UTU Publishing Group

Bulletin of Biomedical Science

Journal Homepage: www.utu.com.sg

MINI-REVIEW

Unveiling the Regulatory Mechanisms of Small/Short Open Reading Frame (smORF)-Encoded Peptides in Liver Diseases

Jun Wang[1], Xiaoxue Yuan[2], Yun Wang[3], Yu Zhang[3], Hongping Lu[4], Yang Zhang[3], Jun Cheng[1, 3, 4]

[1]Peking University Ditan Teaching Hospital, Beijing 100015, China
[2]Municipal Center of Infectious Diseases of Beijing, Beijing 100015, China
[3]Liver Disease Center, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China
[4]Hebei Utu Pharmaceutical Com. Ltd, Shijiazhuang, Hebei Province 052165, China

ARTICLE INFO

Received 15 May 2023
Accepted 12 July 2023
Published 01 August 2023

KEY WORDS:
Small open-reading frame(s);
Upstream open-reading frame(s);
Small/short open-reading frame(s)-encoded peptide(s);
SEP(s);
Liver disease(s)

Abstract

From the discovery of upstream open-reading frame(s) (uORFs) and post-translational regulation to functional small/short open reading frames (smORFs)-encoded peptides (SEPs), mounting evidence suggests the necessity of updating genome databases to include small proteins derived from smORFs that are often overlooked. Functional studies across various species have demonstrated that SEPs can play pivotal regulatory roles in numerous basic biological activities. However, our understanding of the regulatory mechanisms underlying SEPs in human diseases remains limited. Given the escalating burden of liver diseases, including hepatitis B virus (HBV) infection, hepatitis C virus (HCV) infection, liver steatosis, liver fibrosis, and liver cancer, among others, it is imperative to investigate the mechanisms involved. The expression of downstream coding sequences is regulated by individual uORFs through complex secondary structures. Given the conservation and essential effects of SEPs across species, comprehending the regulation of SEPs holds significant potential for unraveling the mechanisms underlying diverse liver diseases. In this review, we provide an overview of recent research and primarily focus on the function and regulation of SEPs in liver diseases.

Abbrevations: uORF(s), upstream open-reading frame(s); smORF(s), small/short open reading frame(s); SEP(s), small/short open reading frames (smORFs)-encoded peptides; HBV, hepatitis B virus; HCV, hepatitis C virus; NASH, non-alcoholic steatohepatitis; aa, amino acid(s); SgrS, sugar transport small RNA; UTRs, untranslated regions; mORF, main ORF; main ORF; NMD, nonsense mediated decay; pgRNA, pregenomic RNA; ARF, alternative reading frame; FFA, free fatty acid; ER, endoplasmic reticulum; eIF2, eukaryotic initiation factor 2; ATF4, activating transcription factor 4; UPR, unfolded protein response; mTORC1, mammalian target of rapamycin complex 1; LIP, liver inhibitory protein; MTP, microsomal triglyceride transfer protein; MOTS-c, mitochondrial open-reading- frame of the twelve S rRNA –c; PPP1R15A, protein phosphatase 1 regulatory subunit 15A; mEH, microsomal epoxide hydrolase; FH, familial hypercholesterolemia; HJV, hemojuvelin; ASNSD1, asparagine synthetase domain- containing 1; LAP, liver activating protein; PRL, phosphatase of regenerating liver

Corresponding author: Jun Cheng, Hebei Utu Pharmaceutical Com. Ltd, Shijiazhuang, Hebei Province 052165, China. chengj0817@sina.cn

Introduction

Within every eukaryotic genome, there exist tens of thousands to millions of small/short open reading frame (smORF) DNA sequences, encoding less than 100 amino acids (aa), surpassing the number of annotated genes. For instance, the mouse genome is estimated to contain a staggering 40,700,000 smORFs (Couso, 2015). Recent omics analyses have revealed the translation of many previously unannotated smORFs, leading to the application of the term "dark matter" to these newly discovered RNAs and proteins (Baboo & Cook, 2014). Moreover, recent research has unveiled a class of small peptides directly encoded by smORFs, distinct from classical peptides that are initially translated into larger precursor proteins and undergo limited proteolytic processing. These smORF-encoded polypeptides are referred to as Small ORF-encoded Peptides (SEPs) (Chu et al., 2015).

SEPs have been identified in diverse organisms, ranging from viruses, yeasts, worms, flies, to humans (Chu et al., 2015; Plaza et al., 2017). The earliest studies on smORFs date back to 1989 when investigation into the class II restriction-modification system of Serratia marcescens revealed a smORF of 252 bp within the gene encoding an endonuclease, where the smORF's translation initiation codon was GUG. However, despite attempts to detect the smORF-encoded protein in phosphocellulose column eluates of S. marcescens or E. coli extracts transformed with relevant plasmids, it remained elusive (Heidmann et al., 1989). Subsequently, the discovery of the first SEP came during research on Escherichia coli to understand the role of non-coding RNA, specifically the sugar transport small RNA (SgrS). It was found that the 5'-untranslated regions (UTRs) of SgrS encoded a SEP, a 43-aa peptide called SgrT, which played a crucial role in cell metabolism by directly binding and inhibiting SgrS to regulate glucose influx (Chu et al., 2015). Human genomes harbor hundreds of smORFs, and it has been reported that smORFs and SEPs may regulate fundamental biological processes, such as pathogen defense, innate immunity, cell communication, and signal transduction, due to their small size (Frith et al., 2006).

Dysregulation of SEPs can contribute to disease phenotypes, and many SEPs have emerged as potential therapeutic targets for various conditions, including heart failure, obesity, diabetes, and cancer (Hassel et al., 2023). However, liver diseases and their complications, recognized as major causes of mortality, demand greater attention in this field. Liver diseases, including hepatitis B virus (HBV) infection, hepatitis C virus (HCV) infection, and non-alcoholic steatohepatitis (NASH), can progress to chronic liver cirrhosis and liver cancer. The absence of definitive therapies for most liver diseases places a heavy burden on patients. Hence, in this review, we present recent findings on smORFs and SEPs, highlighting their unexpected significance and ongoing efforts to unravel their potential regulation in different liver diseases. This exploration aims to provide novel insights into the underlying pathological processes and identify potential drug targets in complex liver diseases.

Characteristics and Regulation of SEPs

1. Distinct Features of SEPs

The analysis of discovered SEPs has revealed several distinguishing characteristics. Proteomics studies have shown that the first downstream in-frame stop codons in smORFs encoding SEPs serve as termination sites. The start codon can be the closest upstream in-frame AUG or a near-cognate non-AUG codon (e.g., CUG) within a Kozak-consensus sequence. In cases where neither of these conditions is met, the codon immediately following an upstream stop codon is designated as the start site. SEPs typically range in length from 8 to 149 amino acids, with the majority being less than 100 amino acids long (Chu et al., 2015).

2. Diversity of smORF Locations and Types

It has been observed that the majority of RNA transcripts containing smORFs encoding SEPs are not derived from the RefSeq database, indicating their presence in unannotated regions of the genome. SEPs have been identified in various types of transcripts, including mRNAs, pre-mRNAs, long non-coding RNAs (lncRNAs), miRNAs, and rRNAs. Additionally, smORFs can be found in upstream (5'-UTR), downstream (3'-UTR), or within the coding sequence of main ORFs, often referred to as uORFs, dORFs, or CDS-ORFs, respectively (Plaza et al., 2017). However, the discovery of uORFs encoding conserved peptides in humans remains limited (Dever et al., 2020).

3. Mechanisms of smORF and SEP Function

Post-transcriptional regulation, particularly through uORFs, plays a crucial role in gene expression control. uORFs are among the most common regulatory RNA elements in mammalian transcripts (Somers et al., 2013). The translation of uORFs can repress the downstream coding sequence through various mechanisms, including ribosomal subunit dissociation, ribosome stalling, nonsense-mediated decay (NMD) induction, and ribosome stalling with dissociation. Conversely, uORF translation can facilitate main ORF translation through reinitiation or ribosomal leaky scanning (Barbosa et al., 2013; Silva et al., 2019). The regulation of uORF-mediated translation repression is influenced by various features, such as the distance between the 5'-cap and the uORF, the AUG context, uORF length and secondary structure, conservation across species, the number and location of uORFs, and the length

of intercistronic regions (Barbosa et al., 2013). Furthermore, studying the potential associations between SEPs and other regulators involved in the same phenotype offers an effective approach to understanding SEP function. Investigating the identified binding proteins of SEPs can provide insights into their functions based on the known impact of the interaction partners (Chu et al., 2015).

4. Regulation of uORFs in Hepatitis B Virus Infection

The genome of the hepatitis B virus (HBV) encodes four major overlapping coding regions, including the core, surface, polymerase, and X genes. The presence of an uncharacterized short uORF, termed C0 ORF, conserved in all HBV subtypes has been identified. The C0 ORF negatively regulates core protein translation and promotes reinitiation of polymerase protein synthesis under optimal conditions. It serves as a determinant of the balance between core and polymerase protein synthesis (Chen et al., 2005; Zong et al., 2017). Similar short uORFs have also been identified in the related duck hepatitis B virus (DHBV), where ribosomes translating the P protein are shunted from the donor site near the 5'-UTR of the pregenomic RNA to receptor sites near the P AUG, facilitating initiation of P protein synthesis (Sen et al., 2004).

5. Regulation of smORFs in Hepatitis C Virus Infection

Hepatitis C virus (HCV) infection induces various changes in cellular gene expression. Studies have demonstrated the translation of uORFs in stress-responsive genes, such as C/EBP homologous protein (CHOP) and protein phosphatase 1 regulatory subunit 15A (PPP1R15A), which are involved in endoplasmic reticulum (ER) stress and HCV replication. These translated uORFs escape translational downregulation and play roles in ER stress and hepatocellular carcinoma (HCC) progression (Gerresheim et al., 2019). The HCV core ORF region can also produce an alternative reading frame (ARF)/core+1 protein, distinct from the canonical polyprotein ORF. The ARF/core+1 protein, consisting of at least 99 amino acids, has been detected in patients with chronic HCV infection, liver cirrhosis, and liver cancer, suggesting its potential involvement in HCV pathogenesis (Niepmann & Gerresheim, 2020).

6. Regulation of uORFs and SEPs in Liver Steatosis

The translation regulation of uORFs plays a rapid and effective role in coping with various types of stress. In non-alcoholic steatohepatitis (NASH), which is characterized by inflammation and cell death induced by free fatty acids (FFAs), the uORFs of stress response genes, such as CHOP, selectively promote their translation by bypassing inhibitory uORFs, leading to the upregulation of genes associated with ER stress and NASH progression (Willy et al., 2015). The eukaryotic initiation factor 6 (eIF6) has also been implicated in the regulation of translation for adipogenic transcription factors involved in lipid metabolism. Its activation leads to enhanced levels of lipogenic and glycolytic enzymes (Brina et al., 2015). Moreover, uORFs have been found to regulate the translation efficiency and protein expression of microsomal triglyceride transfer protein (MTP), an essential component in the assembly of triglyceride-rich lipoproteins (Suzuki et al., 2016). Additionally, a mitochondrial SEP named MOTS-c has been identified, which regulates insulin sensitivity and metabolic homeostasis, offering potential insights into NASH pathogenesis (Lee et al., 2015).

7. Regulation of uORFs and SEPs in Drug-Induced Liver Injury

The induction of detoxification pathways is a significant mechanism for liver chemoprevention. Microsomal epoxide hydrolase (mEH), a critical xenobiotic-metabolizing enzyme, is regulated by uORFs present in its mRNA. These uORFs inhibit mEH protein translation both in cis and in trans configurations, highlighting their role in regulating the detoxification process (Nguyen et al., 2013). Furthermore, drug-induced endoplasmic reticulum (ER) stress can lead to liver injury. The activation of the unfolded protein response (UPR) pathway, initiated by the effector PKR-like ER kinase (PERK), involves the translation of ATF4 mRNA, which contains specific uORFs. This translation induction leads to the synthesis of chaperones, proteins involved in autophagy, protein secretion, and amino acid metabolism, and contributes to the adaptive response to ER stress (Foufelle & Fromenty, 2016).

8. Regulation of uORFs and SEPs in Liver Genetic Diseases and Cancer

Mutations in uORFs of proto-oncogenes and tumor suppressor genes have been associated with malignancies. The balance between CCAAT/enhancer-binding protein α (C/EBPα) and C/EBPβ isoforms, regulated by mTOR kinase signaling and uORF translation, has been implicated in cell differentiation and cancer progression (Wethmar et al., 2010). Exonic cancer mutations can affect uORF start codons, disrupting their regulatory role in genes such as MYC, BCL-2, tumor protein p53, phosphatase and tensin homolog, mutS homolog 5 (MSH5), and others (McGillivray et al., 2018). Additionally, the uORFs present in the longer transcript of the oncogene MDM2 dampen its translation (Brown et al., 1999). The magnesium-regulated phosphatase of regenerating liver (PRL) has been found to be posttranscriptionally regulated through a uORF, suggesting its potential involvement in liver cancer pathogenesis (Hardy et al., 2019). Recent studies have identified numerous SEPs in liver cancer cells, providing further insights into the role of SEPs in liver cancer biology

(Wang et al., 2021).

9. Regulation of uORFs and SEPs in Liver Fibrosis

While limited information is available regarding the function and regulation of SEPs in liver fibrosis, a 96 amino acid cytoplasmic peptide named ASDURF (ASNSD1-SEP) has been identified from the uORF of asparagine synthetase domain-containing 1 (ASNSD1). The conservation of this SEP across mammals suggests its potential involvement in liver fibrosis (Chu et al., 2015; Slavoff et al., 2013).

Conclusions and Prospects

The characterization and regulatory mechanisms of small open reading frames (smORFs) and small peptides (SEPs) in liver diseases represent an exciting and rapidly evolving field of research. These SEPs and smORFs play crucial roles in the pathogenesis of various liver diseases, offering potential avenues for the development of novel therapeutic strategies. The presence of a vast reservoir of smORFs and the conservation of SEPs within human genes emphasize their essential regulatory functions in liver diseases. However, it is important to acknowledge that our current knowledge of SEPs is limited by the methodologies employed. Therefore, there is a high likelihood that many more SEPs remain undiscovered. Advancements in detection techniques are crucial to uncovering additional SEPs and gaining a deeper understanding of their roles in diverse biological processes.

Systematic screening of smORFs holds great promise for discovering disease-associated SEPs and facilitating the development of innovative therapeutic approaches. By exploring the functional and regulatory properties of these SEPs, new targets can be identified and harnessed to modulate disease pathways effectively. Furthermore, the exploration of SEP-protein interactions emerges as a valuable strategy for targeting pathways involved in liver diseases. Leveraging the knowledge gained from these interactions can guide the development of targeted interventions. However, it is essential to note that our understanding of SEP functions from unannotated genomes is still limited, warranting further investigation.

In conclusion, the study of SEPs and smORFs in liver diseases represents a frontier in biomedical research. Continued exploration of the regulatory landscape of upstream open reading frames (uORFs) and SEPs in liver diseases will enhance our understanding of their functional significance and provide novel therapeutic opportunities. By improving detection strategies and uncovering the hidden potential of SEPs, we can unlock new insights into liver disease pathogenesis and pave the way for innovative treatments that target these enigmatic regulatory elements.

References

[1] Baboo S, Cook PR. “Dark matter” worlds of unstable RNA and protein. Nucleus. 2014;5(4):281–6.

[2] Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 2013;9(8):e1003529.

[3] Benito-Vicente A, Uribe K, Jebari S, Galicia-Garcia U, Ostolaza H, Martin C. Familial hypercholesterolemia: the most frequent cholesterol metabolism disorder caused disease. Int J Mol Ssc. 2018;19(11):3426.

[4] Brina D, Miluzio A, Ricciardi S, Clarke K, Davidsen PK, Viero G, Tebaldi T, Offenhäuser N, Rozman J, Rathkolb B, Neschen S, Klingenspor M, Wolf E, Gailus-Durner V, Fuchs H, Hrabe De Angelis M, Quattrone A, Falciani F, Biffo S. eIF6 coordinates insulin sensitivity and lipid metabolism by coupling translation to transcription. Nat Commun. 2015;6:8261.

[5] Brown CY, Mize GJ, Pineda M, George DL, Morris DR. Role of two upstream open reading frames in the translational control of oncogene mdm2. Oncogene. 1999;18(41):5631–7.

[6] Chen A. Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons. Nucleic Acids Res. 2005;33(4):1169–81.

[7] Chen A. Prospects for inhibiting the post-transcriptional regulation of gene expression in hepatitis B virus. World J Gastrenterol. 2014; 20(25):7993-8004.

[8] Chu Q, Ma J, Saghatelian A. Identification and characterization of sORF-encoded polypeptides. Cri Reviews Biochem Mol Biol. 2015;50(2):134–41.

[9] Couso JP. Finding smORFs: getting closer. Genome Biol. 2015;16(1):189.

[10] Dever TE, Ivanov IP, Sachs MS. Conserved upstream open reading frame nascent peptides that control translation. Annu Rev Genet. 2020;54:237–64.

[11] Foufelle F, Fromenty B. Role of endoplasmic reticulum stress in drug-induced toxicity. Pharmacol Res Perspect. 2016; 4(1):e00211.

[12] Frith MC, Forrest AR, Nourbakhsh E, Pang KC, Kai C, Kawai J, Carninci P, Hayashizaki Y, Bailey TL, Grimmond SM. The abundance of short proteins in the mammalian proteome. PLoS Genet. 2006;2(4):e52.

[13] Gerresheim G, Bathke J, Michel A, Andreev D, Shalamova L, Rossbach O, Hu P, Glebe D, Fricke M, Marz M, Goesmann A, Kiniry S, Baranov P, Shatsky I, Niepmann M. Cellular gene expression during hepatitis C virus replication as revealed by ribosome profiling. Int J Mol Sci. 2019;20(6):1321.

[14] Hardy S, Kostantin E, Wang SJ, Hristova T, Galicia-Vázquez G, Baranov PV, Pelletier J, Tremblay ML. Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism. Proc Natl Acad Sci USA. 2019;116(8):2925–34.

[15] Hassel KR, Brito-Estrada O, Makarewich CA. Microproteins: overlooked regulators of physiology and disease. iScience. 2023;26(6):106781.

[16] Heidmann S, Seifert W, Kessler C, Domdey H. Cloning, characterization and heterologous expression of the Sma I restriction-modification system. Nucl Acids Res. 1989;17(23):9783–96.

[17] Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim S-J, Mehta H, Hevener AL, de Cabo R, Cohen P. The

Mitochondrial- derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metabolism. 2015;21(3):443–54.

[18] McGillivray P, Ault R, Pawashe M, Kitchen R, Balasubramanian S, Gerstein M. A comprehensive catalogue of predicted functional upstream open reading frames in humans. Nucleic Acids Res. 2018;46(7):3326–38.

[19] Nguyen HL, Yang X, Omiecinski CJ. Expression of a novel mRNA transcript for human microsomal epoxide hydrolase (EPHX1) is regulated by short open reading frames within its 5′-untranslated region. RNA. 2013;19(6):752–66.

[20] Niepmann M, Gerresheim GK. Hepatitis C virus translation regulation. Int J Mol Sci. 2020;21:2328.

[21] Onofre C, Tomé F, Barbosa C, Silva AL, Romão L. Expression of human hemojuvelin (HJV) Is tightly regulated by two upstream open reading frames in HJV mRNA that respond to iron overload in hepatic cells. Mol Cell Biol. 2015;35(8):1376–89.

[22] Plaza S, Menschaert G, Payre F. In search of lost small peptides. Annu Rev Cell Dev Biol. 2017;33:391–416.

[23] Sen N, Cao F, Tavis JE. Translation of duck hepatitis B virus reverse transcriptase by ribosomal shunting. J Virol. 2004;78(21):11751–7.

[24] Silva J, Fernandes R, Romão L. Translational regulation by upstream open reading frames and human diseases. Adv Exp Med Biol. 2019;1157:99-116.

[25] Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol. 2013;9(1):59–64.

[26] Somers J, Pöyry T, Willis AE. A perspective on mammalian upstream open reading frame function. Int J Biochem Cell Biol. 2013;45(8):1690–700.

[27] Sözen MM, Whittall R, Öner C, Tokatlı A, Kalkanoğlu HS, Dursun A,Coşkun T, Öner R, Humphries SE. The molecular basis of familial hypercholesterolaemia in Turkish patients. Atherosclerosis. 2005;180(1):63–71.

[28] Suzuki T, Brown JJ, Swift LL. Identification of a novel transcript and regulatory mechanism for microsomal triglyceride transfer protein. Buratti E. PLoS ONE. 2016;11(1):e0147252.

[29] Wang B, Hao J, Pan N, Wang Z, Chen Y, Wan C. Identification and analysis of small proteins and short open reading frame encoded peptides in Hep3B cell. J Proteomics. 2021;230:103965.

[30] Wethmar K, Smink JJ, Leutz A. Upstream open reading frames: molecular switches in (patho)physiology. Bioessays. 2010;32(10):885–93.

[31] Willy JA, Young SK, Stevens JL, Masuoka HC, Wek RC. CHOP links endoplasmic reticulum stress to NF-κB activation in the pathogenesis of nonalcoholic steatohepatitis. Mol Biol Cell. 2015;26(12):2190–204.

[32] Zidek LM, Ackermann T, Hartleben G, Eichwald S, Kortman G, Kiehntopf M, Leutz A, Sonenberg N, Wang Z, Von Maltzahn J, Müller C, Calkhoven CF. Deficiency in MTORC 1-controlled C/EBPβ-mRNA translation improves metabolic health in mice. EMBO Reports. 2015;16(8):1022–36.

[33] Zong L, Qin Y, Jia H, Ye L, Wang Y, Zhang J, Wands JR, Tong S, Li J. Differential regulation of hepatitis B virus core protein expression and genome replication by a small upstream open reading frame and naturally occurring mutations in the precore region. Virology. 2017;505:155–61.