomal subunit.

2. Results

2.1. Structure of the bicistronic cassettes

In view of the less dramatic effect of inter cistronic sequence (ICS) on IRES mediated translation initiation, to test the effect of ICS length on gene expression we constructed nine expression vectors containing bicistronic cassettes with variable length of ICS sequences (Fig. 1A and B). Because the order of the genes on the bicistronic cassette has a great influence on their expression levels [9] we placed the hLDLRcDNA sequence that we intend to use as a therapeutic gene for functional complementation of hLDLR deficiency, 5' to the IRES element. Sequential deletions were created using a modified ExoIII/S1 nuclease protocol [15] and the DNA fragments that were generated containing the hLDLRcDNA with different length of spacers were subcloned into the pIRES2-eGFP (Clontech) mammalian expression vector. The pIRES2-eGFP expression vector was chosen because it contains the IRES of the encephalomyocarditis virus (Emcv) which has a high translation efficiency compared to other IRES sequences including those from hepatitis A and C viruses, poliovirus, human rhinovirus, and foot-and-mouth disease virus [16–19]. In addition to its high efficiency, Emcv IRES posses a broad tissue tropism [20] which makes it the most widely, used in gene therapy and gene transfer protocols [21–23]. This vector also permits high levels of plasmid DNA production and its multi-cloning sites are helpful for gene cloning. The eGFP incorporated in this vector is a red-shifted variant [24,25] of wild type eGFP, which has been optimised for brighter fluorescence and higher expression in mammalian cells.

Moreover, sequences flanking the eGFP have been converted to carry a Kozak consensus translation initiation site [26] that increases further translation efficiency in eukaryotic cells. The AUG initiator codon of the eGFP is located 12 bases downstream of the IRES element, i.e., such a short distance is nevertheless sufficient to significantly affect the eGFP expression, regardless of the conservation of the optimal eGFP AUG context (acc AUG gug). Indeed, flow cytometry analyses have previously shown a more than 10 fold difference in eGFP peak intensity. The eGFP was used as a marker for clear and rapid detection for monitoring gene expression. As the fluorescence activity for eGFP requires IRES no substrates, cofactors or additional gene products the transfected cells can be sorted immediately by flow cytometry without the use of antibiotic for selection when such a construct is used for future in vivo studies.

2.2. Analyses of hLDLR and eGFP expression by western blotting and flow cytometry

Comparison of expression from the first and the second gene between cells transfected with bicistronic constructs containing variable length of spacer was performed by Western blotting and flow cytometry. Western analysis ensures that the expressed proteins are of the correct, mature sizes and the flow cytometry permits accurate assessment of gene transfer and expression at a single cell level. Immunoprobing with an anti-hLDLR antibody (Fig. 2A) shows the mature and immature forms of the hLDLR protein from cell extracts transfected with each of the nine constructs. The hLDLR migrates as two bands of molecular masses of 160 kDa and 120 kDa of which the 160 kDa mature band was more prominent in all samples. No signals were detected from the negative controls (lane 10, 12, and 13). As expected, the level of hLDLR protein from cells transfected with the mono-Cistronic pLDLR-C3 vector (lane 11) was higher than the amount of hLDLR protein produced from all other bicistronic cassettes. To generate quantitative data for comparative analysis of these blots, each signal on the blots was measured by densitometry and normalised against β-tubuline (Fig. 2B and C) using National Institutes of Health (NIH) software by ImageJ, a Java program [27]. No significant differences in hLDLR expression were detected from samples transfected with constructs containing different spacer sizes. Immunoprobing of the membrane with anti-eGFP antibody in Fig. 2B shows the production of eGFP protein of the expected molecular mass of approximately 30 kDa from cells transfected with clones containing ICS sequences ranging from 188bp to 18bp. Unlike hLDLR expression, the levels of eGFP expression increase with decreasing length of the spacer sequence, which is also confirmed by densitometry after normalisation with β-tubuline (Fig. 2D). The highest bicistronic level of eGFP expression is detected in extracts from cells transfected with a construct containing a 34bp spacer (lane 8). The expression decreased somewhat again when the spacer is shorten to 18bp (lane 9).

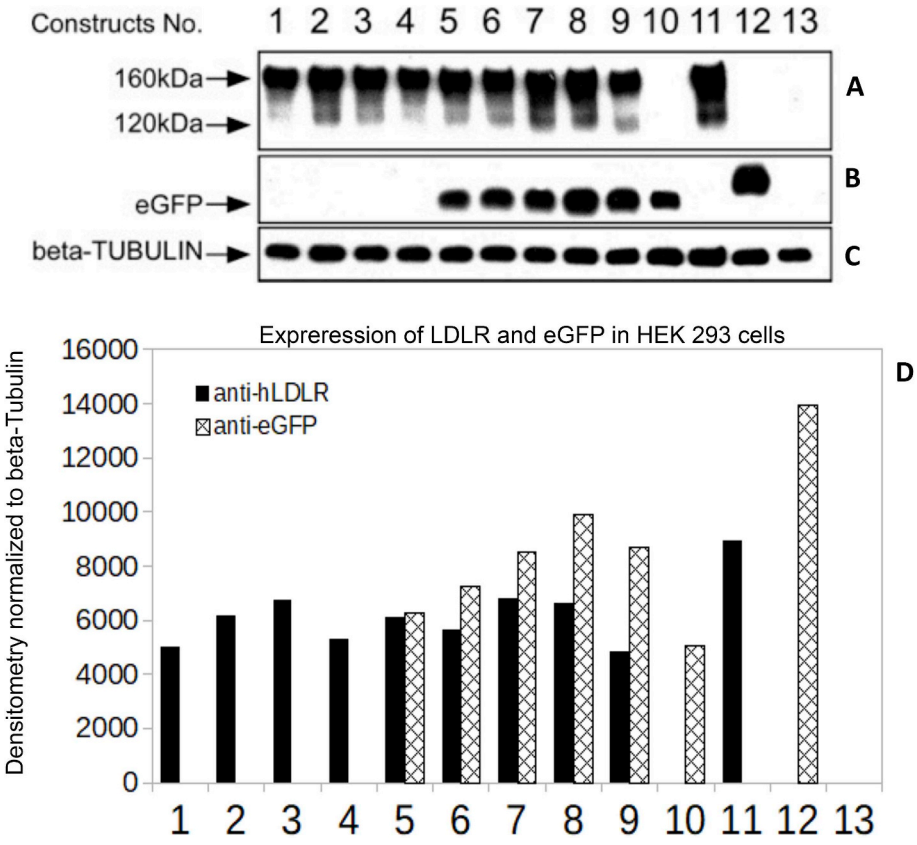


Fig. 2. Comparative expression of LDLR and eGFP proteins from HEK 293 cells transfected with various pLDLR-IRES-eGFP constructs. 5 mg of HEK 293 cellular extracts from cells transfected with 10 μg DNA of each of the corresponding construct were loaded onto each lane of 7% (***A***) and 13% SDS-PAGE (***B*** and **C**). A) Immunoblot with rabbit anti-human LDLR antibody shows the mature (160 kDa) and immature (120 kDa) forms of the hLDLR protein from all cell extracts transfected with hLDLRcDNA containing plasmids. No signals were detected from extracts transfected with pIRES-2-eGFP™ (lane 10), peGFP-C3™ (lane 12) and mock negative controls (lane 13). B) Immunoblotting with anti-eGFP antibody shows the production of eGFP protein of the expected 30 kDa molecular mass from cell extracts transfected with cassettes containing spacers of less than 216bp. The levels of eGFP expression were, generally, increasing with decreasing length of the spacer sequence. The eGFP signals were undetectable in cells extracts transfected with constructs IRES sizes: (1, 445bp, 2, 372bp, 3, 247bp, and 4, 216bp). ***D***) Quantitative densitometry analysis of protein expressions of LDLR and eGFP.

However, no eGFP signals are detectable from cells transfected with constructs with spacer lengths of 445bp, 372bp, 247bp, and 216bp (lane 1–4). In contrast, the amount of eGFP protein from cells transfected with the mono-Cistronic peGFP-C3™ plasmids (lane 12) is higher than the amount of eGFP protein produced from all other bicistronic cassettes. It is noteworthy that mono-Cistronic cassette containing IRES-eGFP mother cassette (lane 10) generates low eGFP protein compared to peGFP-C3 mono-Cistronic vector (lane 12) and also to bicistronic vectors containing 18bp, 34bp 113bp, 141bp and 188bp spacers (lane 5–9). No other proteins were recognised from the cell extracts immunoprobed with these antibodies. Flow cytometry of the transfected cells (Fig.3 A & 3B) show the mean fluorescence intensity and percentage of eGFP positive cells, which confirm the data obtained by Western blotting. Expression of eGFP is almost undetectable from cells transfected with constructs with ICS lengths of 445bp, 372bp and barely detectable from cells transfected with constructs with spacer lengths 247bp, and 216bp. The expression of eGFP peaks at a spacer length of 34bp and starts to decrease when the spacer is shortened to 18bp. In summary, the data provided above show that the length of the ICS sequences between the stop codon of hLDLRcDNA and the IRES element correlate directly with the levels of expression of the second Cistron (eGFP) with optimum expression found with a spacer sequence length of 34bp. furthermore, The Western blot and densitometry define bands and analysis were determined by LICOR Western Blot Analysis Software Free Image Studio Lite. United States.

2.3. Analyses of hLDLR and eGFP transcripts by northern blotting and qRT-PCR

Comparative analysis of transcription initiations between cells transfected with bicistronic constructs containing variable spacer lengths were performed using Northern blotting and qRT-PCR. These analyses should exclude the presence of a putative intrinsic promoter activity in the spacer sequences which might promotes eGFP expression independently. Northern analysis should also ensure that the expressed proteins are generated from the correct sizes of bicistronic mRNA transcripts. Northern blot analysis with the PCR generated hLDLRcDNA probe (Fig. 4A) revealed production of the correct ~4 kb 5'-LDLR-IRES-eGFP-3' single mRNA transcript from cells transfected with constructs vector 6 to 9 containing 141bp, 113bp, 34bp, and 18bp ICS sequences. Indeed, this blot indicates that both hLDLR and eGFP genes from the respective bicistronic cassettes were transcribed as one transcriptional unit and independently translated into two proteins. In contrast, cells transfected with bicistronic cassettes 1 to 4 with ICS of 445bp, 372bp, 247bp, and 216bp, showed the production of the 2.6 kb hLDLR mRNA transcripts only. Unexpectedly, the amount of the 2.6 kb hLDLR mRNA transcripts produced from these bicistronic cassettes was as high as from the control mono-Cistronic hLDLR construct vector 12 compared to those with shorter ICS sequence. This indicates that the long spacer might function as a transcriptional termination signal leading to the generation of mono-Cistronic transcripts instead of the bicistronic one.

Although, the Cistron located at the 5′ of IRES is preferentially produced and at a high levels. Unexpectedly, cells transfected with bicistronic cassette with spacer of 188bp (construct 5) showed the production of both the 2.6 kb hLDLR mRNA and the 4 kb 5'-LDLR-IRES-eGFP-3′ mRNA transcripts. This suggest that at this spacer length, transcription termination might occur prematurely, although, inefficiently leading to the generation of the two different sizes of transcripts. We next ruled out the presence of a putative intrinsic promoter activity in the constructs or the spacer sequences which might promotes eGFP expression independently by blotting the membrane using the PCR generated eGFP probe (Fig. 4B). As expected, the correct ~4 kb 5'-LDLR-IRES-eGFP-3′ single mRNA transcript from cells transfected with constructs 5 to 9 containing 188bp, 141bp, 113bp, 34bp, and 18bp ICS was detected. No signals were detected from cells transfected with constructs 1 to 4 containing 216bp, 247bp, 372bp, and 445bp spacers. In addition, no eGFP mono-Cistronic mRNA signal was detected from any of the constructs, which excludes the presence of a putative

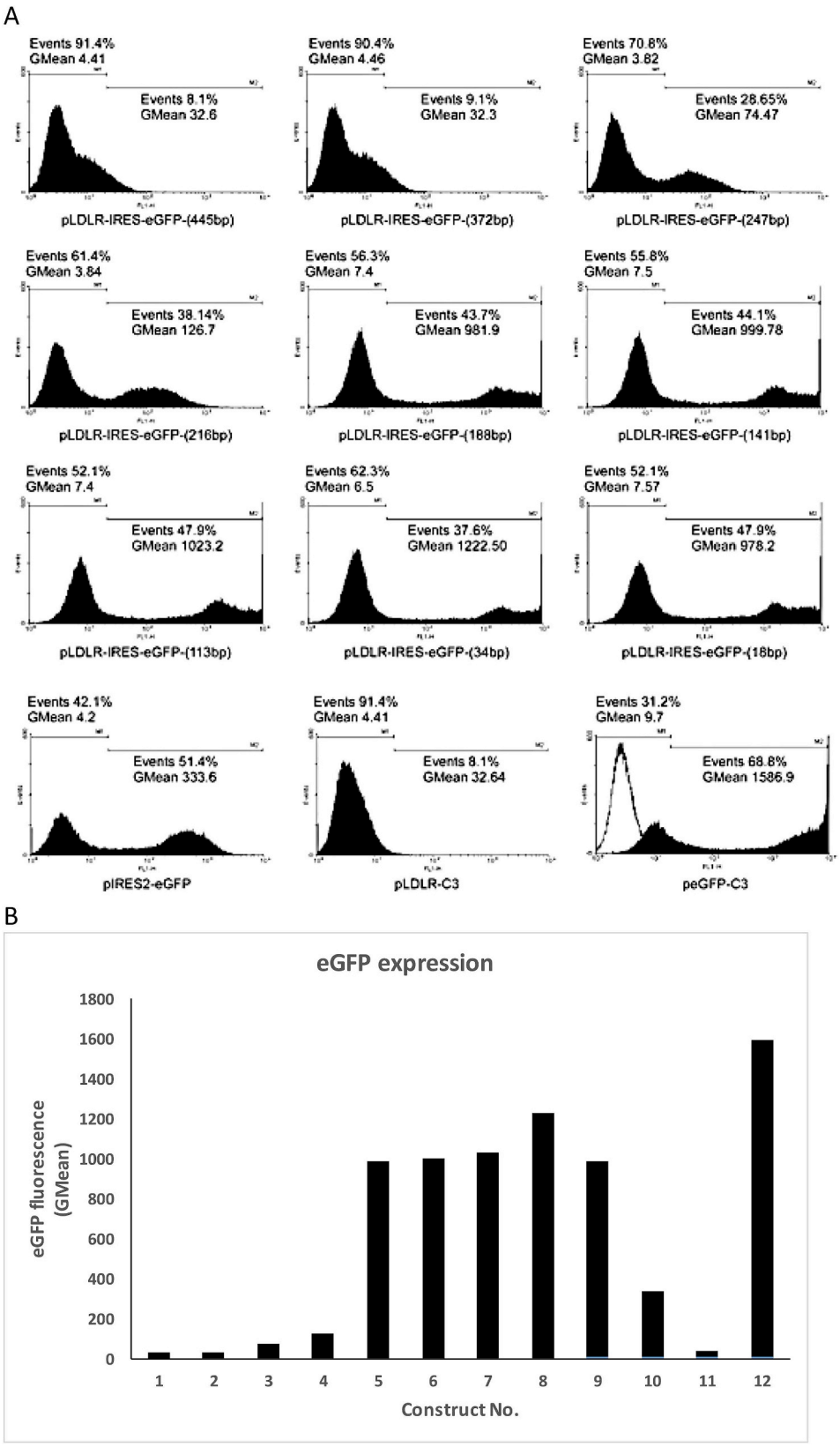


Fig. 3. Flow cytometry analysis of IRES-mediated eGFP expression from various bicistronic constructs. ***A***) 10000 events were recorded and the percentage of positive cells (given as % values on histograms) and the geometric mean of eGFP fluorescence (given as numbers on histograms) were calculated using Cell Quest software. ***B***) Bar diagram showing expression of eGFP in various constructs.

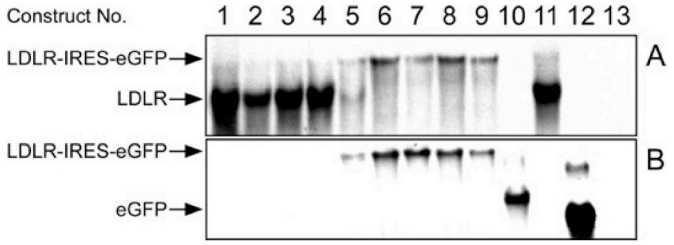


Fig. 4. Comparative analysis of mRNA transcripts produced from HEK 293 cells transfected with various constructs. 10 µg of total RNA prepared from transfected HEK 293 were loaded onto each lane of a denaturing formaldehyde gel. The blot was analysed using the hLDLR probe **(A)** then re-probed with the eGFP probe **(B).** The bicistronic 4 kb mRNA transcript is only produced from the construct 5, 6, 7, 8, and 9. Construct 1, 2, 3, and 4 produced LDLR transcript only. Construct 5 produced two transcripts.

intrinsic promoter activity in the hLDLR or the ICS sequences. The amount of single mRNA transcripts from cells transfected with bicistronic cassettes was generally lower, compared to cells transfected with the mono-Cistronic control constructs.

To generate quantitative data for accurate comparative analysis of the above blots, qRT-PCR was used with primers, which amplify the hLDLR, and eGFP sequence relative to the endogenous ribosomal protein L19 transcript. The qRT-PCR analysis (Fig. 5) confirms the data obtained by Northern blotting. Expression of eGFP is barely detectable from cells transfected by constructs with spacer lengths of 445bp, 372bp, 247bp, and 216bp, respectively, which confirms that longer spacers may enhance transcription termination. The expression of eGFP is increased significantly at a spacer length of 188bp, peaks at 34bp and starts to decrease when the spacer is shortened to 18bp.

2.4. Analyses of biological activity and in vivo expression of optimised cassette

In order to test the functionality of the LDLR-IRES-eGFP-8 optimised cassette, we transfected the LDLR-deficient CHOldlA7 cells and confirmed correct localisation of both the hLDLR and the eGFP transgenic proteins. We next investigated the ability of the transgenic hLDLR to bind and internalise fluorescently labelled-LDL (Dil-LDL™) in vitro. Living CHOldlA7 cells were transfected with the pLDLR-IRES-eGFP-8 vector and assayed microscopically (Fig. 6). Only CHOldlA7 cells which had been transfected with the pLDLR-IRES-eGFP-8 (panel A) and those transfected with pLDLR-C3 (panel B) expressing native LDLR were able to internalise Dil-labelled LDL particles. No internalisation of Dil-LDL™is observed in non-eGFP expressing CHOldlA7 cells shown on the same slides and CHOldlA7 cells transfected with the peGFP-C3™ plasmid (panel C). Assuming that in vivo experiments are best suited to test the efficacy of the optimised 5'-hLDLR-IRES-eGFP-3′ cassette, a

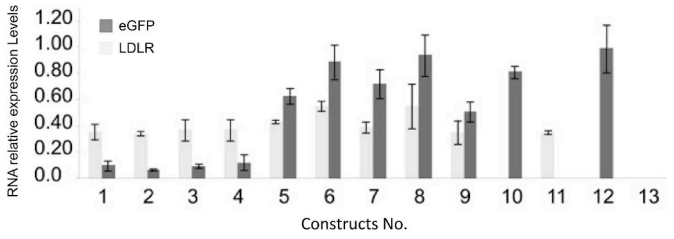


Fig. 5. Comparative analysis using qRT-PCR with primers, which amplify the hLDLR, and eGFP sequence relative to the endogenous ribosomal protein L19 transcript. The RNA relative expression levels of eGFP are barely detectable from cells transfected with constructs with spacer lengths of 445bp, 372bp, 247bp, and 216bp (constructs 1 to 4). The expression of eGFP is increased significantly at a spacer length of 188bp and peaks when the spacer length is 34bp (construct 8).

pilot experiment was conducted by injecting 20 µg of plasmid DNA into adult MF1 mice fed a normal chow diet. Vector delivery was performed using hydrodynamic gene transfer method [1,2]. The peGFP-C3™ plasmid was injected as an eGFP control. Injected mice were sacrificed approximately 72 h s after vector delivery and liver cells analysed microscopically confirmed eGFP-expressing cells (approximately 15–20% of liver mass.

2.5. Interaction between IRES RNA structure and 40S human ribosomal subunit

There was no structure available for IRES (Emcv). To the best our knowledge, this is the first time we predict the IRES RNA structure model of all its domains and Emcv IRES RNA protein complex using homology Modeling through MD simulation that's also showed an interactions with the 40S human ribosomal subunit during translation initiation (Fig. 7A–E). The topography of the initiation complex was assessed using the predicted protein domains ribosome and IRES: ribosome interactions as well as contact data information. It has been found that the interactions of eIF4G and eIF3 with type II IRES RNAs are the most useful [28,29]. These two aspects are essential for the recruitment of Emcv IRES RNAs to ribosomes [30]. EIF4G acts in conjunction with eIF3 as a connection between the IRES RNA and the 40S human ribosomal subunit. The C-terminal portion holds the binding sites for both the IRES RNA as well as eIF3 [31]. It binds to the 40S human ribosomal subunit in the absence of other initiation factors and has been shown to make multiple contacts with the Emcv IRES RNA [32]. Although the exact location of the eIF3 binding sites on Emcv IRES RNA is not clear, Cryo-EM studies of eIF4- G: eIF3 and eIF3: 40S complexes place the Emcv-associated eIF4G near the ribosomal E site [33]. Due to the lack of data for this complex near the E site of the ribosome, Emcv IRES: eIF4G interactions were not modeled (Fig. 7E), consistent with the models of eIF4G: eIF3 and eIF3: 40S complexes. With helix-1 located near the E site, the AUG start codon can be positioned near the P site without disturbing the rest of the model. This placement was consistent with data analysed from to eprinting analysis of Emcv IRES RNA bound to human ribosomes, which demonstrated protections of 12 nucleotides downstream of the first start codon (824–836). The initiation codon of IRES RNA model was located in the P site.

This particular arrangement was plausible in Emcv IRES RNA because translation started at this codon and would permit the formation of a cross-link between the IRES RNA and rpS25. In Emcv IRES RNA, translation initiates either at the first start codon (462-AUG- 464) or the preferred start at codon present after 12 nucleotides downstream. For the second AUG codon to reach the P site, the ribosome scanned the Emcv RNA [33]. Translation factors eIF1 and eIF1A promoted scanning by inducing an open conformation of the ribosomal subunit. The Emcv IRES was expected to be outside the ribosome by the time the second AUG codon enters the P site. Translation initiation site selection is determined, in part, by the context of the nucleotide sequence surrounding the first AUG-MET codon encountered by the scanning ribosomal subunit. A results in the scanning ribosome initiating translation from this weak AUG at a low frequency or bypassing it completely in favor of a stronger downstream AUG start site. The Met-charged initiator tRNAi brought to the P-site of the small ribosomal subunit by eukaryotic Initiation Factor 2 (eIF2). It hydrolyzes GTP, and signals for the dissociation of several factors from the small ribosomal subunit, which results in the association of the large subunit (or the 60S subunit). The complete ribosome (80S) then commences translation elongation, during which the sequence between the 'start' and 'stop' codons is translated from mRNA into an amino acid sequence.

2.6. Long range RNA tertiary interactions within IRES elements

The intramolecular interaction in domains of IRES (Emcv), showed

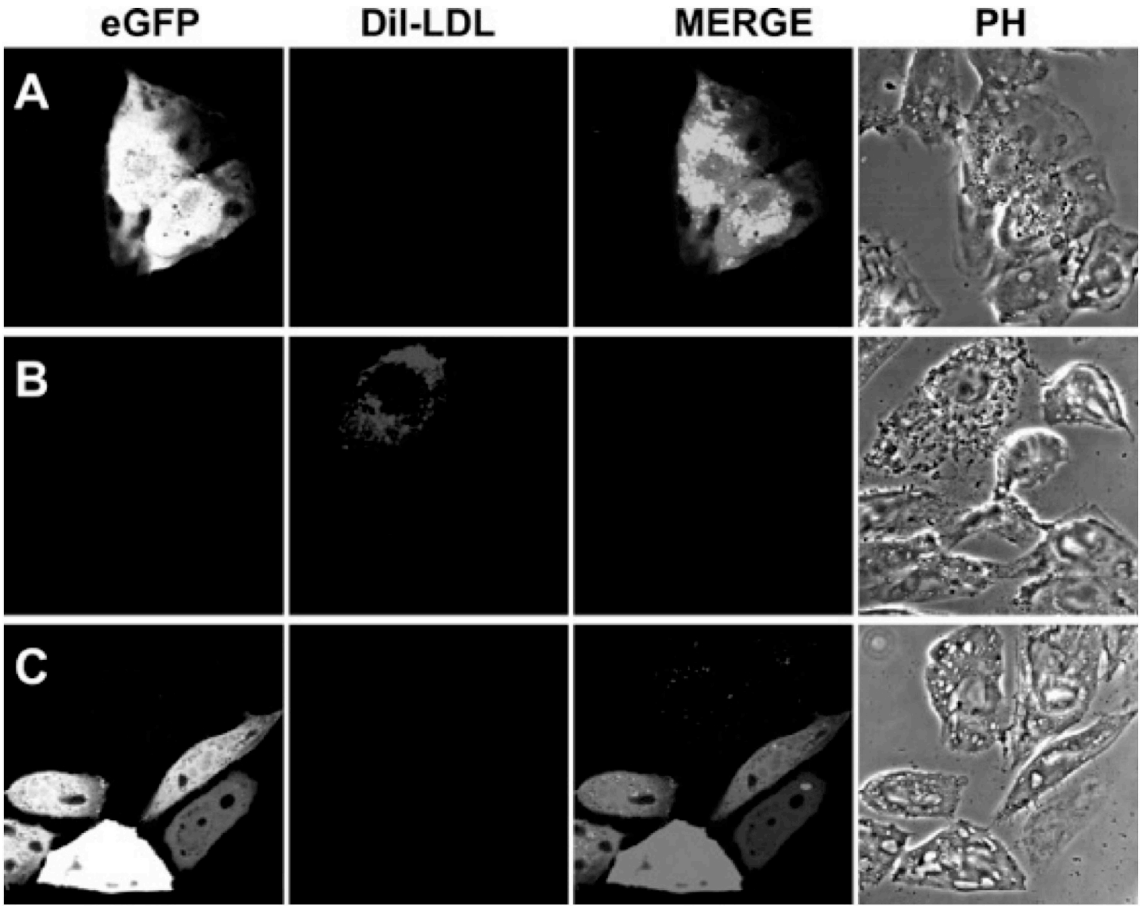


Fig. 6. Internalisation of Dil-labelled LDL by living transfected CHOldlA7 over-expressing the optimised 5'-hLDLR-IRES-eGFP-3' cassette. Cells on panel A to C were transfected with pLDLR-IRES-eGFP-8, pLDLR-C3, and pEGFP-C3™ respectively. Only eGFP expressing cells on panel A or those expressing native LDLR (panel B) were able to internalise Dil-labelled LDL particles. No internalisation was detected in non-eGFP expressing cells on same slides or in those transfected with the pEGFP-C3™ plasmid (panel c). PH is phase contrast.

that the specific binding receptor of in tetra loop was unknown, although the tertiary contact was required for efficient IRES activity, the specific binding receptor of in tetra loop was unknown. We selected a final Model IRES (Emcv) and found two receptor candidates (G231/C241, and C232/G240) base pairs interacting with A180 and A181 residues in GUAA tetra loop. The trajectory for IRES (Emcv) shows that the two adenosines retained a distance < 3 Å. In contrast, only the first adenosine A180 of other Models retain a distance < 4 Å during the initial 10, and 20 ns, respectively. In Model IRES (Emcv), the average distance between C232/G240 pair and A180 is 2.0 ± 0.50 Å while C231/G241 pair and A181 was 2.0 ± 0.21 Å. These findings targeted receptors of A180 and A181 residues, respectively. The tertiary interaction of IRES (Emcv), we considered the Leontis/Westhof nomenclature and analysed the three edges Watson Crick, Hoogsteen and Sugar edge for potential hydrogen bonding interactions. The measured minimum distances between the Sugar edge of each C232 to G240 and C231 to G241 base pair with three edges of each A180 and A181 over the 100ns time show tightly formed hydrogen bonding interactions for the sugar edge/Watson–Crick between the C232/G240 pair and A180 and sugar edge/Hoogsteen edge tertiary interactions between G231/C241 pair and A181. In addition, we observed tertiary contacts between U179 and A234 residues via trans Watson–Crick/Watson–Crick edge interactions at 22 ns? These long-range interactions occurred sequentially at 7, 20 and 22 ns, involving A180, A181 and U179, respectively. These co-operative long-range interactions helped in the stabilization of the IRES domain.

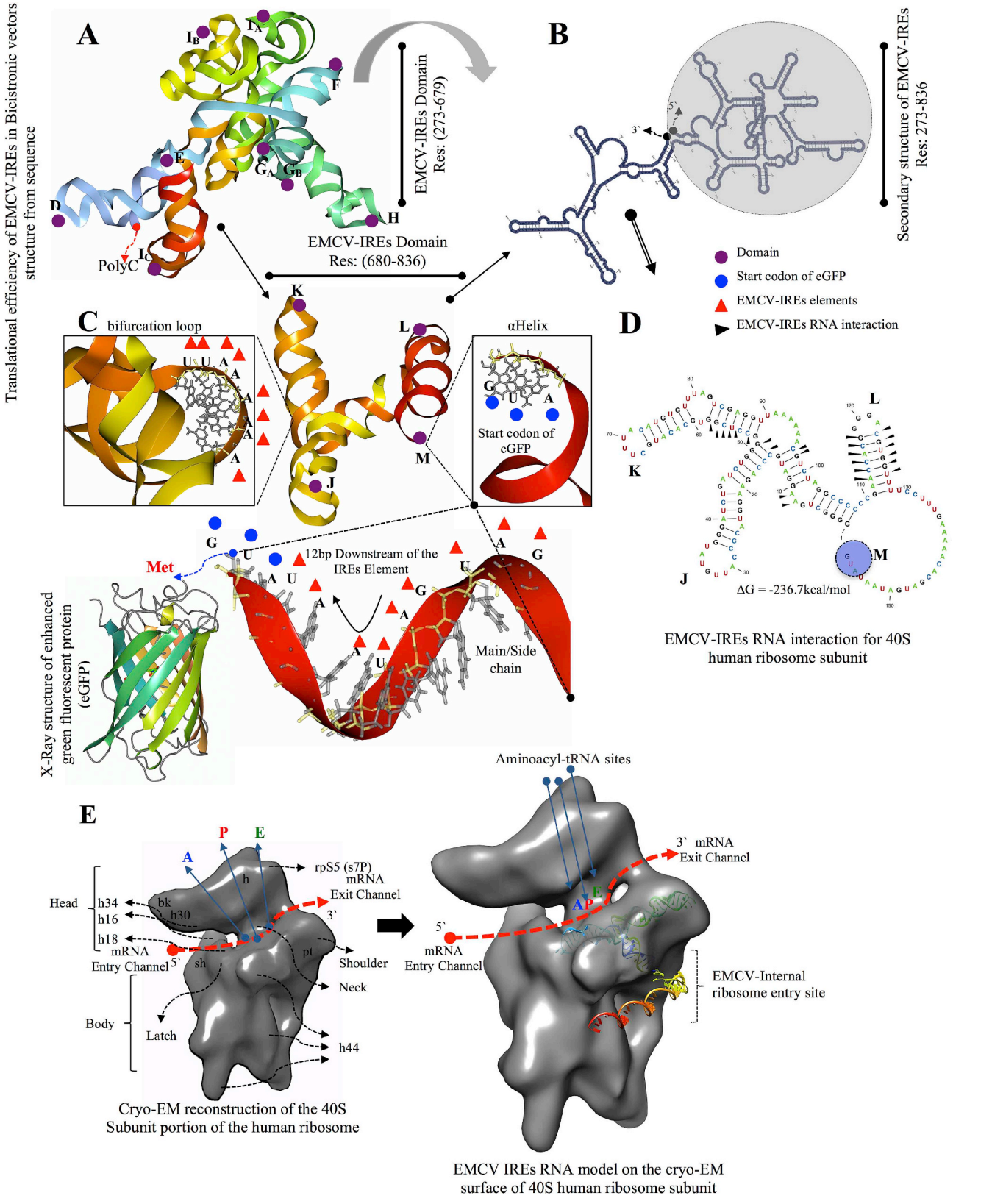
3. Nine constructs of spacer sequence and thermodynamic free energy

IRES calculate spacer sequence of 1–9 constructs minimum free energy of the thermodynamic ensemble mRNAs structures. The genetic code of the sequences introduced has to be in the correct form. It was necessary then to change all the timelines present in the sequences to the uracile residues. The optimal secondary structure symbolized with brackets and points, and it also shows the minimum free energy in kcal/mol of each IRES spacer sequence construct structure (Fig. 8). The more negative results are more stable thermodynamically equilibrium structure. The secondary structure predicted by RNA-Composer is due to the pattern predicted structures corresponded to the real IRES structure of our target. The pattern we had only used one nucleotide upstream the loop-bulge. For this reason, the secondary structure was not as stable as if more base pairings had been included in the stem. In all nine IRES constructs predict the molecular dynamic of tertiary structure shown in the clear thermodynamic energy scale level. The enthalpy and entropy are state functions, so changes in their values depend on the path between as initial and final state. We did for the free energy construct thermodynamic cycles for the changes in enthalpy and entropy for the good free energies and stability of the IRES spacer sequence construct.

4. Discussion

When an IRES is used for the construction of a multi-cistronic cassette in gene transfer and expression vectors, it has been frequently observed that the DNA sequence subcloned at the 3′ to the IRES element

is poorly expressed compared to the sequence subcloned at the 5′ of the IRES [34]. In most of the reported articles this phenomenon is attributed to the low efficiency of IRES mediated translation initiation compared to cap dependent translation. A possible explanation is a higher affinity of translation factors for the cap structure than for the IRES element. Therefore, transcription factors may become less available for the internal initiation of translation [35]. Alternatively, the IRES may require additional host *trans*-acting factors, which modulate its function in a cell or tissue type specific manner [36]. Thus, high levels of IRES mediated internal translation initiation can be



(caption on next page)

Fig. 7. Emcv (IRES) structural and functional annotation. A) Three-dimensional model structures of IRES (Emcv) RNA domains modeling structures consists in to two form of (domains Residues. 273–697); D, E, IA, IB, IC, GA, GB, H, F) is represent in the helical form. B) The complete domains of IRES 2D structure of the RNA with highlighting in grey Color for all domains. C) Other (domains Residues. 680–836; J, K, L, M). Domains are in different colors as in the secondary structure schematic depicted at the lower right. The 5′ and 3′ ends of domains along with conserved and variable subdomains are shown. The main helical stalks are shown in orange and yellow. The locations of helices and the tetra loop are shown in red. Bifurcation loop residues represent are in red Color. The images were generated using QTMG. X-ray crystal structure of eGFP start coding with MET and 12bp downstream in (red) of the IRES element were representing different colors blue are AUG stop codon and initiate of start codon eGFP MET. D) Secondary structure of IRES (Emcv) RNA interaction for 40S human ribosome subunit. The interacting residues are highlighted in red Color. E) Hypothetical placement of protein-constrained IRES (Emcv) RNA on the Cryo-EM surface representing of the human 40S subunit. Cryo-EM reconstruction of the 40S subunit portion of the human ribosome with traveling of mRNA 5′ to 3` (entry to exit channel) and also showing interaction with IRES (Emcv) entry in the APE site. Coloring of the IRES (Emcv) RNA ribbon colors were in green (J), blue (K), yellow (L) and orange (M). The first start codon (green) is indicated, and the naked human 40S subunit (EMDB ID 1092) was retrieved from the EM-Database. These images were generated using UCSF Chimera.

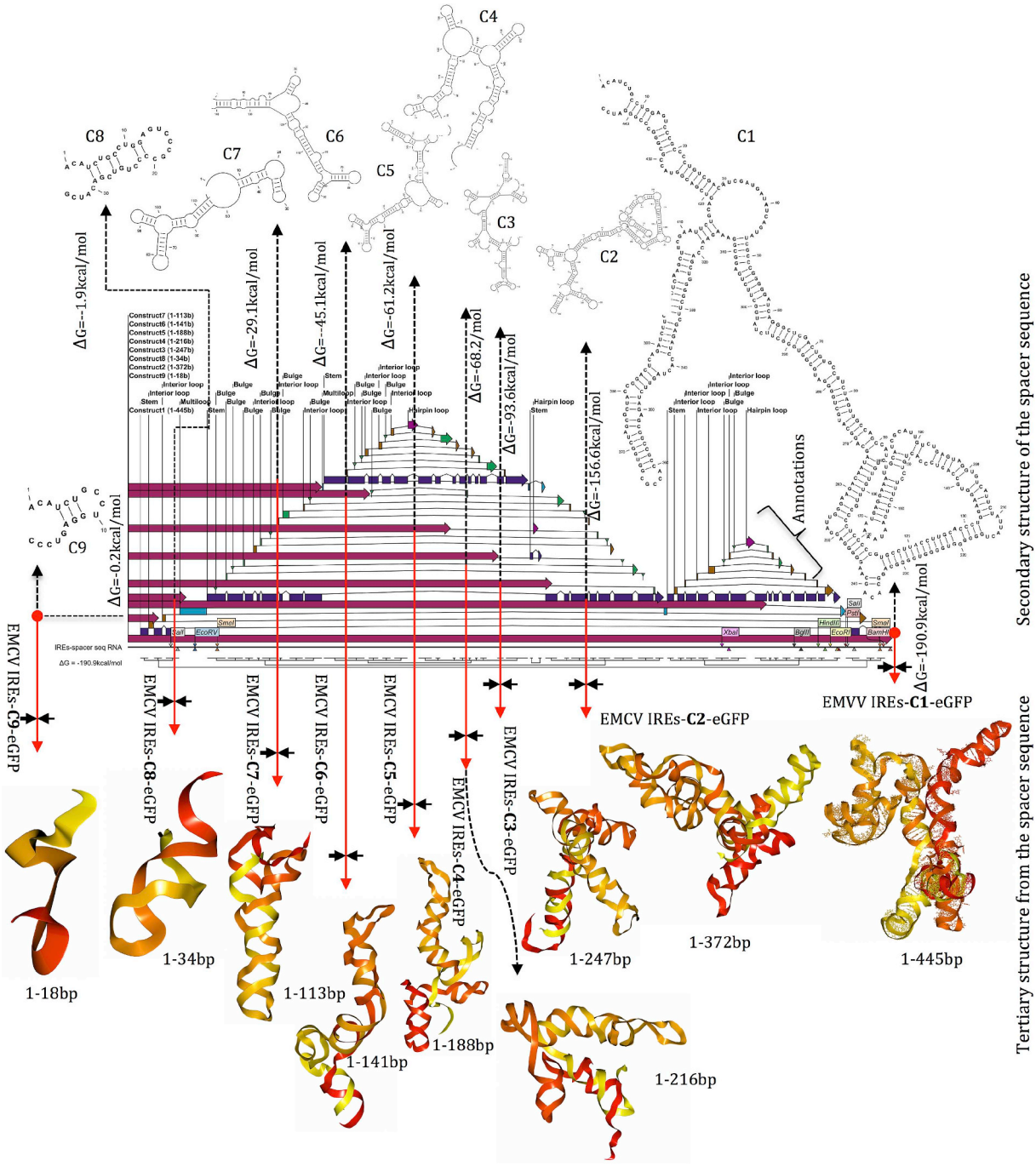


Fig. 8. A constructs of 1–9 spacer sequence and free energies. The different thermodynamic ensemble free energy of all constructs [1–9] of spacer sequence and the molecular dynamic structure were determined by simulation. ***A)*** The 2D secondary structure of EMCV-IRES constructs (C1−C9) as shown in RNA 2D view black. ***B)*** The secondary structure of EMCV IRES in 1–9 as shown in red bar with denoted different thermodynamic free energy. ***C)*** The tertiary EMCV-IRES structures as shown in the helical state.

anticipated depending on the nature of the IRES sequence and its tissue tropism [16]. Another important factor that might have a strong influence on IRES dependent translation initiation is the length and possibly the nature of the ICS sequence flanking the IRES element at both its 5′ and 3' ends [9].

In the present investigation, an IRES derived from the Emcv genome was used to mediate eGFP marker gene translation initiation. We have successfully constructed nine bicistronic cassettes containing sequentially deleted non-coding ICS sequence, between the stop codon of the hLDLRcDNA and the IRES/eGFP elements. We then analysed the expression levels of the respective transgenes systematically at the DNA, RNA and protein levels. By tailoring the size of the ICS sequence at the 5' of IRES, we showed for the first time that IRES dependent expression of the second gene could be significantly increased to the level of the cap-dependent expression of the first gene. Maximum expression efficiency of the eGFP was obtained when the ICS was 18bp to 141bp in length. However, when the ICS was 216bp or longer, the expression levels of eGFP were reduced significantly and become completely undetectable at a spacer length of 445bp. These results are in agreement with data published by Attal et al. (1999a) who found that IRES elements from Emcv as well as from poliovirus function optimally when about 100 nucleotides were added after the termination codon of the first Cistron (luciferase). These authors also showed that IRES elements become totally inefficient when added after nucleotide (300–500) spacer [9]. In their constructs the second Cistron was the chloramphenicol acetyl transferase gene (CAT). In a similar study using the IRES elements from poliovirus and from SV40, respectively in bicistronic cassettes containing the firefly luciferase gene as the first Cistron and the CAT gene as the second Cistron, Attal et al. (1999b) also showed that maximum efficiency in the expression of the second Cistron was achieved when the spacer fragment was composed of 30–90 nucleotides. This expression was profoundly reduced when the ICS fragment contained only 8bp and also when it was 300bp long. Finally, expression was undetectable when the spacer fragment was 500 nucleotides [37]. However, despite the remarkable strong influence of the length of ICS sequence on gene expression of the second Cistron in Attal studies, the levels of expression from the second gene were never similar or exceeding those of the first gene.

It has been frequently argued that the low level of second Cistron expression is due to the low efficiency of IRES functions in the cell in questions. We showed here that the low levels of expression from the second gene when ICS of 218bp or longer is used, is due to the generation of a single transcriptional unit containing the LDLR Cistron only, suggesting that long ICS sequence may serve as an insulator for transcription termination. To exclude the effect of tissue tropism, we have transfected our constructs into Hela and CHOldlA7 cells and obtained similar results. We have also noticed that the amount of mRNA containing the hLDLR and eGFP Cistrons does not correlate tightly with the levels of translation of both Cistrons. We therefore cannot exclude that the sequentially deleted sequences might also contain an element that influences mRNA instability. However, we do not have detailed data to support this assumption, which is nevertheless worth future investigation.

To determine whether the phenomenon illustrated above can be generalized to other genes in this arrangement, expression from first and second gene was also compared using a bicistronic plasmid with hLDLRcDNA 5′ from a 445bp ICS identical to construct 1 and the Hygromycin gene in 3′ position. Hygromycin activity in HEK 293, Hela and CHOldlA7 cells transfected by pLDLR-IRES-Hyg was undetectable. We have also constructed another bicistronic cassette with the human Olig2cDNA 5′ and the dsRed2 gene 3′ to a different 400bp ICS sequence. The fluorescent of dsRed2 protein was undetectable in HEK 293, Hela and human fetal mesenchymal stem cells transfected by this construct. Shorting the spacer by 200bp permits high expression of dsRed protein without affecting the expression level of Olig2 transcription factor. We are also aware that permanent phenotypic corrections of LDLR deficiency by gene therapy may require long-lasting expression of the transgenic LDLR in the relevant animal model to judge its efficiency, therefore, we have constructed and examined several viral plasmids based on the lentiviral backbone under the control of ubiquitous CMV promoter. However, the production of viral particles in HEK 293T producer cells following a standard three-plasmid co-transfection protocol failed to generate infectious vector. Owing to the utilisation of a previously characterised liver specific promoter (LSP) [38], Kankkonen and colleagues demonstrated for the first time successful construction and production of high titre (1×10^9 IU) third-generation HIV-1 based lentiviral vectors encoding rabbit LDLR [42].

Comparative sequence analysis [39] is the standard procedure for predicting of RNA secondary structure with an all-atom model. Over 99% base pairs accuracy can be obtained in ribosomal RNA secondary structures, predicted by comparative sequence analysis similar to the one exist in high-resolution crystal structures [40]. However, comparative sequence analysis required IRES a large number of homologous sequences from various databases. In the absence of necessary homologous sequences, minimum free energy IRES RNA structure prediction can be applied to predict the structure of a single RNA sequence with an average of 90% accuracy through molecular dynamics (MD) simulation [41]. This accuracy is sufficient to be used as a starting point to build an alignment for comparative sequence analysis [42] for the target IRES RNA sequence. Identify a specific structure at the 3′ border of the IRES that prevents the correct positioning of the AUG-MET initiation codon in the close vicinity of the ribosomal P site. Taken together, these results demonstrate the presence of an RNA determinant in IRES that allows ribosomal 40S subunit binding, although it is not sufficient to correctly position the AUG in the ribosomal P site. This allows us to propose a new model for initiation complex assembly on IRES and to refine the role-attributed stabilizing and promoting the conformational changes.

We predicted a model of IRES (Emcv) and determined that IRES RNA secondary structures, which is reliable method to model, the biologically meaningful tertiary structure [43]. The base-pairs information was utilized for molecular structure prediction using RNAComposer, which facilitates RNA structures as well as three-dimensional model with form of RNA helical information [44], based on reliable algorithms. The machine translation principle and operates on the RNA FRABASE database acting as the dictionary relating to RNA secondary structure and tertiary structure elements. Furthermore, Covariation-modeling analysis was not limited to defining secondary structure but may also be used to identify possible long-range interactions [45]. Enzymatic and chemical probing of mutant IRES (Emcv) RNAs suggested that the GNRA tetra loop capping helix might form a tertiary contact with 240-GCACG-244 in the helix [34]. Our attempts to identify canonical RNA-RNA tertiary interactions for constraining the model were successful. We used data symbolizing Polypyrimidine tract binding protein (PTB) interactions of Emcv IRES RNAs. PTB enhanced the translation initiation directed by IRES RNAs of Emcv and hepatitis A virus (HAV) [46]. It has been showed that the Emcv dependence on PTB is conditional [47]. IRES-RNA self-folding structural element due to the intramolecular RNA–RNA interacting region has thus been suggested to contribute significantly to the IRES structural organization and stability of domain, and to the critical function of IRES activity [30]. IRES-mediated translation initiation is closely linked to structural organization in domain, specifically the apical region formed by two four-way junctions enabling the RNA–RNA intramolecular interactions. Thus, we focused on the apical region of IRES domain to decipher the spatial arrangement of the RNA fold that is a prerequisite essential step to understand the initiation mechanism of translation. The contact between the 40S human ribosome subunit and IRES RNA was visualized by Cryo-EM. Cross-linking of rpS5 to rpS25 has been suggested in between these proteins [48]. Recent X-ray diffraction studies of the structure of eukaryotic ribosomes have also confirmed this proximity

and indicate that rpS5 was located at the solvent side of the head of the 40S subunit [49]. We therefore proposed that helix1 domain was similar to these ribosomal proteins. The translation factors of eIF1 and eIF1A promoted scanning by inducing an open conformation of the ribosomal subunit. The IRES was expected to be outside the ribosome by the time the second AUG codon enters the P site.

A recent study found that the competition between HCV domain III and eIF3 for binding with 40S subunit may result in the reduction of 43S complex formation and may thus favor translation of HCV mRNAs [50]. We did not find any variants with changes within the eIF3 binding sites in domain III. However, the relevance of identified mutations, especially at position 183, for eIF3 interaction and translation cannot be excluded. The thermodynamic equilibrium method predicts the IRES RNA structure with the lowest equilibrium free energy, as the sum of independent contributions of stacked and base pair instem and loops [51]. Although Lu et al., 2009 refers that the Maximization of expected accuracy provides a better approximation than free energy minimization [52], in the selected IRES structure, thermodynamic and weight scoring scheme provides the best estimation of RNA 2D structure based on a minimum free energy calculations.

However, The AUG codon initiate of MET site selection for the main preference for initiation and dependent in part on the context of the AUG codon. In this case, eIF2 does not contribute in the formation of a ternary complex but, slightly, may it bind directly to the ribosomal A-site and facilitate binding of the Met-tRNAi to the ribosomal P site during translation initiation. EIF2 activity is not subject to direct regulation and consequently may not contribute to the control of mRNA translation.

Future work will aim to subclone our optimised 5′-hLDLR-IRES-eGFP-3′ cassette into a vector with liver specific promoter (LSP) promoter. The use of LSP to drive the expression of our optimised bicistronic cassette may permit physiological expression of LDLR, thus, avoiding the pathological intracellular lipid accumulation [53,54] that may damage the transduced cells in the relevant animal model. In addition, it should permit easy evaluation of vector bio-distribution following gene delivery.

5. Materials and methods

5.1. Plasmid construction

The hLDLRcDNA with the 347bp non-coding spacer sequence was excised from pBC-hLDLR by digestion with *HindIII* and *SacI* and ligated into *HindIII/SacI* digested pLitmus28™ (NEB, UK) to produce SpLitmus-hLDLR. The hLDLRcDNA and the 445bp downstream spacer sequences were derived from the pRSVh-LDLR plasmid. Using a novel PCR-based method, we were able to create several clones of Splitmus-LDLR with sequentially deleted sequences downstream from the LDLR open reading frame (LDLR-ORF) with a unique *HindIII* and *KpnI* restriction sites. After restriction of these clones with *HindIII* and *KpnI*, the fragments containing the LDLR-ORF and the non-coding “spacer” sequence downstream were blunt ended with Klenow and then ligated into *BglII/BamHI* digested and blunt ended pIRES-2-eGFP™ expression vector. All expression vectors containing bicistronic cassettes were successfully constructed and their integrity confirmed by DNA sequence analyses. Expression of the transgenic proteins from these vectors is driven by the promoter cytomegalovirus (CMV). The integrity of the hLDLR-insert in the isolated pLDLR-IRES-eGFP clones was examined by restriction digestion and DNA sequence analysis. The pLDLR-C3 mono-Cistronic hLDLR expressing plasmid, which serves as a positive control for hLDLR expression, was constructed by removing the IRES-eGFP cassette from the pLDLR-IRES-eGFP-1 construct using *SalI* and *NotI* enzymes. The plasmid DNA was then blunt ended by Klenow and relegation with T4 DNA ligase (NEB, UK).

5.2. Gene transfer into cultured cells

Human embryonic kidney epithelial cells *(HEK 293)* and LDLR-deficient Chinese hamster ovary (CHOldlA7) cells were cultured with Dulbecco's modified Eagle's medium (Invitrogen or Sigma, UK), supplemented with 10% fetal calf serum (FCS). All cells were incubated at 37 °C with 5% CO2 in a humidified incubator. Cells were seeded into 15 cm^2 dishes at 60–70% confluence with a density of 4–6x10^6 cells and were transfected the following day with 10 μg of the expression vector DNA using FuGene6 (Roche, UK) according to the manufacturer's instructions. Forty-eight hours later, expression of fluorescent proteins was examined with a fluorescence microscope. All transfection experiments were repeated at least three times with similar results.

5.3. Flow cytometry analysis

Flow cytometry analysis was carried out for the detection of cellular expression of eGFP using a FACScan flow cytometer (Beckton-Dickinson, UK) with the FL1 detector channel. The data were acquired and analysed with CellQuest software (BD). 48 h post-transfection transfected HEK 293T cells were rinsed with PBS, then analysed for GFP expression. 10000 events were recorded. Untransfected and Mock transfected cells were used as controls. Results were presented as a percent of positive cells. Mean fluorescence intensity was used as an indicator of relative expression of eGFP on given cell population.

5.4. Preparation of total RNA and northern blot analysis

Total RNA was extracted with TRIzol based on the manufacturer's instructions (Invitrogen, UK) and treated with *DNase* I (NEB, UK). Thirty micrograms total RNA from each sample were mixed with 2.5 parts of RNA loading buffer in a total volume of 35 μl, heated for 3 min at 100 °C and chilled on ice. Samples were then separated on a 1% denaturing agarose-formaldehyde RNA gel prepared by microwave melting of 3g agarose (Invitrogen, UK) in 200 mls DEPC-treated H_2O. Boiled agarose was allowed to cool to 60 °C, then 56 mls of 5× MOPS buffer and 50 mls of 37% formaldehyde were added). After 4 h s of electrophoresis at 100 V, the gel was soaked in 0.05 M NaOH for 20–30min with gentle agitation, rinsed with DEPC-treated H_2O and soaked in 20× SSC for 45min. Single-strand RNA was then transferred from the agarose-formaldehyde gel to a solid nylon membrane, Hybond-N^+ by the capillary method. Gene images random prime labelling™ kit (Amersham, LIFE SCIENCE, UK) was used for labelling of 50 ng-purified PCR generated DNA probe following the manufacturer's instructions. The blotted membrane was pre-soaked in 5× SSC buffer before hybridisation and then placed in a hybridisation bag containing 50 mls of pre-hybridisation solution for 60 min at 60 °C with gentle agitation. The labelled probe was then added into a fresh hybridisation buffer and hybridisation was carried out at 60 °C for 18 h s. Following hybridisation and washing, the gene images CDP-star detection™ kit (Amersham, LIFE SCIENCE, UK) was used for detection of hybrids according to the manufacturer's instructions.

5.5. Reverse transcription (RT) and quantitative polymerase chain reaction (Q-PCR)

cDNA synthesis was accomplished with 1 μg of total RNA in the presence and absence of reverse transcriptase (RT). The final reaction mixture contained 1 μg RNA, 10 units RNasin RNase inhibitor, 10 mM dNTP mix, 0.5 μg random primers (Promega, UK), 5× reverse-transcriptase buffer, with (RT+) or without (RT-), 100 units M-MLV reverse transcriptase (Promega, UK), in a final volume of 20 μl. Quantitative expression analysis of the housekeeping gene encoding the ribosomal L19 protein gene (GenBank accession no. NM_000981) and the genes of interest *Ldlr* (GenBank accession no. AB209409), *e*GFP and *IRES* sequence was determined by Q-PCR using an ABI PRISM™ 7700 Sequence

Detector following the ABI protocol (Applied Biosystems, UK). The Q-PCR assays were run in triplicate on 96-well plates. Primers (5'-3' sequences used for quantitative analysis are: L19-F 5′-gcggaagggtacagccaat-3'; L19-R 5′-tttgcgccggctgc-3'; LDLR-F3 5′-tgaggtccacatttgccaca-3'; LDLR-R3 5′-catcctccagactgaccatctg-3'; eGFP-F1 5′-gtccgccctgagcaaaga-3'; eGFP-R1 5′-tcacgaactccagcaggacc-3'; IRES-F1 5′-gccggtgtgcgtttgtctat-3'; IRES-R1 5′-ttccgggccctcacatt-3′ were designed according to real-time PCR requirements using the Primer Express 1.7 (Applied Biosystems, UK) software. All primer sets were free of primer-dimer products. SYBR® Green was used in the Q-PCR assays. A standard curve using a control placental cDNA pool was obtained in each assay. Thermal cycle parameters included an initial incubation at 50 °C for 2 min for 1 cycle; Taq activation at 95 °C for 10 min for 1 cycle; repetitive denaturation at 95 °C for 15 s and annealing at 60 °C for 1 min, for 40 cycles. Following completion of the quantitative amplification of the samples of interest, values were obtained and analysed using the Sequence Detector 1.7 software (Applied Biosystems, UK). All triplicate cycle threshold (Ct) values had to be within 1 Ct of each other and the correlation coefficient (r^2) above 0.98. The quantitative values for each triplicate were averaged and the relative expression of the genes of interest was determined by a ratio of their expression to that of the housekeeping gene for the same sample.

5.6. Western blot analysis of protein expression

Forty-eight hours after transfection, cells were harvested by scrubbing and re-suspended in 30μl/10^6 cells ice-cold lysis buffer (10 mM Tris pH 7.5, 1 mM EDTA, 1% NONIDET P-40 (Sigma, UK), protease inhibitor cocktail at indicated dilutions (Roche, UK). Cell suspensions were then centrifuged at 13,000 × *g* for 20 min at 4 °C. Supernatant were re-centrifuged again at 40,000 × *g* for 45 min at 4 °C. The protein content of cell lysate supernatants was assayed by the bicinchoninic acid (BCA) method using a commercial assay kit from Pierce with BSA as a standard. Five micrograms protein of each lysate were combined with equal volumes of Laemmli sample buffer, boiled for 5 min and were then separated on a 13% or 7% SDS-PAGE gel for GFP and LDLR protein detection, respectively. Gels were then blotted onto PVDF membranes with a Hoefer apparatus (200 mA). After 1 h incubation at room temperature in blocking solution (5% dried skimmed milk in PBS-Tween-20 0.1%), membranes were incubated overnight at 4 °C with specific primary antibodies, diluted as specified below with blocking solution. After three 5-min washes in PBS-Tween-20, blots were incubated at room temperature for 1 h with peroxides-conjugated goat anti-rabbit antibody (Abcam) or rabbit anti-mouse polyclonal antibody (Dako, UK) diluted 1:1000 in blocking solution. Following the final wash, detection on autoradiography hyper-films was performed after inducing a chemiluminescence reaction with the Amersham Super signal detection kit. Densitometric analysis was performed using ImageJ. Primary antibodies used in this study were anti-LDLR (polyclonal, 1:2000; Research Diagnostic, USA) and anti-GFP (polyclonal, 1:2000; abcam, UK) and anti-beta-tubulin (1:2000; Sigma, UK).

5.7. Emcv-IRES molecular modeling

The models were constructed using the sequence of the hLDLR-IRES (Emcv) RNA from (GenBank accession number X74312 positions 273–836). Its was consisted of different domains (starts from 5' polyC residue 273, D, E, F, G^A, G^B, H, I^A, I^B, J, K, L M (AUG-end of the Stop codon). The models of Emcv-IRES were determined by MD simulation methods. Briefly, the secondary structure information was used as input for RNA-Composer [53] installed on an IBM Blue Gene/L supercomputer. Based on the Structural Classification of RNA (SCOR) database [55] initial models were modified. The coordinates of appropriate structures were taken from the Protein Data Bank (PDB) [56] and were further analysed using CCP4/QTMG [57,58]. Experimental data obtained from the literature and the PDB were used to validate the modeled structure. Finally, bond angles and lengths were corrected to produce biologically feasible conformations.

5.8. Molecular dynamics simulations for hLDLR-IRES (Emcv)-eGFP region

Each system was solvated with the explicit TIP3P water model within a water box of dimension 10 Å on each surface. Simulations were performed while using the Amber force field [58–60] with Na ions to neutralize the system charge. We minimize the system in two steps, first over the water and ion molecules holding domains fixed and, second, with all constraints removed. The minimization was performed using the Powell conjugate gradient algorithm. The initial equilibrations were achieved over 60 ps at constant temperature (300K) and pressure (1atm). Pressure was maintained at 1atm using the Langevin piston method, with a piston period of 100fs, damping constant of 50fs and piston temperature of 300K. Temperature coupling was enforced by velocity reassignment every 2ps. Both minimization and equilibration are performed using the NAMD program [61]. For the production run, we simulated a conventional MD trajectory for 100ns with the force field using the NAMD package. The system was simulated at constant temperature (300K) and volume using weakly coupled Langevin dynamics of non-hydrogen atoms, with a damping coefficient of $c = 10\ ps-1$ with a 2fs time step maintaining bonds to all hydrogen atoms rigid. Non-bonded interactions were truncated at 12 and 14 Å for van der Waals and electrostatic forces, respectively. Periodic boundary conditions were applied, and the particle mesh Ewald method was used to calculate electrostatic interactions. All simulations using the NAMD package were run on IBM Blue Gene/L supercomputer at the computational center for bioinformatics and genomics lab, Makkah.

6. Conclusion

We have presented a comparative analysis of protein expression patterns between different bicistronic cassettes containing variable length of ICS sequence. By using this approach to generate intervening ICS sequences of tailored size, we have overcome the previously repeatedly reported problem of insignificant second Cistron expression and have produced a functionally optimised 5'-hLDLR-IRES-eGFP-3′ expression cassette containing a 34bp spacer.

Author contributions

Conceived and designed the experiments: FA, ZA, MT, HM, MC, and SA. Analysed the data: AB, BB, SW, CC, MT, and WK. Wrote the paper: FA, and ZA.

Author disclosure statement

Authors have no conflict of interest.

Acknowledgments

This work was supported by a grant from Umm Al-Qura University, and partially by grant number (KACST AR-30-125), from the King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia. We would like to acknowledge of Dr. Frank Schnieders, Hamburg, Germany, for his kind gift of hLDLRcDNA and the 445bp downstream spacer sequences were derived from the pRSVh-LDLR plasmid.

References

[1] J.R. Holt, D.C. Johns, S. Wang, Z.Y. Chen, R.J. Dunn, E. Marban, et al., Functional expression of exogenous proteins in mammalian sensory hair cells infected with adenoviral vectors, J. Neurophysiol. 81 (4) (1999) 1881–1888 PubMed PMID: 10200223.

[2] A.T. Hoque, X. Liu, H. Kagami, W.D. Swaim, R.B. Wellner, B.C. O'Connell, et al., Construction and function of a recombinant adenovirus encoding a human

aquaporin 1-green fluorescent protein fusion product, Cancer Gene Ther. 7 (3) (2000) 476–485, https://doi.org/10.1038/sj.cgt.7700146 PubMed PMID: 10766354.
[3] W.J. Kollen, A.E. Mulberg, X. Wei, M. Sugita, V. Raghuram, J. Wang, et al., High-efficiency transfer of cystic fibrosis transmembrane conductance regulator cDNA into cystic fibrosis airway cells in culture using lactosylated polylysine as a vector, Hum. Gene Ther. 10 (4) (1999) 615–622, https://doi.org/10.1089/10430349950018689 PubMed PMID: 10094204.
[4] B.R. Cullen, P.T. Lomedico, G. Ju, Transcriptional interference in avian retroviruses–implications for the promoter insertion model of leukaemogenesis, Nature 307 (5948) (1984) 241–245 PubMed PMID: 6363938.
[5] M. Emerman, H.M. Temin, Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism, Cell 39 (3 Pt 2) (1984) 449–467 PubMed PMID: 6096005.
[6] M. Emerman, H.M. Temin, Quantitative analysis of gene suppression in integrated retrovirus vectors, Mol. Cell Biol. 6 (3) (1986) 792–800 PubMed PMID: 3022130; PubMed Central PMCID: PMCPMC367579.
[7] X. Zhu, I. Chung, M.R. O'Gorman, P.R. Scholl, Coexpression of normal and mutated CD40 ligand with deletion of a putative RNA lariat branchpoint sequence in X-linked hyper-IgM syndrome, Clin. Immunol. 99 (3) (2001) 334–339, https://doi.org/10.1006/clim.2001.5022 PubMed PMID: 11358428.
[8] M. Kozak, Adherence to the first-AUG rule when a second AUG codon follows closely upon the first, Proc. Natl. Acad. Sci. U. S. A. 92 (15) (1995) 7134 PubMed PMID: 7624384; PubMed Central PMCID: PMCPMC41486.
[9] J. Attal, M.C. Theron, L.M. Houdebine, The optimal use of IRES (internal ribosome entry site) in expression vectors, Genet Anal 15 (3–5) (1999) 161–165 PubMed PMID: 10596757.
[10] S. Lopez de Quinto, E. Martinez-Salas, Parameters influencing translational efficiency in aphthovirus IRES-based bicistronic expression vectors, Gene 217 (1–2) (1998) 51–56 PubMed PMID: 9795130.
[11] S. Lopez de Quinto, E. Martinez-Salas, Involvement of the aphthovirus RNA region located between the two functional AUGs in start codon selection, Virology 255 (2) (1999) 324–336 PubMed PMID: 10069958.
[12] J. Pelletier, N. Sonenberg, Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA, Nature 334 (6180) (1988) 320–325, https://doi.org/10.1038/334320a0 PubMed PMID: 2839775.
[13] B.L. Semler, M.L. Waterman, IRES-mediated pathways to polysomes: nuclear versus cytoplasmic routes, Trends Microbiol. 16 (1) (2008) 1–5, https://doi.org/10.1016/j.tim.2007.11.001 PubMed PMID: 18083033.
[14] X. Zuo, J. Wang, P. Yu, D. Eyler, H. Xu, M.R. Starich, et al., Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3' UTR of turnip crinkle virus, Proc. Natl. Acad. Sci. U. S. A. 107 (4) (2010) 1385–1390, https://doi.org/10.1073/pnas.0908140107 PubMed PMID: 20080629; PubMed Central PMCID: PMCPMC2803139.
[15] F. Al-Allaf, O. Tolmachov, M. Themis, C. Coutelle, Accurate size gauging of ExoIII/S1-generated deletions by PCR analysis of ligation mixtures, Anal. Biochem. 339 (2) (2005) 348–350, https://doi.org/10.1016/j.ab.2004.11.008 PubMed PMID: 15797577.
[16] A.M. Borman, P. Le Mercier, M. Girard, K.M. Kean, Comparison of picornaviral IRES-driven internal initiation of translation in cultured cells of different origins, Nucleic Acids Res. 25 (5) (1997) 925–932 PubMed PMID: 9023100; PubMed Central PMCID: PMCPMC146526.
[17] H.F. Gallardo, C. Tan, M. Sadelain, The internal ribosomal entry site of the encephalomyocarditis virus enables reliable coexpression of two transgenes in human primary T lymphocytes, Gene Ther. 4 (10) (1997) 1115–1119, https://doi.org/10.1038/sj.gt.3300506 PubMed PMID: 9415319.
[18] N. Ramesh, S.T. Kim, M.Q. Wei, M. Khalighi, W.R. Osborne, High-titer bicistronic retroviral vectors employing foot-and-mouth disease virus internal ribosome entry site, Nucleic Acids Res. 24 (14) (1996) 2697–2700 PubMed PMID: 8758998; PubMed Central PMCID: PMCPMC146003.
[19] J.C. Saiz, S. Lopez de Quinto, N. Ibarrola, F.X. Lopez-Labrador, J.M. Sanchez-Tapias, J. Rodes, et al., Internal initiation of translation efficiency in different hepatitis C genotypes isolated from interferon treated patients, Arch. Virol. 144 (2) (1999) 215–229 PubMed PMID: 10470249.
[20] A.M. Borman, J.L. Bailly, M. Girard, K.M. Kean, Picornavirus internal ribosome entry segments: comparison of translation efficiency and the requirements for optimal internal initiation of translation in vitro, Nucleic Acids Res. 23 (18) (1995) 3656–3663 PubMed PMID: 7478993; PubMed Central PMCID: PMCPMC307262.
[21] M. Azzouz, E. Martin-Rendon, R.D. Barber, K.A. Mitrophanous, E.E. Carter, J.B. Rohll, et al., Multicistronic lentiviral vector-mediated striatal gene transfer of aromatic L-amino acid decarboxylase, tyrosine hydroxylase, and GTP cyclohydrolase I induces sustained transgene expression, dopamine production, and functional improvement in a rat model of Parkinson's disease, J. Neurosci. 22 (23) (2002) 10302–10312 PubMed PMID: 12451130.
[22] R.A. Morgan, L. Couture, O. Elroy-Stein, J. Ragheb, B. Moss, W.F. Anderson, Retroviral vectors containing putative internal ribosome entry sites: development of a polycistronic gene transfer system and applications to human gene therapy, Nucleic Acids Res. 20 (6) (1992) 1293–1299 PubMed PMID: 1313966; PubMed Central PMCID: PMCPMC312173.
[23] M. Pizzato, E. Franchin, P. Calvi, R. Boschetto, M. Colombo, S. Ferrini, et al., Production and characterization of a bicistronic Moloney-based retroviral vector expressing human interleukin 2 and herpes simplex virus thymidine kinase for gene therapy of cancer, Gene Ther. 5 (7) (1998) 1003–1007, https://doi.org/10.1038/sj.gt.3300670 PubMed PMID: 9813672.
[24] B.P. Cormack, R.H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP), Gene 173 (1 Spec No) (1996) 33–38 PubMed PMID: 8707053.
[25] T.T. Yang, L. Cheng, S.R. Kain, Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein, Nucleic Acids Res. 24 (22) (1996) 4592–4593 PubMed PMID: 8948654; PubMed Central PMCID: PMCPMC146266.
[26] S.K. Jang, H.G. Krausslich, M.J. Nicklin, G.M. Duke, A.C. Palmenberg, E. Wimmer, A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation, J. Virol. 62 (8) (1988) 2636–2643 PubMed PMID: 2839690; PubMed Central PMCID: PMCPMC253694.
[27] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (7) (2012) 671–675 PubMed PMID: 22930834.
[28] S. Lopez de Quinto, E. Lafuente, E. Martinez-Salas, 9, IRES Interaction with Translation Initiation Factors: Functional Characterization of Novel RNA Contacts with eIF3, eIF4B, and eIF4GII. RNA vol. 7, (2001), pp. 1213–1226 PubMed PMID: 11565745; PubMed Central PMCID: PMCPMC1370167.
[29] V.G. Kolupaeva, T.V. Pestova, C.U. Hellen, I.N. Shatsky, Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA, J. Biol. Chem. 273 (29) (1998) 18599–18604 PubMed PMID: 9660832.
[30] O. Fernandez-Miragall, R. Ramos, J. Ramajo, E. Martinez-Salas, Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation, RNA 12 (2) (2006) 223–234, https://doi.org/10.1261/rna.2153206 PubMed PMID: 16373480; PubMed Central PMCID: PMCPMC1370902.
[31] E. Damoc, C.S. Fraser, M. Zhou, H. Videler, G.L. Mayeur, J.W. Hershey, et al., Structural characterization of the human eukaryotic initiation factor 3 protein complex by mass spectrometry, Mol. Cell. Proteomics 6 (7) (2007) 1135–1146, https://doi.org/10.1074/mcp.M600399-MCP200 PubMed PMID: 17322308.
[32] C.S. Fraser, J.Y. Lee, G.L. Mayeur, M. Bushell, J.A. Doudna, J.W. Hershey, The j-subunit of human translation initiation factor eIF3 is required for the stable binding of eIF3 and its subcomplexes to 40 S ribosomal subunits in vitro, J. Biol. Chem. 279 (10) (2004) 8946–8956, https://doi.org/10.1074/jbc.M312745200 PubMed PMID: 14688252.
[33] G.J. Belsham, Dual initiation sites of protein synthesis on foot-and-mouth disease virus RNA are selected following internal entry and scanning of ribosomes in vivo, EMBO J. 11 (3) (1992) 1105–1110 PubMed PMID: 1339342; PubMed Central PMCID: PMCPMC556552.
[34] H. Mizuguchi, Z. Xu, A. Ishii-Watabe, E. Uchida, T. Hayakawa, IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector, Mol. Ther. 1 (4) (2000) 376–382, https://doi.org/10.1006/mthe.2000.0050 PubMed PMID: 10933956.
[35] D.D. Anthony, W.C. Merrick, Eukaryotic initiation factor (eIF)-4F. Implications for a role in internal initiation of translation, J. Biol. Chem. 266 (16) (1991) 10218–10226 PubMed PMID: 2037575.
[36] M.K. Merrill, E.Y. Dobrikova, M. Gromeier, Cell-type-specific repression of internal ribosome entry site activity by double-stranded RNA-binding protein 76, J. Virol. 80 (7) (2006) 3147–3156, https://doi.org/10.1128/JVI.80.7.3147-3156.2006 PubMed PMID: 16537583; PubMed Central PMCID: PMCPMC1440377.
[37] J. Attal, M.C. Theron, C. Puissant, L.M. Houdebine, Effect of intercistronic length on internal ribosome entry site (IRES) efficiency in bicistronic mRNA, Gene Expr. 8 (5–6) (1999) 299–309.
[38] J. Wang, P. Cieplak, Q. Cai, M.J. Hsieh, J. Wang, Y. Duan, et al., Development of polarizable models for molecular mechanical calculations. 3. Polarizable water models conforming to Thole polarization screening schemes, J. Phys. Chem. B 116 (28) (2012) 7999–8008, https://doi.org/10.1021/jp212117d PubMed PMID: 22712654; PubMed Central PMCID: PMCPMC3493849.
[39] R.R. Gutell, J.C. Lee, J.J. Cannone, The accuracy of ribosomal RNA comparative structure models, Curr. Opin. Struct. Biol. 12 (3) (2002) 301–310 PubMed PMID: 12127448.
[40] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, Science 289 (5481) (2000) 905–920 PubMed PMID: 10937989.
[41] D.H. Mathews, M.D. Disney, J.L. Childs, S.J. Schroeder, M. Zuker, D.H. Turner, Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, Proc. Natl. Acad. Sci. U. S. A. 101 (19) (2004) 7287–7292, https://doi.org/10.1073/pnas.0401799101 PubMed PMID: 15123812; PubMed Central PMCID: PMCPMC409911.
[42] J.M. Diamond, D.H. Turner, D.H. Mathews, Thermodynamics of three-way multibranch loops in RNA, Biochemistry 40 (23) (2001) 6971–6981 PubMed PMID: 11389613.
[43] J. Burks, C. Zwieb, F. Muller, I. Wower, J. Wower, Comparative 3-D modeling of tmRNA, BMC Mol. Biol. 6 (2005) 14, https://doi.org/10.1186/1471-2199-6-14 PubMed PMID: 15958166; PubMed Central PMCID: PMCPMC1168896.
[44] M. Popenda, M. Szachniuk, M. Antczak, K.J. Purzycka, P. Lukasiak, N. Bartol, et al., Automated 3D structure composition for large RNAs, Nucleic Acids Res. 40 (14) (2012) e112, https://doi.org/10.1093/nar/gks339 PubMed PMID: 22539264; PubMed Central PMCID: PMCPMC3413140.
[45] N. Larsen, C. Zwieb, SRP-RNA sequence alignment and secondary structure, Nucleic Acids Res. 19 (2) (1991) 209–215 PubMed PMID: 1707519; PubMed Central PMCID: PMCPMC333582.
[46] R. Gosert, K.H. Chang, R. Rijnbrand, M. Yi, D.V. Sangar, S.M. Lemon, Transient expression of cellular polypyrimidine-tract binding protein stimulates cap-independent translation directed by both picornaviral and flaviviral internal ribosome entry sites in vivo, Mol. Cell Biol. 20 (5) (2000) 1583–1595 PubMed PMID: 10669736; PubMed Central PMCID: PMCPMC85342.
[47] A. Kaminski, M.T. Howell, R.J. Jackson, Initiation of encephalomyocarditis virus

RNA translation: the authentic initiation site is not selected by a scanning mechanism, EMBO J. 9 (11) (1990) 3753–3759 PubMed PMID: 2170120; PubMed Central PMCID: PMCPMC552132.

[48] D.R. Tolan, R.R. Traut, Protein topography of the 40 S ribosomal subunit from rabbit reticulocytes shown by cross-linking with 2-iminothiolane, J. Biol. Chem. 256 (19) (1981) 10129–10136 PubMed PMID: 7275971.

[49] A. Ben-Shem, N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova, M. Yusupov, The structure of the eukaryotic ribosome at 3.0 A resolution, Science 334 (6062) (2011) 1524–1529, https://doi.org/10.1126/science.1212642 PubMed PMID: 22096102.

[50] C. Sun, J. Querol-Audi, S.A. Mortimer, E. Arias-Palomo, J.A. Doudna, E. Nogales, et al., Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation, Nucleic Acids Res. 41 (15) (2013) 7512–7521, https://doi.org/10.1093/nar/gkt510 PubMed PMID: 23766293; PubMed Central PMCID: PMCPMC3753635.

[51] M. Zuker, P. Stiegler, Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, Nucleic Acids Res. 9 (1) (1981) 133–148 PubMed PMID: 6163133; PubMed Central PMCID: PMCPMC326673.

[52] Z.J. Lu, J.W. Gloor, D.H. Mathews, Improved RNA secondary structure prediction by maximizing expected pair accuracy, RNA 15 (10) (2009) 1805–1813, https://doi.org/10.1261/rna.1643609 PubMed PMID: 19703939; PubMed Central PMCID: PMCPMC2743040.

[53] G. Cichon, T. Willnow, S. Herwig, W. Uckert, P. Loser, H.H. Schmidt, et al., Non-physiological overexpression of the low density lipoprotein receptor (LDLr) gene in the liver induces pathological intracellular lipid and cholesterol storage, J. Gene Med. 6 (2) (2004) 166–175, https://doi.org/10.1002/jgm.473 PubMed PMID: 14978770.

[54] J. Heeren, D.S. Steinwaerder, F. Schnieders, G. Cichon, M. Strauss, U. Beisiegel, Nonphysiological overexpression of low-density lipoprotein receptors causes pathological intracellular lipid accumulation and the formation of cholesterol and cholesteryl ester crystals in vitro, J. Mol. Med. (Berl.) 77 (10) (1999) 735–743 PubMed PMID: 10606209.

[55] P.S. Klosterman, M. Tamura, S.R. Holbrook, S.E. Brenner, SCOR: a structural classification of RNA database, Nucleic Acids Res. 30 (1) (2002) 392–394 PubMed PMID: 11752346; PubMed Central PMCID: PMCPMC99131.

[56] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The protein Data Bank, Nucleic Acids Res. 28 (1) (2000) 235–242 PubMed PMID: 10592235; PubMed Central PMCID: PMCPMC102472.

[57] E. Potterton, S. McNicholas, E. Krissinel, K. Cowtan, M. Noble, The CCP4 molecular-graphics project, Acta Crystallogr D Biol Crystallogr 58 (Pt 11) (2002) 1955–1957 PubMed PMID: 12393928.

[58] L. Potterton, S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, et al., Developments in the CCP4 molecular-graphics project, Acta Crystallogr D Biol Crystallogr 60 (Pt 12 Pt 1) (2004) 2288–2294, https://doi.org/10.1107/S0907444904023716 PubMed PMID: 15572783.

[59] A.D. MacKerell Jr., N. Banavali, N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, 257-65, Biopolymers 56 (4) (2000), https://doi.org/10.1002/1097-0282(2000)56:4<257 PubMed PMID: 11754339.

[60] Q. Wang, X. Gong, Z. Gong, X. Ren, Z. Ren, S. Huang, et al., The mesenchymal stem cells derived from transgenic mice carrying human coagulation factor VIII can correct phenotype in hemophilia A mice, J Genet Genomics 40 (12) (2013) 617–628, https://doi.org/10.1016/j.jgg.2013.11.002 PubMed PMID: 24377868.

[61] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al., Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (16) (2005) 1781–1802, https://doi.org/10.1002/jcc.20289 PubMed PMID: 16222654; PubMed Central PMCID: PMCPMC2486339.