Rafaela Lacerda Santos

Mestre

Non-canonical translation initiation of proteins with potential relevance in colorectal cancer

Dissertação para obtenção do Grau de Doutor
em Biologia, Especialidade Genética Molecular

Orientador: Doutora Luísa Romão, Investigadora Principal com Habilitação, Instituto Nacional de Saúde Doutor Ricardo Jorge

Co-orientador: Doutor Pedro Viana Baptista, Professor Associado com Agregação, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri:

Presidente: Prof. Doutor Jorge Joaquim Pamies Teixeira

Vogais: Prof. Doutora Isabel Maria Godinho de Sá Nogueira
Prof. Doutor Manuel António da Silva Santos
Prof. Doutor Paulo Henriques Carasquinho de Matos
Prof. Doutora Maria Alexandra Marques Moreira
Prof. Doutora Luísa Maria Ferreira Romão Loison

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro, 2016

Non-canonical translation initiation of proteins with potential relevance in colorectal cancer

Polonius: What do you read, my lord?

Hamlet: Words, words, words.

in Hamlet*, by William Shakespeare*

Agradecimentos

E cinco anos passados, eis que tudo se resume a isto: palavras, palavras, palavras. Nunca estas palavras poderão traduzir – já que esta é uma tese que fala de tradução – tudo o que se passou na minha vida, e tudo aquilo por que a minha vida passou, nos últimos anos. Fazer um doutoramento é como traduzir uma proteína dentro de uma célula. Analisemos este percurso como se de um RNA mensageiro se tratasse. Tudo começa na extremidade da região 5' não-traduzida. O ribossoma chega, olha, reconhece o *cap*, diz que é por ali. Foi o que lhe ensinaram, foi o caminho que aprendeu. E pensa que tudo vai ser como lhe ensinaram: percorre toda a região não-traduzida (tem algumas estruturas secundárias para desfazer, mas nada com que não esteja a contar; disseram-lhe que iria ser assim, ele sabia que iria ser assim), chega ao codão de iniciação – AUG, é aqui, só pode ser aqui – e começa a traduzir. Uma metionina – é sempre o primeiro aminoácido, não há que enganar, não há que duvidar – depois outro, e mais outro, e outros tantos até encontrar um codão de terminação, pára, o péptido é libertado e vai cumprir a sua missão. Parece simples, parece lógico, não parece muito fácil, mas tudo estava previsto. É trabalhoso, mas tudo estava previsto. Todavia, na vida, tal como na célula (como se a célula não fosse vida), nem sempre a mensagem está ali à nossa frente, tão clara, tão nítida, que seja só fazer o que nos ensinaram e continuar como se não mais tivera passado. Às vezes, não há *cap*. E agora? O ribossoma tem de «entrar à bruta» no meio de um RNA mensageiro que não conhece. Acha que o AUG é aqui, mas ali há outro, e acolá mais um, e mais à frente ainda surge um outro. Qual é o correcto? Experimenta este, não traduz o que quer; experimenta o outro, não é bem isto, mas já é um começo. E, de repente, uma estrutura secundária grande, assustadora, intricada, impossível de desfazer. E agora? Lá ao fundo já vêm outros ribossomas, não pode ficar ali parado. E então decide-se: salta para outro AUG, vai tentar a sua sorte mais à frente. Pode ser que funcione. Já avista um codão de terminação. As coisas estão a correr bem, demasiado bem. Um *stop* tão cedo? É prematuro. E então percebe que o que fez até ali tem de ir para o lixo. É preferível assim: destruir o que já está feito, mas prevenir o erro. E tem de começar tudo outra vez. Agora já conhece algumas alternativas, agora já vai olhar para a mensagem que tem de traduzir de outra maneira: vai antecipar as estruturas secundárias, vai ter tempo para decidir se quer desfazê-las ou passar-lhes ao lado. Às vezes, simplesmente já não há mais ATP para gastar e ainda é preciso seguir em frente. E, ao fim, depois de muitas idas e vindas, depois de muitos «desfazer-para-voltar-a-fazer», depois de já não saber se ainda é um ribossoma ou apenas duas subunidades que já não se conseguem juntar novamente, vê a sua mensagem traduzida. E vê o seu péptido ganhar uma função e transformar-se numa proteína que vai por aí até encontrar um sítio onde seja útil, até encontrar uma célula onde possa ter alguma serventia. Até ser mandada para o lixo, porque já não serve para nada. Porque apareceu uma nova proteína que funciona melhor. E o ribossoma olha para a sua proteína, aquela que acabou de produzir, e pensa que, se calhar, ficou alguma estrutura secundária por desfazer, que, se calhar, não devia ter começado

naquele AUG, que, se calhar… O trabalho está feito. Outros ribossomas virão, outras proteínas surgirão, mas este ribossoma ainda pode funcionar. Ainda pode traduzir. Há tantas mensagens que ainda falta decifrar. Há tantas proteínas que ainda falta pôr a funcionar. Este ribossoma não desiste. Mesmo quando tem as suas subunidades separadas, mesmo quando acha que não as vai conseguir juntar novamente. Sempre aparece um factor de iniciação que lhe dá uma mãozinha, sempre surge um IRES que lhe estende o braço. E é a esses factores de iniciação e a esses IRES que este ribossoma quer agradecer o ter conseguido, apesar de todas as estruturas secundárias, manter unidas as suas subunidades.

Agradeço às instituições que permitiram que este trabalho fosse realizado: Fundação para a Ciência e a Tecnologia, pela bolsa de doutoramento; Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT-UNL), por me ter aceitado como aluna; Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA), por me ter permitido desenvolver este trabalho nas suas instalações; BioISI, pelo apoio financeiro prestado para ajuda a deslocações a congressos; Alfagene pelo apoio financeiro para ajuda a deslocações a congressos, também. Claro que as instituições são feitas por pessoas e, como tal, é a elas que devo estes agradecimentos.

Agradeço à Glória Isidro, coordenadora do Departamento de Génica Humana (DGH) do INSA pela sua celeridade e pragmatismo na resolução de problemas, pelo apoio logístico sempre prestado e pela boa-disposição sempre presente.

Ao João Lavinha, investigado principal e coordenador da Unidade de Investigação & Desenvolvimento do DGH, pelas estimulantes conversas de cariz científico, filosófico ou político que fomos mantendo ao longo destes anos.

Aos Professores Isabel Sá-Nogueira (coordenadora do Programa Doutoral em Biologia) e Pedro Viana Baptista (professor do Departamento de Ciências da Vida e meu co-orientador), da FCT-UNL, por todo o apoio prestado e ajuda na resolução de todos os problemas logísticos e burocráticos que foram surgindo.

À Luísa Romão, minha orientadora, agradeço por me ter recebido no seu laboratório e ter supervisionado a progressão do meu trabalho ao longo destes últimos anos. Seria motivo suficiente de agradecimento, pois todos precisamos de um orientador que nos acolha e nos providencie as condições adequadas à realização do nosso trabalho. No entanto, esta relação foi além disto e, entre risos e puxões-de-orelhas, agradeço-lhe por nunca ter desistido de mim, por tudo o que me ensinou e me incentivou a aprender sozinha. Quando a mãe-pata tem muitos patinhos de quem cuidar, alguns, às vezes, acham que ficam para trás, porque ela não os leva ao colo para atravessar o rio. Parece difícil, parece, ao patinho, que a mãe-pata não está a fazer tudo o que podia para ele chegar à outra margem. Mas o que o patinho não sabe é que a mãe-pata está a fazer a melhor coisa que poderia fazer – ensiná-lo a nadar sozinho e a chegar à outra margem pela sua própria membrana interdigital. E é este o maior agradecimento que eu tenho para fazer à minha mãe-pata laboratorial – o ter-me ensinado a usar as minhas membranas e o caminho para a outra margem, e não me ter levado ao colo. Por ter lá estado

nos bons momentos, mas por não ter abandonado o barco quando tudo faria prever que se afundasse. E quero fazer-lhe um agradecimento especial pela dedicação e força que me deu na recta final para acabar tudo e concluir esta fase tão peculiar da minha vida.

Gostaria, também, de agradecer à Margarida Gama-Carvalho e ao Peter Jordan, membros da minha comissão de acompanhamento de tese, por todas as proveitosas discussões e sugestões para o desenrolar deste trabalho. À Margarida agradeço, ainda, por me ter dado a conhecer o fantástico mundo das bases de dados de RNA e proteína e que tão úteis foram para desenvolver esta investigação.

Ao Alexandre Teixeira e à Ana Ramos agradeço por me terem ensinado a dar os primeiros passos no laboratório até conseguir começar a correr sozinha. Ao Alexandre, estou grata, também, por todo o incentivo que me deu durante o tempo que convivemos no laboratório e por todas as conversas e desconversas que fomos tendo e que vieram iluminar um caminho, por vezes, tão sombrio.

Ao Marco Candeias agradeço as discussões de resultados e o optimismo pueril com que encara cada experiência. Mostrou-me que qualquer pequenino resultado positivo, mesmo mergulhado num imenso oceano de resultados negativos, pode ser motivo de regozijo e um incentivo à perseverança.

Manifesto, aqui, a minha incomensurável gratidão à Juliane Menezes pela ajuda que me deu a concluir experiências, discutir resultados e pensar novas experiências. Por todas as horas e canetas coloridas que gastou a discutir resultados comigo e os próximos passos a dar. Por me ter arrumado a bancada e os papéis vezes sem conta e impedido que o meu espaço «virasse o samba do criolo doido». Só *isto* bastaria para ela ter um lugar de relevo nesta página, para ela ter direito a todas as minhas palavras de agradecimento. No entanto, no caso da Juliane, e correndo o risco de utilizar um bacoco lugar-comum, não há palavras (ou melhor, há-as, mas não caberiam todas aqui) para descrever o quão fundamental ela foi para que este trabalho fosse levado até ao fim. E, apesar de toda a nossa colaboração científica ter sido fundamental para que este trabalho se materializasse, o que eu sinto mesmo necessidade de lhe agradecer é o ela ter sido – o ela *ser* – a pessoa que esteve sempre comigo em todos os momentos bons, em todos os momentos maus, em todos os momentos em que foi preciso sacudir-me pelos ombros e obrigar-me a ver o lado positivo, em todos os momentos em que foi preciso levantar-me quando tudo estava a empurrar-me para baixo, em todos os momentos em que perdi o foco e ela foi lá limpar-me os óculos com toda a veemência. Em todos os momentos. E fê-lo sem pedir nada em troca, sem cobrar honorários. E isso fez-me voltar a acreditar que há pessoas que podem fazer coisas pelos outros sem esperar um retorno, sem querer uma recompensa, simplesmente porque gostam e se preocupam e querem ver os outros bem e felizes. E eu espero que continue a ser assim o resto da vida – que possamos fazer tudo uma pela outra sem esperar nada em troca, simplesmente porque nos queremos ver felizes e queremos ficar felizes com a felicidade da outra. Simplesmente porque somos amigas com um grande «A» maiúsculo. E assim seremos até sermos velhinhas e senis. Henry Ford

disse: «Coming together is a beginning, keeping together is progress, working together is success.» Em relação à Juliane, eu acrescento: *becoming friends is bliss*.

Há outra pessoa a quem eu tenho de (não tenho, mas quero muito fazê-lo) agradecer. Chama-se Francesca Brito e apareceu no laboratório numa altura em que o ambiente estava bipolarizado e eu quase obrigada a escolher um dos pólos quando o que queria era ficar no equador. Veio com a inocência própria de quem não sabe o que se passa e trouxe-me uma golfada de ar fresco para conseguir voltar a respirar. Dar-lhe uma mãozinha para acabar a sua tese de mestrado fez-me voltar a ter a certeza de que, afinal, ainda havia coisas pelas quais lutar e pelas quais valia a pena voltar a sorrir. E é sobretudo pela estabilidade que me trouxe e por me ter obrigado a acreditar em mim e na minha capacidade de fazer as coisas bem que eu tanto tenho a agradecer-lhe. E que assim seja até a Alzheimer nos atacar e nos esquecermos uma da outra. E, nessa altura, como ela costuma dizer, «conhecermo-nos novamente e ficarmos amigas outra vez».

À Cláudia Onofre e ao Paulo Jorge, agradeço o apoio técnico e a discussão científica, respectivamente, e *vice-versa*, em algumas situações. À Cláudia, agradeço, principalmente, o pragmatismo e a eficiência (o chamado «desenrascanço») com que sempre me ajudou e que tão úteis se mostraram quando a prática resolveu não obedecer à teoria e, antes, vogar à sua vontade. Ao Paulo, agradeço, sobretudo, a divagação e o idealismo (o chamado «enrascanço»), e o fervor dos pensamentos soltos em catadupa com que sempre discutiu a Ciência que fomos fazendo e que tão proveitosos se revelaram quando a teoria decide não justificar a prática e ficar aquém da sua verdadeira utilidade. Aos dois, agradeço, também, as conversas e as gargalhadas, o apoio e a amizade que me deram durante estes anos, a troco de nada, e que, espero, fiquem para o resto das nossas vidas. Sempre a troco de nada.

Aos colegas que passaram pelo laboratório durante o tempo em que lá estive – Cristina Barbosa, Isabel Peixeiro, Ana Morgado, Bruno Silva, Gerson Asper, Cláudia Estima, entre outros – agradeço o apoio, técnico e/ou moral, que me foram dando ao longo do tempo. À Cláudia, agradeço, ainda, a ajuda a concluir algumas experiências relativas ao transcrito *MLH1*. Ao Bruno, agradeço as dicas que me foi dando para conseguir «acordar do pesadelo» que foi clonar as regiões 5' não-traduzidas dos transcritos *UPF1* e *AGO1*.

A todas as meninas (sempre só meninas) do laboratório de Oncobiologia, agradeço por todo o apoio técnico e reagentes que nos foram facultando. Em especial, à Patrícia Barros, por tudo o que me ensinou a fazer e, principalmente, por tudo o que me ensinou a não fazer. E por todas as nossas conversas na sala de cultura de tecidos, por vezes sérias e profundas, mas, outras tantas, tão deliciosamente ridículas que nos dão vontade de ficar a trabalhar até tarde, mesmo quando o corpo não aguenta mais.

Ao Paulo Matos, agradeço os conselhos e sugestões sempre tão eficazes e objectivos que me permitiram vencer alguns obstáculos.

Ao meu muito querido José Furtado, além de agradecer os conselhos e os ensinamentos sobre a cultura de tecidos, quero agradecer a boa-disposição, a gargalhada que ecoa pelos

corredores do DGH e que nos anima logo pela manhã, quando ainda não chegou mais ninguém, o «serviço de psicologia» sempre disponível para me dar a mão e ajudar a levantar do chão e mostrar que para a frente é que é o caminho, o nunca desistir e o mostrar-me que é assim que tem de ser.

Ao José Ferrão, agradeço todo o apoio técnico prestado que atenuou o peso do vasto trabalho que havia para fazer, mas, também, todos os momentos de diversão e de conversa (muitos em colaboração com a Patrícia e o Paulo Jorge, na mesmíssima sala de cultura de tecidos) que levantam o jugo do trabalho que não vai correndo como se desejaria.

À Comissão de Natal 2014 – Iris Caetano, Sónia Pedro, Pedro Loureiro, Dina Carpinteiro, João Lavinha – e aos seus convidados especiais – José Ferrão e Ricardo Faria – por, todos juntos, termos feito a festa de Natal mais memorável que o DGH já viu em toda a sua existência (se por bons motivos, ficará ao critério de quem lá esteve…)

Ao Laurentino Simão, agradeço ter sido o meu companheiro em todas as reuniões da Sociedade Portuguesa de Genética Humana em que participámos, e em todas as visitas de estudo que recebemos no DGH e que tanto gosto me deram fazer.

Aos colegas da Unidade de Inovação e Tecnologia, agradeço todo o apoio técnico na análise das sequências dos meus (muitos) produtos de clonagem. Em especial, ao Daniel Sampaio e à Catarina Silva, que, além de sempre solícitos e muito profissionais, me trataram com grande cuidado e amizade.

Gostaria de deixar aqui, também, um agradecimento muito especial à Isabel Junceiro, responsável pela sala de tratamento de material, por me ter sempre facilitado o acesso a tudo de que necessitei e nunca ter deixado que o meu trabalho se estragasse. Além disso, foram fundamentais o apoio e o incentivo que me foi dando.

Às colegas do secretariado, Isabel Simões e Carina Costa, agradeço o apoio logístico e administrativo, que tanto facilitaram a minha permanência no DGH, e as conversas divertidas e as gargalhadas com que sempre me presentearam.

Estou igualmente grata aos restantes colegas do DGH que nunca deixaram de se preocupar comigo, que me deram todas as palavras de encorajamento e que ficarão para sempre comigo. Gostaria de destacar algumas pessoas com as quais desenvolvi uma maior proximidade: Susana Gomes, Célia Ventura, Cristina Alves, Hildeberto Correia, Maria João Silva, Henriqueta Louro, Bárbara Marques, Elizabeth Silva e Conceição Silva.

Ser aluna da Universidade Nova de Lisboa permitiu-me o acesso aos cursos da NOVA Escola Doutoral que tanto conhecimento prático me trouxeram. Gostaria de agradecer a todos os professores, especialmente aos dos cursos de Comunicação – Ana Sanchez, António Granado e Joana Lobo Antunes – e a todos os colegas que enriqueceram a minha aprendizagem, em particular ao Osvaldo Estrela, que me acompanhou em todos.

Agradeço, também, a todos os meus amigos fora do laboratório por se preocuparem comigo, por me apoiarem e me darem ânimo. Quando mais não seja, por me perguntarem constantemente «Então, como é que está a correr? Já está quase? Vá, força, está quase a

acabar». Sem nenhuma ordem especial: Elizabeth Bochmann, Gordon Cope, Alexandra Bochmann, Rui Franco, Celia Williams, Mariana Mourato, João Loureiro, Carla Barroso, Célia Barros, Sérgio Barros, Sandro Dias, Sofia Carmo, Filomena António, Marta Mendes e Luciano Nobre. Um agradecimento muito grande e especial vai para o meu amigo Norman MacCallum, que me apoiou intensamente e esteve sempre a dar-me alento para trabalhar e ultrapassar as dificuldades que foram surgindo ao longo destes anos. É bom ter amigos, é ainda melhor quando estes se preocupam connosco e nos querem ver bem e felizes.

Agradeço à minha tia Ana pelo apoio e por estar sempre por perto em todos os momentos – bons, maus, assim-assim. Além disso (como se não bastasse), agradeço-lhe por ter ajudado a «patrocinar» a minha ida ao congresso de Heidelberg, que tanto prazer e conhecimento me proporcionou.

À minha prima Vanda, agradeço o apoio e o incentivo que sempre me deu em tudo o que faço. Agradeço as nossas conversas, o desabafar de problemas encafuados cá dentro e o sorriso e o pensamento positivo que vêm sempre à superfície, mesmo quando tudo está prestes a desabar.

Agradeço à minha Madrinha por ser um exemplo, um modelo a seguir. Por ter sido, e continuar a ser, uma combatente, uma resistente, uma força incomensurável de trabalho, esforço e dedicação para mudar e melhorar o seu mundo e o Mundo de nós todos. Agradeço-lhe por, do fundo dos seus noventa intensos anos, me fazer pensar todos os dias «quando for grande, quero ser como ela», apesar de, por maior que eu venha a ser, nunca vir a ser como ela.

E os agradecimentos estão mesmo a acabar. É possível que tenha faltado alguém que, de alguma maneira, tenha contribuído para que este trabalho se concretizasse, mas, se não me lembro, é porque não foi assim tão importante. De que me lembre, faltam «apenas» as três pessoas mais importantes da minha vida e que, sem elas, eu já não estaria aqui, inteira, para conseguir coligir esta lista de agradecimentos: a minha mãe e o meu pai, e o Eduardo.

Diz-se que os filhos não pedem para nascer, não pedem para ter os pais que têm. Se eu não tivesse nascido, haveria de revolver este mundo, o outro e todos os que encontrasse pelo caminho, para nascer, se isso significasse nascer desta mãe e deste pai. Se eu tivesse nascido e não tivesse nascido desta mãe e deste pai, haveria de morrer e voltar a nascer tantas vezes quantas as que fossem necessárias até acertar nesta mãe e neste pai. Sempre me estenderam os seus quatro braços gigantes para me embrulhar e amparar, para me aquecer e amortecer as pancadas que vêm de fora. Sempre, sempre estiveram ao meu lado, para o que desse e viesse, com chuva, com sol, com tempestades a varrer tudo à sua frente. Sempre estiveram lá para me apoiar e me dar força, mas, sempre, sempre, sem perder o Norte, o rigor e a razão. Enaltecendo o bom que fui fazendo, mas apontando o mau e o errado, para que se não voltasse a repetir. E é sobretudo isto que eu lhes agradeço: o equilíbrio e o Norte que sempre me deram e, principalmente, durante estes últimos cinco anos, em que a minha bússola se avariou tantas vezes. Agradeço-lhes por nunca terem desistido de fazer de mim uma pessoa

um bocadinho melhor todos os dias, mesmo quando a fúria e a raiva se colavam a mim e arrastavam tudo à sua frente sem distinguir a quem magoam. Agradeço-lhes por nunca deixarem de acreditar e por me fazerem acreditar, também, que, afinal, é possível; que, afinal, eu consigo, mesmo quando era eu quem já não acreditava, mesmo quando era eu quem queria desistir. Agradeço-lhes por servirem de saco-de-pancada, mas não fazerem caso das nódoas negras que foram surgindo tantas vezes. Agradeço-lhes, todos os dias, por adubarem a minha vida e fazerem de mim uma pessoa um bocadinho melhor a cada dia que passa. Mesmo quando, às vezes, parece que já não vale a pena, eles nunca desistem, eles nunca param. Sempre. Para sempre.

Também não se escolhe o amor. Também não se escolhe com quem se partilha a vida, com tudo o que ela traz de bom e de mau. Simplesmente surge e pronto e temos de lidar com isso sem manual de instruções. Mas quando temos alguém que está aqui todos os dias para nos receber de braços abertos, sem esperar um retorno, sem esperar uma recompensa, então, dá-nos vontade de voltar para casa a correr e irmos encufuar-nos debaixo de umas mãos grandes e quentinhas que nos protegem e fazem esquecer o frio áspero que nos esfiapa até cada fio desaparecer. E, essas mãos, encontrei-as no Eduardo. No Eduardo, que, há quase dez anos, está sempre aqui, ao meu lado, atrás de mim, à minha frente, à minha volta, dependendo de onde vem o frio. No Eduardo, que recebe os meus murmúrios, os meus lamentos, as minhas vicissitudes e os toma para ele, vestindo-os de cores mais suaves para que não magoem tanto os sentidos. No Eduardo, que nunca virou as costas às raivas e às mágoas que eu não soube deixar à porta de casa, e as acolheu e depurou para que não nos infestassem. No Eduardo, que me ouviu apresentar-lhe temas de que nunca ouvira falar e não se fartou, que leu e reviu textos (e esta tese toda, de uma ponta à outra) de assuntos que não lhe dizem nada com os mesmos interesse, atenção e dedicação com que lê uma peça de Pinter. Simplesmente porque gosta de mim. Não sei porque é que gosta de mim, mas gosto que assim seja. E gosto ainda mais de gostar dele. Poderia dizer que o amo com todo o meu coração, mas a fragilidade do órgão não se adequa à situação. Prefiro dizer que o amo com todo o meu fígado, pois é um órgão muito mais robusto, capaz de resistir às agressões mais hediondas e, mesmo assim, conseguir funcionar e regenerar-se.

Falta só agradecer ao *Peúgas*, que me acompanhou em tantos dias solitários a ler e a escrever, que me animou, e anima, todos os dias com os seus miminhos peludos e os seus miados cheios de ternura.

Summary

Eukaryotic gene expression is a very intricate process comprising several tightly regulated steps. One of those is *translation*, whose complex initial phase has been considered the rate-limiting step of protein synthesis. The canonical mechanism of translation initiation consists of recruiting 40S ribosomal subunits and several initiation factors to the 5' terminal cap structure of the messenger RNA (mRNA), and subsequent scanning of the entire 5' untranslated region (5'UTR), until the first AUG in a good initiation context is reached. However, several transcripts are able to maintain their protein expression levels under conditions impairing canonical translation initiation by using mechanisms that allow them to bypass the need of cap recognition and/or 5'UTR scanning.

The aim of this work was to identify proteins that can be translated via non-canonical mechanisms of translation initiation. For that purpose, we thoroughly searched both literature and available databases for proteins whose characteristics suggest they might be good candidates to be translated via non-canonical mechanisms. Based on their characteristics and expression patterns, we selected human up-frameshift 1 (UPF1), human Argonaute RNA-induced silencing complex catalytic component 1 (AGO1), and human MutL homolog 1 (MLH1) transcripts for further experimental validation. We cloned the 5'UTR of each of the selected candidates in a bicistronic system, pR_F, in which the 5' cistron, RLuc, encodes the *Renilla* luciferase protein as an internal control for transfection efficiency, and the 3' cistron, FLuc, encodes the firefly luciferase protein. In this system, FLuc indicates the amount of protein synthesized under the control of the sequence cloned upstream FLuc AUG. The negative control for non-canonical translation initiation mechanisms is the human β-globin (*HBB*) 5'UTR and the positive controls for cap-independent translation activity are the v-myc avian myelocytomatosis viral oncogene homolog (*c-Myc*) IRES sequence (cellular control) and *Encephalomyocarditis* virus (*EMCV*) IRES sequence (viral control). We transfected HeLa (cervical cancer-derived cell line), NCM460 (normal intestinal mucosa-derived cell line) and

HCT116 (colorectal cancer-derived cell line) cells with each of the aforementioned constructs and assessed relative FLuc expression levels by luminometry assays.

Regarding *UPF1* 5'UTR, in all tested cell lines, there was a significant increase (11–27 fold) in relative FLuc expression levels from *UPF1* 5'UTR-containing plasmid compared to those from the empty and *HBB* 5'UTR-containing plasmids, indicating *UPF1* 5'UTR mediates FLuc expression. Transfection of promoterless constructs indicated the presence of a cryptic promoter within *UPF1* 5'UTR. To rule out false-positive results, we transfected cells with *in vitro* transcribed, capped and polyadenylated mRNA containing either *UPF1* 5'UTR or the counterpart controls. Transfection of HeLa, NCM460 and HCT116 cells with *UPF1* 5'UTR-containing transcript resulted in a significant increase in relative FLuc expression levels compared to those from the empty transcript — 2.1-fold in HeLa cells, 2.4-fold in NCM460 cells and 2.5-fold in HCT116 cells. Besides, the increase in relative FLuc expression levels from *UPF1* 5'UTR-containing transcript is similar to that of those from *c-Myc* IRES-containing transcript (the cellular positive control for cap-independent translation activity) in all tested cell lines — 2.8-fold in HeLa cells and 3.1-fold in both NCM460 and HCT116 cells. These results indicate that *UPF1* 5'UTR can mediate cap-independent translation in every tested cell line. To find which part of the sequence is required for mediating cap-independent translation, we performed a deletional and mutational analysis of the sequence and verified that cap-independent translation activity was ceased when the first 100 nucleotides, or the last 125, were absent or altered, showing they are required for such activity. By subjecting cells to several stress stimuli, we observed that such activity is maintained under conditions impairing canonical translation initiation. We also produced *in vitro* monocistronic transcripts without a regular cap structure containing either *UPF1* 5'UTR or each of the control sequences cloned upstream FLuc AUG. We observed a significant increase in relative FLuc expression levels in cells transfected with transcripts containing *UPF1* 5'UTR or each of the positive controls, compared to the empty and *HBB* 5'UTR controls. Altogether, these results clearly indicate that *UPF1* 5'UTR is able to mediate cap-independent translation initiation.

As far as *AGO1* 5'UTR is concerned, we observed a significant 2.8-fold increase in relative FLuc expression levels in HeLa cells transfected with *AGO1* 5'UTR-containing plasmid compared to those observed in cells transfected with the empty plasmid. Such expression levels were, however, significantly lower than those measured from *c-Myc* IRES-containing plasmid — 5.8-fold. False-positive results were ruled out, as *AGO1* 5'UTR sequence neither contains cryptic promoters nor does it promote alternative splicing events, which could mask a putative cap-independent translation activity. By subjecting transfected cells to stress conditions impairing cap-dependent translation initiation, we understood that the identified cap-independent translation activity was not only maintained but also enhanced upon knock-down of eukaryotic initiation factor (eIF) 4E, the cap-binding protein. Furthermore, subjecting cells to conditions of cap-mediated translation inhibition does not affect FLuc expression levels under the control of *AGO1* 5'UTR, but inhibiting the eIF4G–eIF4E interaction significantly reduces

such activity, suggesting *AGO1* 5'UTR-mediated translation may be dependent on eIF4G. However, in cells transfected with *in vitro* transcribed, capped and polyadenylated bicistronic mRNA containing *AGO1* 5'UTR, the relative FLuc expression levels were similar to those from cells transfected with either empty or negative control transcripts. This result indicates that *AGO1* 5'UTR sequence is not able to mediate internal cap-independent translation initiation in conditions in which it does not go through a nuclear experience. Nonetheless, in cells transfected with monocistronic transcripts lacking cap structure, relative FLuc expression levels mediated by the *AGO1* 5'UTR were significantly higher (4.7-fold) than those from the negative controls, indicating that *AGO1* 5'UTR can mediate translation initiation in the absence of the cap structure, when the mRNA 5' end is free. Together, these results indicate that *AGO1* 5'UTR sequence mediates a non-canonical cap-independent eIF4G-dependent mechanism of translation initiation that seems to be enhanced by a free 5' end.

As for *MLH1* 5'UTR, it contains a cryptic promoter. We evaluated the activity of such promoter in HeLa, NCM460, and HCT116 cells and observed that it is much more active in NCM460 cells than in cancer cells and that in colorectal cancer (CRC) cells it is more active than in cervical cancer cells. We also observed an influence of the previously described colorectal cancer-associated c.-28A>T mutation and c.-93G>A polymorphism on translation: a decrease in relative FLuc expression levels in NCM460 cells and a decrease in relative FLuc expression levels in HeLa cells, respectively. Concerning *MLH1* 5'UTR putative cap-independent translation activity, we only observed a significant increase in relative FLuc expression levels in HCT116 cells. By subjecting cells to several stress stimuli, we observed that such activity is maintained in HCT116 cells but not in the other tested cell lines, suggesting the putative cap-independent translation initiation mediated by *MLH1* 5'UTR does not occur in the latter.

Thus, understanding how the synthesis of the selected proteins is regulated will allow us to understand the biological relevance of such mechanisms and to what extent they may provide tools for the development of new therapies for several diseases caused by deregulation of protein synthesis.

Keywords: eukaryotic gene expression, translation initiation, cap-independent translation initiation, internal ribosome entry site (IRES), cap-independent translation enhancer (CITE), UPF1, AGO1, MLH1

Resumo

A expressão génica nos eucariotas é um processo muito intricado, que compreende múltiplos passos firmemente regulados. Um deles é a *tradução*, cuja complexa etapa inicial tem sido vista como o passo limitante da síntese proteica. O mecanismo canónico da iniciação da tradução consiste no recrutamento de subunidades 40S do ribossoma, juntamente com vários factores de iniciação, para a estrutura *cap* localizada na extremidade 5' do mRNA, e subsequente rastreio de toda a região 5' não-traduzida (UTR, do inglês *untranslated region*) até atingir o primeiro codão de iniciação num contexto favorável. Contudo, vários transcritos conseguem manter os seus níveis de expressão proteica em circunstâncias que condicionam a iniciação canónica da tradução; para o efeito, utilizam mecanismos que lhes permitem suprimir a necessidade de reconhecimento da estrutura *cap* e/ou de rastreio da 5'UTR.

Este trabalho teve como objectivo a identificação de proteínas que possam ser traduzidas por meio de mecanismos não-canónicos de iniciação de tradução. Com esse propósito, pesquisámos minuciosamente a literatura e as bases de dados disponíveis, de modo a encontrar proteínas cujas características sugiram que elas possam ser boas candidatas a ter a sua tradução mediada por meio de mecanismos não-canónicos. Com base nas suas características e padrões de expressão, seleccionámos os transcritos correspondentes às proteínas UPF1 (do inglês *human up-frameshift 1*), AGO1 (do inglês *human Argonaute RNA-induced silencing complex catalytic component 1*) e MLH1 (do inglês *human MutL homolog 1*) para ulterior validação experimental. Clonámos a 5'UTR de cada um dos candidatos num sistema bicistrónico, pR_F, em que o cistrão a 5', RLuc, codifica a proteína luciferase da *Renilla* e funciona como controlo interno para a eficiência da transfecção, e o cistrão a 3', FLuc, a luciferase do pirilampo. Neste sistema, a expressão de FLuc indica a quantidade de proteína sintetizada sob o controlo da sequência clonada a montante do AUG da FLuc. O controlo negativo para mecanismos não-canónicos de tradução foi a 5'UTR do transcrito da β-globina humana (*HBB*) e os controlos positivos para a tradução independente da estrutura *cap* foram a

sequência do elemento IRES (do inglês *internal ribosome entry site*) do transcrito *c-Myc* (do inglês *v-myc avian myelocytomatosis viral oncogene homolog*), como controlo celular, e a sequência do elemento IRES do vírus *Encephalomyocarditis* (*EMCV*), como controlo viral. Transfectámos células HeLa (linha celular derivada do cancro do colo do útero), NCM460 (linha celular derivada da mucosa intestinal normal) e HCT116 (linha celular derivada de cancro colorrectal), com cada um dos constructos supracitados e avaliámos os níveis de expressão relativa da FLuc através de testes de luminometria.

Em relação à 5'UTR do transcrito *UPF1*, verificou-se, em todas as linhas celulares, um aumento significativo (11–27 vezes) dos níveis de expressão relativa da FLuc, proveniente do plasmídeo que contém a 5'UTR do transcrito *UPF1*, comparativamente aos provenientes dos plasmídeos vazio e contendo a 5'UTR do transcrito *HBB*, o que indica que a 5'UTR do *UPF1* é capaz de mediar a expressão de FLuc. A transfecção de constructos sem promotor revelou a presença de um promotor críptico na 5'UTR do *UPF1.* De forma a eliminar falsos resultados positivos, transfectámos células HeLa, NCM460 e HCT116, com mRNA produzidos *in vitro* — transcritos, com estrutura *cap* adicionada e poliadenilados —, contendo ora a 5'UTR do *UPF1* ora os controlos congéneres. A transfecção destas células com transcritos contendo a 5'UTR do *UPF1* resultou num aumento significativo dos níveis de expressão relativa de FLuc, em comparação com os do transcrito vazio — 2,1 vezes em células HeLa, 2,4 em células NCM460 e 2,5 em células HCT116. Além disso, o aumento significativo dos níveis de expressão relativa da FLuc do transcrito portador da 5'UTR do *UPF1* é semelhante àquele dos do transcrito portador do IRES do *c-Myc* (o controlo celular positivo para actividade de tradução independente da estrutura *cap*) em todas as linhas celulares testadas — 2,8 vezes em células HeLa e 3,1, tanto em células NCM460 como em HCT116. Estes resultados mostram que a 5'UTR do *UPF1* é capaz de mediar a tradução independente da estrutura *cap* em todas as linhas celulares testadas. Para determinar que parte da sequência é necessária a esta actividade, conduzimos uma análise deleccional e mutacional da sequência e verificámos que a tradução independente da estrutura *cap* era reduzida a níveis semelhantes aos do plasmídeo vazio quando os primeiros 100 nucleótidos, ou os últimos 125, estavam ausentes ou alterados, revelando a importância destas sequências neste processo. Ao sujeitar as células a diversos estímulos de *stress*, observámos que tal actividade é mantida sob condições que diminuem a eficiência da iniciação canónica da tradução. Produzimos, também, transcritos *in vitro*, sem estrutura *cap* funcional, contendo quer a 5'UTR do *UPF1* quer cada uma das sequências de controlo clonadas a montante do AUG da FLuc. Observámos um aumento significativo dos níveis de expressão relativa da FLuc em células transfectadas com transcritos contendo a 5'UTR do *UPF1* — ou a sequência IRES de cada um dos controlos positivos —, comparativamente aos observados em células transfectadas com o transcrito vazio ou o controlo negativo [transcrito contendo a 5'UTR do *HBB*]. Em geral, estes resultados demonstram claramente que a 5'UTR do *UPF1* é capaz de mediar a iniciação da tradução de forma independente da estrutura *cap.*

No que respeita à 5'UTR do *AGO1*, verificámos um aumento significativo — de 2,8 vezes — dos níveis de expressão relativa da FLuc em células HeLa transfectadas com plasmídeos contendo a referida sequência, em comparação com os verificados em células transfectadas com o plasmídeo vazio. No entanto, tais níveis de expressão revelaram-se significativamente mais baixos do que os medidos a partir do plasmídeo contendo o IRES do *c-Myc* — 5,8 vezes mais, em comparação com o plasmídeo vazio. Foram excluídos quaisquer falsos positivos, já que a 5'UTR do *AGO1* não contém promotores crípticos nem promove *splicing* alternativo capazes de mascarar uma pretensa actividade de tradução independente da estrutura *cap*. Ao sujeitar células transfectadas com as referidas construções em condições de *stress*, limitando, assim, a iniciação da tradução dependente da estrutura *cap*, percebemos que a actividade de tradução independente da estrutura *cap* se mantém — e, até, melhora —, com o *knock-down* do factor eucariótico de iniciação 4E (eIF, do inglês *eukaryotic initiation factor*), a proteína de ligação à estrutura *cap*. Ademais, a sujeição de células a condições de inibição de tradução mediada pela estrutura *cap* não afecta os níveis de expressão da FLuc sob o controlo da 5'UTR do *AGO1*, mas a inibição da interacção de eIF4G com eIF4E reduz significativamente tal actividade, o que indica que esta poderá ser dependente de eIF4G. Todavia, em células transfectadas com mRNA bicistrónicos produzidos *in vitro* — transcritos, com estrutura *cap* adicionada e poliadenilados —, contendo a 5'UTR do *AGO1*, os níveis de expressão relativa da FLuc mostraram-se semelhantes aos das células transfectadas com transcritos de controlo negativo ou vazio. Esta conclusão indica que a sequência 5'UTR do *AGO1* não é capaz de mediar a iniciação interna da tradução independente da estrutura *cap* em condições em que não passa por uma experiência nuclear. Porém, em células transfectadas com transcritos monocistrónicos sem estrutura *cap*, os níveis de expressão relativa da FLuc são significativamente mais elevados (4,7 vezes) do que os dos controlos negativos, mostrando que a 5'UTR do *AGO1* pode mediar a iniciação da tradução, na ausência de estrutura *cap*, quando a extremidade 5' do mRNA está livre. No cômputo geral, estes resultados mostram que a sequência 5'UTR do *AGO1* medeia um mecanismo não-canónico de tradução independente da estrutura *cap*, mas dependente do eIF4G, que parece ser potenciado por uma extremidade 5' livre.

Quanto à 5'UTR do *MLH1*, registámos a presença de um promotor críptico. Avaliámos a actividade deste promotor em células HeLa, NCM460 e HCT116 e constatámos que é muito mais activo em células NCM460 do que em células cancerígenas; e que em células de cancro colorrectal é mais activo do que em células de cancro do colo do útero. Apercebemo-nos, também, de uma influência da mutação c.-28A>T e do polimorfismo c.-93G>A [descritos como associados ao cancro colorrectal] na tradução — diminuição dos níveis de expressão relativa da FLuc em células NCM460 e aumento nos níveis de expressão relativa da FLuc em células HeLa, respectivamente. Quanto à presumível actividade de tradução independente da estrutura *cap* mediada pela 5'UTR do *MLH1*, observámos um aumento significativo dos níveis de expressão relativa da FLuc apenas em células HCT116. Ao sujeitarmos cada uma das três

linhas celulares a diferentes estímulos de *stress*, verificámos que esta actividade permanece em HCT116 mas não, necessariamente, nas outras duas.

Assim, a compreensão de como é regulada a síntese das proteínas inicialmente seleccionadas permitir-nos-á compreender a relevância biológica de tais mecanismos — e como poderão eles contribuir para o desenvolvimento de terapias para múltiplas doenças resultantes da desregulação da síntese proteica.

Palavras-chave: expressão génica eucariótica, iniciação da tradução, tradução independente da estrutura *cap*, local de entrada interno do ribossoma, elemento potenciador da tradução independente da estrutura *cap*, UPF1, AGO1, MLH1.

Texto escrito ao abrigo do acordo ortográfico de 1945.

Contents

Figure Index

Table Index

List of abbreviations, acronyms and symbols

µg microgram

µl microlitre

µM micromolar

4E-BP eIF4E-binding protein

4EGI-1 eiF4E–eIF4G interaction inhibitor

A adenine

Å Ångström

aa amino acid

ABC ATP binding cassette

ABCE1 ATP binding cassette subfamily E member 1

AD Alzheimer's disease

ADP adenosine triphosphate

AGO1 Argonaute RISC catalytic component 1

AGO2 Argonaute RISC catalytic component 2

AMP adenosine monophosphate

AMPK AMP-activated protein kinase

A_n polyadenylated tail

Apaf-1 Apoptotic protease activating factor 1

APP amyloid-β precursor protein

A-site aminoacyl site

ATM ataxia telangiectasia mutated kinase

ATP adenosine triphosphate

BAG-1 Bcl-2-associated athanogene 1

Bag-l BCL-2-associated athanogene

BCL2 B-cell limphoma 2

Bcl-xL B-cell lymphoma-extra large

BiP immunoglobulin heavy-chain binding protein

bp base pair

BSA bovine serum albumin

C cytosine

CAT-1 Cationic amino acid transporter

CBC cap-binding complex

CBP cap-binding protein

CC cervical cancer

CDKN2A/p16INK4a cyclin-dependent kinase inhibitor 2A

cDNA mRNA-complementary deoxyribonucleic acid

cIAP2 cellular inhibitor of apoptosis protein

CITE cap-independent translation enhancer

CK2α casein Kinase 2 subunit α

c-Myc v-myc avian myelocytomatosis viral oncogene homolog

$CoCl_2$ Cobalt chloride

CRC colorectal cancer

CRM1 chromosomal maintenance 1

cryo-EM cryo–electron microscopy

CTD carboxyl-terminal domain

C-terminal carboxyl terminus

DAP5 death associated protein

DCP1a decapping mRNA 1a

ddH_2O bidestilled water

DENR density-regulated protein

DM2 myotonic dystrophy type 2

DMEM Dulbecco's Modified Eagle's Medium

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

dsRNA double stranded ribonucleic acid

eEF eukaryotic elongation factor

EGR2 early growth response 2

eIF eukaryotic initiation factor

EJC exon–exon junction complex

ER endoplasmic reticulum

eRF eukaryotic release factor

E-site exit site

EV71 enterovirus 71

FBS foetal bovine serum

FGF fibroblast growth factor

FL-p53 full length p53

FLuc firefly luciferase

FPKM fragments per kilobase gene model and million reads

G guanine

G(5')ppp(5')A non-functional cap analogue structure

g/a_AGO1_F 5’ G-capped/5’ A-capped *AGO1* 5’UTR-containing monocistronic transcript

g/a_EMCV_F 5’ G-capped/5’ A-capped *EMCV* IRES-containing monocistronic transcript

g/a_F 5’ G-capped/5’ A-capped empty monocistronic transcript

g/a_HBB_F 5’ G-capped/5’ A-capped *HBB* 5’UTR-containing monocistronic transcript

g/a_MLH1_F 5’ G-capped/5’ A-capped *MLH1* 5’UTR-containing monocistronic transcript

g/a_MYC_F 5’ G-capped/5’ A-capped *c-Myc* IRES-containing monocistronic transcript

g/a_UPF1_F 5’ G-capped/5’ A-capped *UPF1* 5’UTR-containing monocistronic transcript

GATA-6 GATA-binding factor 6

GCN2 general control nonderepressible 2 kinase

GDP guanosine 5’ diphosphate

GEF guanine exchange factor

Gli glycine

Glu glutamic acid

GSC germline stem cell

GTP guanosine 5’ triphosphate

h hour

HBB haemoglobin, beta

HCT116 Human pre-metastatic colorectal carcinoma-derived cell line

HCV hepatitis C virus

HDM human double minute homolog

HeLa Human cervical cancer-derived cell line

Hh-Gli Hedgehog-Gli

HIF1α hypoxia inducible factor 1 alpha subunit

HIV-1 human immunodeficiency virus type 1

HMGA2 high mobility group A protein 2

HNPCC hereditary nonpolyposis colon cancer

HnRNP heterogeneous nuclear ribonucleoproteins

HNSCC head and neck squamous cell carcinoma

HSP70 heat shock protein 70

HuR human antigen R

IAP inhibitor of apoptosis protein

IDR idarubicin

IGF1R Insulin-like growth factor 1 receptor

IGF2BP1 insulin-like growth factor 2 mRNA binding protein 1

IgG immunoglobulin G

IL interleukin

IP Immunoprecipitation

IRES internal ribosome entry site

ITAF IRES *trans*-acting factor

KD knock-down

KRAS Kirsten rat sarcoma viral oncogene homolog

KRT17 cytokeratin 17

L1/L2 PIWI linker regions

La La autoantigen

LAR Luciferase Assay Reagent

LEF-1 lymphoid enhancer factor-1

M molecular marker

m^{6}A N^6-methyladenosine

m^{7}G 7-methylguanosine

MAGOH mago-nashi homolog

MCT-1 Monocarboxylate transporter 1

MDM2 murine double minute 2

Met methionine

mg milligram

Mid middle

min minute

miR microRNA

MIRES m^{6}A-induced ribosome engagement sites

ml millilitre

MLH1 mutL Homolog 1

mm millimetre

MMR mismatch repair

mRNA messenger ribonucleic acid

mRNP messenger ribonucleoprotein

MSH2 mutS homolog 2

MSI microsatellite instability

mTOR mammalian target of rapamycin

NCM460 Human normal intestinal mucosa-derived cell line

NF nuclear factor

NIH3T3 mouse embryo fibroblast-derived cell line

nM nanomolar

NMD nonsense-mediated mRNA decay

NRAS neuroblastoma resistance to audiogenic seizures viral oncogene homolog

NRF2 Nuclear factor (erythroid-derived 2)-like 2 related factor

ns non-significant

NSCLC non-small cell lung cancer

nt (s) nucleotide (s)

NTC normal termination codon

NTD amino terminal domain

N-terminal amino terminus

ºC degrees Celsius

OCT4 octamer-binding protein 4

ODC ornithine decarboxylase

OIS oncogene-induced senescence

ORF open reading frame

p38-MAPK p38 mitogen-activated protein kinases

p53 tumour protein 53

p70-S6 ribosomal protein S6 kinase B1

PABPC1 poly(A)-binding protein, cytoplasmic 1

PARP poly(ADP-ribose) polymerase 1

PAZ PIWI/Argonaute/Zwille

PBMC peripheral blood mononuclear cells

PBS phosphate-buffered saline

PCBP1 Poly(rC)binding protein 1

PCR polymerase chain reaction

PITSLREp58 cyclin-dependent kinase

PIWI P-element-induced whimpy testes

PKA RII-β protein kinase cAMP-dependent type II regulatory subunit beta

PKR Protein kinase RNA-activated

pmol picomole

PNA peptide nucleic acids

PNRC2 proline-rich nuclear receptor co-regulatory protein 2

Pol polymerase

Poly(A) polyadenylated

P-p70 S6 phosphorylated ribosomal protein S6 kinase B1

p-R_101-275_F *UPF1* 5'UTR nts 101–275-containing bicistronic plasmid

pR_101-275_F *UPF1* 5'UTR nts 101–275-containing bicistronic plasmid

pR_1-100_F *UPF1* 5'UTR nts 1–100-containing bicistronic plasmid

p-R_1-100_F *UPF1* 5'UTR nts 1–100-containing bicistronic plasmid

pR_1-150_F *UPF1* 5'UTR nts 1–150-containing bicistronic plasmid

p-R_1-150_F *UPF1* 5'UTR nts 1–150-containing bicistronic plasmid

pR_1-200_F *UPF1* 5'UTR nts 1–200-containing bicistronic plasmid

p-R_1-200_F *UPF1* 5'UTR nts 1–200-containing bicistronic plasmid

pR_1-50_F *UPF1* 5'UTR nts 1–50-containing bicistronic plasmid

p-R_1-50_F *UPF1* 5'UTR nts 1–50-containing bicistronic plasmid

p-R_151-275_F *UPF1* 5'UTR nts 151–275-containing bicistronic plasmid

pR_151-275_F *UPF1* 5'UTR nts 151–275-containing bicistronic plasmid

p-R_201-275_F *UPF1* 5'UTR nts 201–275-containing bicistronic plasmid

pR_201-275_F *UPF1* 5'UTR nts 201–275-containing bicistronic plasmid

pR_51-275_F *UPF1* 5'UTR nts 51–275-containing bicistronic plasmid

p-R_51-275_F *UPF1* 5'UTR nts 51–275-containing bicistronic plasmid

pR_AA_F *UPF1* 5'UTR mutated at nts 39–40-containing bicistronic plasmid

pR_AAT-ATT_F *UPF1* 5'UTR mutated at nts 161–163 and 209–211-containing bicistronic plasmid

pR_AGO1_F *AGO1* 5'UTR-containing bicistronic plasmid

p-R_AGO1_F *AGO1* 5'UTR-containing bicistronic plasmid

pR_ATA_F *UPF1* 5'UTR mutated at nts 98–100-containing bicistronic plasmid

pR_EMCV_F *EMCV* IRES-containing bicistronic plasmid

p-R_EMCV_F *EMCV* IRES-containing bicistronic plasmid

pR_F empty bicistronic plasmid

p-R_F promoterless bicistronic plasmid

pR_HBB_F *HBB* 5'UTR-containing bicistronic plasmid

p-R_HBB_F *HBB* 5'UTR-containing bicistronic plasmid

pR_MLH1_F *MLH1* 5'UTR-containing bicistronic plasmid

p-R_MLH1_F *MLH1* 5'UTR-containing bicistronic plasmid

pR_MYC_F *c-Myc* IRES-containing bicistronic plasmid

p-R_MYC_F *c-Myc* IRES-containing bicistronic plasmid

pR_UPF1_F *UPF1* 5'UTR-containing bicistronic plasmid

p-R_UPF1_F *UPF1* 5'UTR-containing bicistronic plasmid

PSF/SFPQ PTB-associated Splicing Factor

P-site peptidyl-site

PTB polypyrimidine tract binding protein

PTC premature termination codon

PTC1-L PTC1 isoform L

PTCH1b patched drosophila homolog 1b

PTX paclitaxel

QGRS Quadruplex forming G-rich sequences

R_101-275_F *UPF1* 5'UTR nts 101–275-containing bicistronic transcript

R_1-100_F *UPF1* 5'UTR nts 1–100-containing bicistronic transcript

R_1-150_F *UPF1* 5'UTR nts 1–150-containing bicistronic transcript

R_1-200_F *UPF1* 5'UTR nts 1–200-containing bicistronic transcript

R_1-50_F *UPF1* 5'UTR nts 1–50-containing bicistronic transcript

R_151-275_F *UPF1* 5'UTR nts 151–275-containing bicistronic transcript

R_201-275_F *UPF1* 5'UTR nts 201–275-containing bicistronic transcript

R_51-275_F *UPF1* 5'UTR nts 51–275-containing bicistronic transcript

R_AA_F *UPF1* 5'UTR mutated at nts 39–40-containing bicistronic transcript

R_AAT-ATT_F *UPF1* 5'UTR mutated at nts 161–163 and 209–211-containing bicistronic transcript

R_AGO1_F *AGO1* 5'UTR-containing bicistronic transcript

R_ATA_F *UPF1* 5'UTR mutated at nts 98–100-containing bicistronic transcript

R_EMCV_F *EMCV* IRES-containing bicistronic transcript

R_F empty bicistronic transcript

R_HBB_F *HBB* 5'UTR-containing bicistronic transcript

R_MLH1_F *MLH1* 5'UTR-containing bicistronic transcript

R_MYC_F *c-Myc* IRES-containing bicistronic transcript

R_UPF1_F *UPF1* 5'UTR-containing bicistronic transcript

RBP ribonucleoprotein

RF release factor

RHA RNA helicase A

RISC RNA-induced silencing complex

RLuc Renilla luciferase

RNA ribonucleic acid

RNPC1 RNA binding motif protein 38

RPL26 ribosomal protein L26

RPMI Roswell Park Memorial Institute–1640

RPS25 ribosomal protein S25

rRNA ribosomal ribonucleic acid

RT reverse transcription

rt room temperature

RUNX2 runt-related transcription factor 2

s second

SDS sodium dodecyl sulphate

SEAP secreted alkaline phosphatase

SF1 superfamily 1

shRNA small/short hairpin RNA

SL stem loop

SMAGP small cell adhesion glycoprotein

SMAR1 scaffold/matrix attachment region-binding protein 1

SMD Staufen1-mediated mRNA decay

SMG suppressor with morphogenetic effect on genitalia
SNAT2 sodium-coupled neutral amino acid transporter 2
snRNP small nuclear ribonucleoprotein
S-ODN1 phosphorothioate antisense oligonucleotide 1
Sp1 specificity protein-1
SURF SMG1–UPF1–eRF1–eRF3
SV40 Simian vacuolating virus 40
T thymine
TBS tris-buffered saline
TCP80 translational control protein 80
TERRA telomeric repeat-containing RNA
TISU translation initiator of short 5'UTR
TNF tumour necrosis factor
tRNA transporter ribonucleic acid
$tRNA_i$ initiator transporter ribonucleic acid
TRPV3 transient receptor potential cation channel subfamily V member 3
TSS transcription start site
Tyr tyrosine
U uridine
uAUG upstream AUG
Unr upstream of N-ras
uORF upstream open reading frame
UPF1 up-frameshift 1regulator of nonsense transcripts yeast homolog
UPF2 UPF2 regulator of nonsense transcripts yeast homolog
UTR untranslated region
UV ultra-violet
v/v volume per volume
VEGF vascular endothelial growth factor
w/v weight per volume
WT1 Wilms' tumour suppressor gene
XIAP X-linked inhibitor of apoptosis protein
XRN1 exoribonuclease 1
YBX1 Y-box binding protein
ZNF9 zinc finger protein 9
γ-H2AX phosphorylated histone H2AX
ΔG Gibbs minimum free energy
ΔN-p53 p53 N-terminal truncated isoform

1.

Introduction

1.1. Overview of eukaryotic gene expression

Eukaryotic gene expression explains how the genetic information stored as DNA (deoxyribonucleic acid) molecules is read out as protein “machines” to be used by the cell. Yet, this process requires a third, intermediate molecule, RNA (ribonucleic acid). This flow of information, from DNA to RNA, and from RNA to protein, is known as the *Central Dogma of Biology*. It states that the coded genetic information encrypted in the DNA is transcribed into individual transportable cassettes, composed of messenger RNA (mRNA) and each of these is programmed for the synthesis of a certain protein or small number of proteins (Lodish *et al.*, 2000). The mechanism whereby the information encoded in the DNA is deciphered into proteins is now quite well understood as there are many exceptions to this rule brought up in recent years. Results of genomic studies revealed that much of the DNA that does not encode proteins encodes various types of functional RNA (ENCODE Project Consortium *et al.*, 2007; Gerstein *et al.*, 2007). However, how gene expression regulation occurs in eukaryotic cells — i.e., how cells determine to make the right proteins at the right time in the right amount — is still a major focus of current research in Molecular Cell Biology.

There are several steps throughout the gene expression regulation pathway. Although they are usually studied as independent events, each of these stages represents a subdivision of a continuous process, with each phase physically and functionally connected to the next (Orphanides and Reinberg, 2002). Such continuous process includes events like transcription, mRNA processing, export and translation, protein folding and transport, which will be briefly described below (Figure 1.1).

Transcription is the production of RNA copies from the DNA template performed by RNA polymerases that add one RNA nucleotide at a time to a growing strand of RNA (Lodish *et al.*, 2000). There are three types of RNA polymerases, and each needs a specific promoter and a set of transcription factors to initiate the process (Cooper, 2000; Weipoltshammer and Schöfer, 2016). RNA polymerase I is responsible for transcription of ribosomal RNA (rRNA) genes, whereas RNA polymerase II (RNA Pol II) transcribes all protein-coding genes and also non-coding RNA, and RNA polymerase III transcribes 5S rRNA, transfer RNA (tRNA) genes, and some small non-coding RNA (Cooper, 2000; Lee *et al.*, 2004; Weipoltshammer and Schöfer, 2016).

Transcription by RNA Pol II originates a pre-mRNA molecule in a process composed of three stages: initiation, elongation and termination. It starts when the preinitiation complex — composed of RNA Pol II and several auxiliary proteins, known as *transcription factors* — recognizes and binds to consensus sequences in the promoter located upstream of the start site for transcription (Proudfoot *et al.*, 2002; Luna *et al.*, 2008; Hocine *et al.*, 2010). In addition, the activity of promoters may be greatly increased by enhancer sequences that can act over distances of several kilobases located either upstream or downstream of the gene to be transcribed. At that time, transcription factors recruit and position RNA Pol II near the transcription start site and, subsequently, elongation occurs after transition to an RNA Pol II

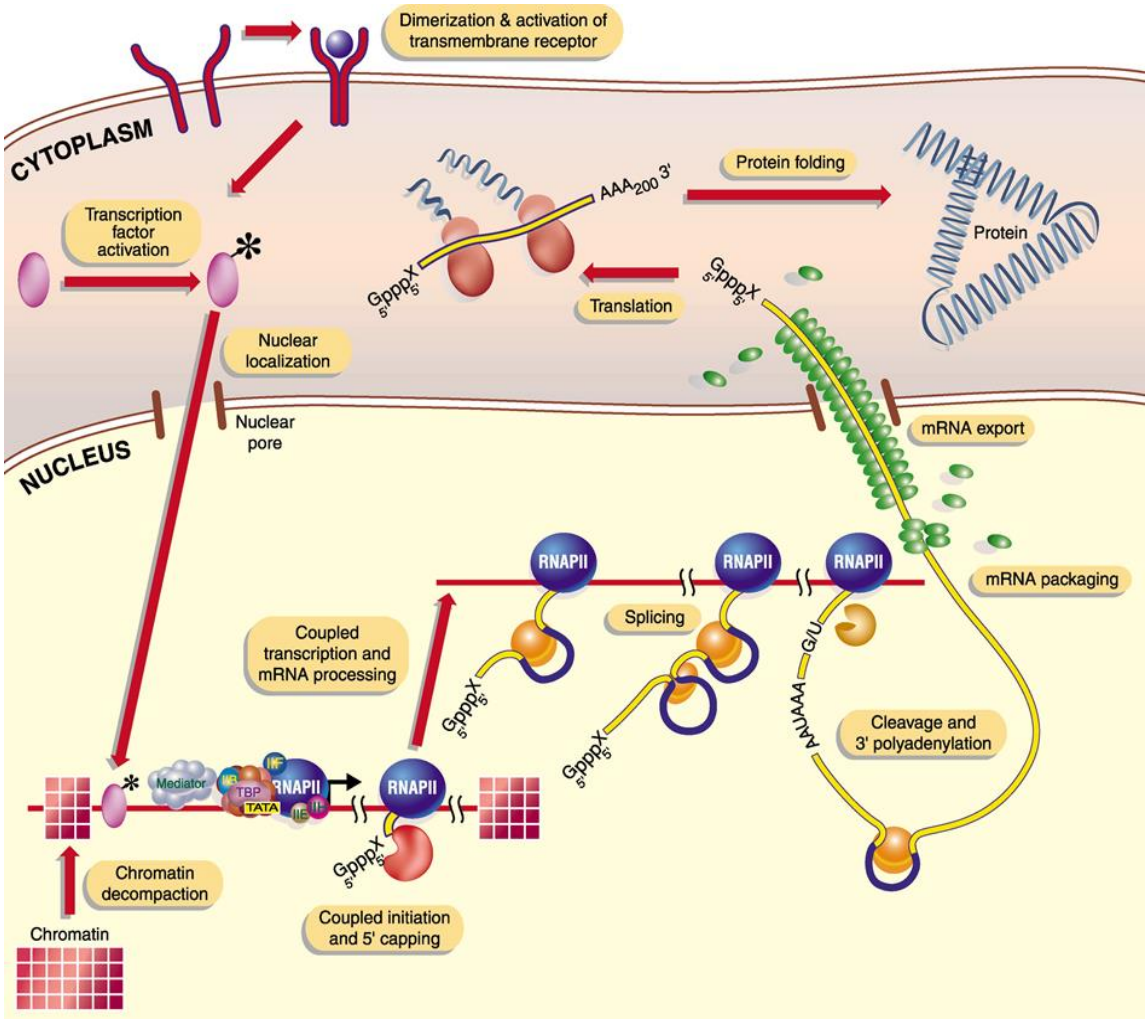


Figure 1.1 — Overview of gene expression. Each step of the gene expression pathway is physically and functionally connected to the next and is a subdivision of a continuous process. Adapted from Orphanides and Reinberg, 2002.

elongation complex. This switch is associated with alterations in the chromatin structure and changes in the RNA Pol II C-terminal domain (CTD) phosphorylation state, which binds to various protein factors that promote transcript maturation and modification (Phatnani and Greenleaf, 2006; Hocine *et al.*, 2010). Then, RNA Pol II proceeds through the remainder of the gene and transcription stops when conserved polyadenylation signals direct cleavage and polyadenylation at the 3' end of the nascent transcript (Luna *et al.*, 2008; Bentley, 2014).

In order to become a mature mRNA, the nascent transcript needs to be processed. In the course of transcription elongation, several mRNA processing events, such as splicing and 5' capping, take place (Luna *et al.*, 2008; Hocine *et al.*, 2010; Bentley, 2014). During 5' capping, a set of enzymatic reactions adds the 7-methylguanosine (m^7G) to the 5' end of the nascent transcript, a structure that helps give the transcript stability by protecting it from 5' to 3' exonuclease degradation (Luna *et al.*, 2008; Hocine *et al.*, 2010; Bentley, 2014). This structure also serves as a binding site for the cap-binding complex (CBC) — which is composed of cap-

binding protein (CBP) 80 and CBP20, and needed in splicing, export and first round of translation — and for eIF4E, which replaces CBC in the subsequent rounds of translation (Neugebauer, 2002; Luna *et al.*, 2008; Hocine *et al.*, 2010). Also in the course of transcription elongation, a process called *splicing* occurs. This process consists of two transesterification reactions catalysed by the spliceosome in which the introns are removed and the neighbouring exons are spliced together (Wahl *et al.*, 2009). The spliceosome is a highly dynamic machine responsible for removing the vast majority of pre-mRNA introns, whose building blocks are the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4/U6, and U5 (Wahl *et al.*, 2009). The formation of this complex occurs at particular splice junctions and depends on certain sequences, including the 5' splice site (that includes an almost invariant sequence GU at the 5' end of the intron), the branch point sequence (which contains a conserved adenosine important to intron removal), the polypyrimidine tract (a variable stretch of pyrimidines, which is thought to recruit factors to the branch point sequence and 3' splice site), and the 3' splice site (terminates the intron at the 3' end with an almost invariant AG sequence) (Neugebauer, 2002; Luna *et al.*, 2008; Wahl *et al.*, 2009; Hocine *et al.*, 2010; Will and Luhrmann, 2011). On the other hand, processing of the nascent RNA must also occur at the 3' end concurrently to transcription termination (Buratowski, 2005). Thus, 3' cleavage and polyadenylation occur if the polyadenylation signal sequence (generally, 5'-AAUAAA-3') is present in the pre-mRNA, leading to the cleavage of the pre-mRNA molecule followed by the addition of a series of ~200 adenines (A), which forms a poly(A) tail that protects the RNA from degradation (Shi and Manley, 2015). Furthermore, specific RNA-binding proteins must be loaded onto nascent transcripts, thus forming mRNPs that are, then, ready to be exported to the cytoplasm (Mor *et al.*, 2010;Katahira, 2015). Once in the cytoplasm, mRNPs can undergo remodelling and the mRNA is ready for translation (Iglesias and Stutz, 2008; Rougemaille *et al.*, 2008).

Translation is a process that takes place in large ribonucleoprotein complexes — the ribosomes — and is typically divided in four phases: initiation, in which there is the localisation of the initiation codon by ribosomal subunits and eIFs; elongation, in which an amino acid is added at a time to the nascent peptide, according to the sequence encoded in the mRNA molecule; termination, which occurs when the ribosome reaches a stop codon, and leads to the release of the polypeptide; and recycling, during which ribosomal subunits must be dissociated, and the mRNA and deacylated tRNA released to regenerate the necessary components for subsequent rounds of translation. (Dever and Green, 2012a; Hinnebusch, 2014). The polypeptide released after translation termination is therefore folded into a three-dimensional structure as a consequence of its amino acid composition and subsequent interaction, resulting in a protein in its native state, which is associated with a particular function (Herczenik and Gebbink, 2008). The proteins are then transported to their corresponding organelle or are exported to other cells, according to their signal peptide, a 5-30 amino acid peptide present at the N-terminal end of most newly synthesized proteins that are destined towards the secretory pathway, in order to fulfil their role in the organism (van Vliet *et al.*, 2003).

1.2. A closer look into translation

Translation is a very important step of eukaryotic gene expression as it plays a crucial role in many fundamental biological processes, including cell growth, development and the response to environmental stresses and other biological cues (López-Lastra *et al.*, 2005). Translational control allows fine-tuning of gene expression by stimulating or repressing the translation of specific mRNA through the reversible phosphorylation of translation factors (Liu and Qian, 2014). Deregulation of translation is therefore a major event that may lead to cell transformation and to the development of diseases such as cancer.

In the following paragraphs, we will describe the different phases of eukaryotic translation, according to the state-of-the-art of the field.

1.2.1. Initiation

Due to its complexity and so many factors involved, initiation has been considered the rate-limiting step of protein synthesis and the recruitment of the ribosome is crucial in translational control (Jackson *et al.*, 2010).

Over the past decades many discoveries regarding how translation initiation occurs in eukaryotes have been made. In 1979, Marilyn Kozak first proposed the scanning model of translation initiation, according to which 40S ribosomal subunits are recruited to the 5' terminal cap structure, scan the entire 5' untranslated region (UTR) of the mRNA in a 5' to 3' direction and initiate translation at the first AUG in a good initiation context. Up until now, many have been the discoveries made in this field and nowadays the knowledge about this mechanism is far broader than it was at the time. Although the events taking place during translation initiation are all dependent upon each other, for the sake of simplicity, we will describe them as (i) formation of the ternary complex; (ii) formation of the 43S preinitiation complex; (iii) binding of the mRNA to the 43S; (iv) scanning of 5'UTR and AUG recognition; (v) assembly of 80S ribosome; and (vi) recycling of eIF2-GTP. Figure 1.2 illustrates these events:

i) *Formation of the ternary complex*. Translation initiation starts with this event. The ternary complex is composed of eukaryotic initiation factor 2 (eIF2) — a hetero-dimer of 3 subunits (α, β and γ), with a total molecular weight of ~125KDa — bound to the Met-tRNA$_i^{Met}$ and GTP by the γ subunit (Erickson and Hannig, 1996). Its assembly is controlled by the guanine nucleotide exchange factor (GEF) eIF2B, a 5-subunit protein that converts eIF2-GDP to the active eIF2-GTP complex before each round of translation (Gomez *et al.*, 2002). GTP is hydrolysed after recognition of the AUG start codon, producing eIF2 bound to GDP, which has a 10-fold reduced affinity for Met-tRNA$_i^{Met}$ (figure 1.2.A) (Kapp and Lorsch, 2004a). This GTP-dependent recognition of the methionine moiety may, in part, prevent unacylated tRNA$_i$ from entering the initiation pathway and is likely to be an important part of the tRNA release mechanism from eIF2 after initiation codon recognition (Kapp and Lorsch, 2004a; Kapp and Lorsch, 2004b).

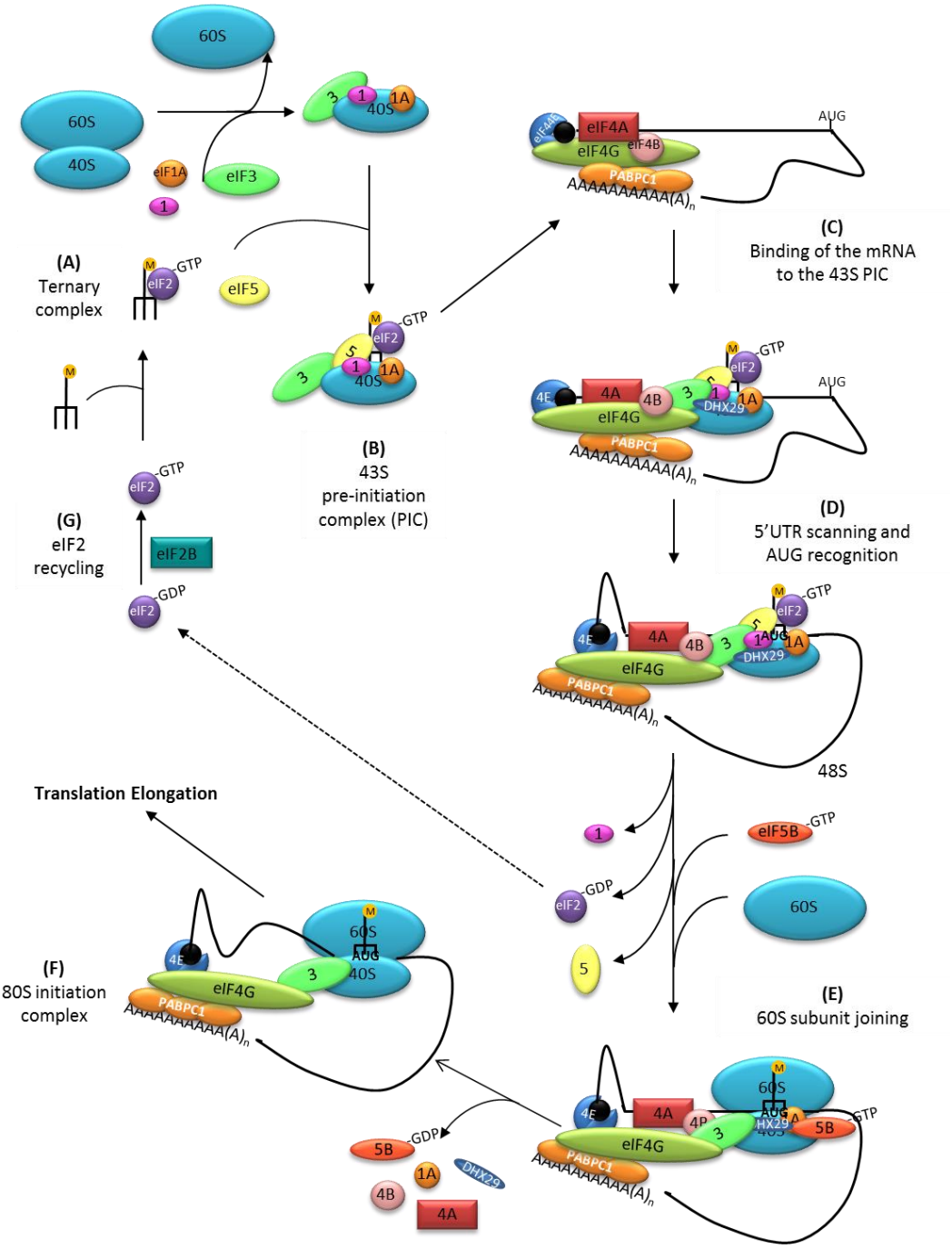


Figure 1.2 — The canonical model of eukaryotic translation initiation. (A) Translation initiation starts with the formation of the ternary complex, composed of eIF2 bound to the Met-tRNAi and GTP. (B) Once the ternary complex is assembled and active, it must bind the 40S ribosomal subunit with the aid of eIF1, eIF1A, eIF3 and eIF5, forming the 43S preinitiation complex. (C) Then, the 43S preinitiation complex must bind the cap structure at the 5' end of the mRNA molecule, so it can scan the mRNA for the initiation codon. (D) The scanning stops when the 43S complex finds the first initiation codon in a favourable context, thus forming a stable complex known as 48S initiation complex. (E) After 48S initiation complex is formed, several events take place so that 60S subunit may join and form the 80S ribosome. This reaction requires eIF5B, which hydrolyses the eIF2-GTP, thereby releasing the initiation factors, including eIF2-GDP, from the small ribosomal subunit, leaving the initiator $tRNA_i$ bound to the start codon. Following eIF2-GDP dissociation, eIF5B-GTP binds to the 40S subunit and accelerates the rate of 60S subunit joining. (F) Once initiation step is finished and the ribosome has entered the elongation phase, eIF2 is recycled as to enable ternary complex formation once again for another round of translation to take place.

ii) Formation of the 43S preinitiation complex. Once the ternary complex is assembled and

active, it must bind the 40S ribosomal subunit. According to current models based on studies in reconstituted eukaryotic systems, this binding is aided by eIF1, eIF1A, eIF3 and eIF5 (Pestova *et al.*, 1998; Valásek *et al.*, 2002; Olsen *et al.*, 2003; Majumdar, 2003; Kolupaeva, 2005). In this way, small factors such as eIF1 and eIF1A might alter the local conformation of the eIF2 binding site. In addition, recent works revealed the crystal structure of 40S subunit in complex with eIF1 alone; eIF1 and eIF1A; and mRNA, tRNA and eIF1A, allowing understanding the location of these factors and tRNA bound to small ribosomal subunit. This provides insight into the details of translation initiation specific to eukaryotes, which will eventually have implications in the mechanism of mRNA scanning (Lomakin and Steitz, 2013). Although eIF1 and eIF1A promote scanning, eIF1 and possibly the C-terminal tail of eIF1A must be displaced from the Peptidyl (P) decoding site to permit base-pairing between Met-tRNA$_i$ and the AUG codon, as well as to allow subsequent phosphate release from eIF2-GDP (Lomakin and Steitz, 2013; Hinnebusch, 2014). On the other hand, a large factor such as eIF3 might distort the conformation of the entire 40S subunit to allow easier access of eIF2 with its attached Met-tRNA$_i$ (Kapp and Lorsch, 2004b). eIF3 is a multi-subunit complex composed of 13 subunits (a–m), whose structure is only now beginning to emerge (Hinnebusch, 2014). 40S binding by eIF3 is enhanced by eIF3j subunit (Kolupaeva, 2005). However, negative cooperativity is observed between the binding of eIF3j and the binding of eIF1, eIF1A and ternary complex with the 40S subunit; so, to overcome this, eIF3 dramatically increases the affinity of eIF1 and eIF3j for the 40S subunit (Sokabe and Fraser, 2014). eIF3 spans the entry and exit channels on the backside of the 40S subunit (Hinnebusch, 2014). Nevertheless, since much of this factor is flexible, it may communicate dynamically with factors bound to the interface side of the 40S subunit. This would agree with its binding to the aminoacyl (A) site and mRNA entry channel of the 40S subunit, via placing eIF3j C-terminal domain directly in the ribosomal decoding centre (Fraser *et al.*, 2007; Sun *et al.*, 2011a). eIF3j also interacts with eIF1A and reduces 40S subunit affinity for mRNA. A high affinity for mRNA is restored after recruitment of initiator tRNA, even though eIF3j remains in the mRNA-binding cleft in the presence of tRNA. These results suggest that eIF3j functions in part by regulating access of the mRNA-binding cleft in response to initiation factor binding (Fraser *et al.*, 2007). eIF5 also affects ternary complex recruitment, as it is crucial in the assembly of the eukaryotic preinitiation complex, serving as an adaptor between eIF3 bound to the 40S subunit and the ternary complex. It is likely to stabilise ternary complex binding to the 40S via simultaneous interactions with both structures (Kapp and Lorsch, 2004b). This activity is mediated by the ability of its C-terminal HEAT domain to interact with eIF1, eIF2 and eIF3 in the multifactorial complex (Yamamoto *et al.*, 2005). The structure resulting from the binding of ternary complex to 40S ribosomal subunit, together with the aforementioned initiation factors, is designated 43S preinitiation complex (figure 1.2.B). Mutations in the eIF5 or eIF3a segments that disrupt interactions among eIF1, eIF5 C-terminal domain, eIF3c N-terminal domain and eIF3a C-terminal domain, and interactions between each of those segments and eIF2β N-terminal domain, impair cell growth that is mitigated by ternary complex overexpression (Hinnebusch, 2014). Also, after substitutions in eIF5 CTD that had weakened its binding to

eIF2β N-terminal tail, a reduced occupancy of eIF2 was shown (Yamamoto *et al.*, 2005). eIF3c NTD mutations probably reduce ternary complex recruitment by weakening the interaction between eIF3c NTD and eIF5 CTD, or the ability of eIF5 CTD to interact with eIF2β N-terminal tail in the multifactorial complex (Karaskova *et al.*, 2012). Although the preassembly of the multifactorial complex components is not required to the stimulatory effect of these components on ternary complex recruitment (Sokabe *et al.*, 2012), it seems likely that the preformed multifactorial complex provides a major pathway to ternary complex recruitment *in vivo* (Hinnebusch, 2014).

iii) Binding of the mRNA to the 43S preinitiation complex. Once it is assembled, the 43S preinitiation complex must bind the cap structure at 5' end of the mRNA molecule, so it can then scan the UTR and reach the 5' proximal AUG. eIF4F is crucial in recognizing the m^7G cap structure, because eIF4E recruits eIF4G/eIF4A to the 5' end (Pestova *et al.*, 2007). Apart from its role in directly binding the cap structure, eIF4E stimulates eIF4A helicase activity. This activity promotes mRNA restructuring in a manner that is independent of its cap-binding function. The eIF4E-binding site in eIF4G functions as an auto-inhibitory domain to modulate its ability to stimulate eIF4A helicase activity, but binding of eIF4E counteracts this auto-inhibition, enabling eIF4G to stimulate eIF4A helicase activity (Feoktistova *et al.*, 2013). eIF4A is a DEAD box RNA helicase, whose ATPase activity is required for duplex unwinding *in vitro* (Pause and Sonenberg, 1992). It is held in its active conformation by eIF4G, which enables it to unwind the 5'UTR of the mRNA and produce a single-stranded binding site for the 43S preinitiation complex near the 5' cap (Oberer *et al.*, 2005; Özeş *et al.*, 2011). Its RNA-unwinding activity is stimulated by eIF4B and eIF4H, two RNA-binding proteins that are thought to play functionally redundant roles in translation initiation (Grifo *et al.*, 1984; Richter-Cook, 1998). While the depletion of eIF4B from mammalian cells resulted in the inhibition of translation initiation, preferably of mRNA with more structured 5'UTR (Shahbazian *et al.*, 2010), eIF4H is less effective in increasing the efficiency in coupling ATP hydrolysis to duplex unwinding by eIF4F (Özeş *et al.*, 2011), because it does not have the C-terminal RNA-binding region found in eIF4B, which is instrumental in stimulating eIF4A helicase activity (Rozovsky *et al.*, 2008). Regarding eIF4G, it is a high-molecular-weight protein that acts as a scaffold for binding eIF4E and eIF4A. In addition, eIF4G helps recruit the 43S preinitiation complex to the mRNA by directly interacting with eIF3. eIF4G binds to eIF3 through eIF3 subunits c, d and e, independently of eIF4A binding to the middle region of eIF4G (Villa *et al.*, 2013). Altogether, at the 5' end of the mRNA, the binding of the preinitiation complex to the mRNA involves the cooperative activities of eIF4F, eIF3, eIF4B and eIF4H (Figure 1.2.C).

iv) Scanning of the 5'UTR and AUG recognition. After proper assembly at the 5' end of the mRNA, the preinitiation complex needs to scan the mRNA to find the initiation codon (Kozak, 1989; Kozak, 2002). If the 5'UTR is unstructured, a minimal 43S complex (comprising only 40S, eIF1, eIF2-Met-$tRNA_i$ and eIF3) is capable of scanning without any requirement for ATP hydrolysis or factors associated with it. However, if — as in most cases — the 5'UTR is at least mildly structured, this scanning process requires the hydrolysis of ATP, eIF1, eIF1A, and

DHX29[1], a protein that binds directly to the 40S subunit and eIF1A (Pestova *et al.*, 1998; Pestova and Kolupaeva, 2002; Pisareva and Pisarev 2016). Both eIF1 and eIF1A, together with the 40S, form a tunnel through which the mRNA slides. Such spatial conformation ensures scanning processivity by keeping the mRNA unstructured and properly oriented for the examination of the nucleotide sequence in the P site by $tRNA_i$, i.e., the $tRNA_i$ attempts to establish Watson-Crick base pairing between its anticodon and a nucleotide triplet of mRNA moving through the P site. The basic loop of eIF1 competes for the P site of the 40S subunit with the Anti-Stem Loop of $tRNA_i$, as they end up displacing each other during mRNA scanning (Lomakin and Steitz, 2013). Moreover, the hydrolysis of ATP can be used by eIF4A or Ded1p (both DEAD-box family members with helicase activity) to actively translocate the ribosome in a 5' to 3' direction, or unwind secondary structures in the mRNA. This leads to a diffusive movement of the ribosome that is prevented from backsliding due to reforming of the unwound structures behind it (figure 1.2.D) (De La Cruz *et al.*, 1997; Kapp and Lorsch, 2004b). Then, the scanning stops when the 43S complex finds the first AUG in a favourable context, i. e. with a purine (usually A) in position -3 and a guanine in position +4 (Kozak, 1989). Once the AUG codon is in the P site, it becomes base-paired with all three nucleotides of the anticodon of the $tRNA_i$, thereby stabilising the conformation of the $tRNA_i$ while allowing it to displace the basic loop of eIF1 (Lomakin and Steitz, 2013), thus forming a stable complex known as 48S initiation complex (figure 1.2.D). Further, eIF1 has an important role in the selection of the start codon as it is required for the 43S preinitiation complex to discriminate between cognate and non-cognate initiation codon. In the absence of this factor, scanning complexes are arrested at good and bad initiation codon contexts with similar efficiency. Accordingly, eIF1 is able to dissociate 48S complexes preassembled at an upstream initiation codon in a bad context, resulting in formation of a stable complex at the next downstream codon in good context (Pestova *et al.*, 1998; Pestova and Kolupaeva, 2002). Together, eIF5A and eIF5B also stimulate 48S initiation complex formation by means of influencing initiation codon selection during ribosomal scanning. eIF5A alone may promote 48S initiation complex formation simply by allowing GTP hydrolysis and AUG recognition at the expense of continued scanning downstream. However, such 48S initiation complexes are less stable due to eIF2-GDP dissociation from Met-$tRNA_i$. So, eIF5B is then required to stabilise Met-$tRNA_i^{Met}$ in the P site, operating only after AUG recognition and release of eIF2-GDP from the 48S initiation complex by its ability to stabilise Met-$tRNA_i^{Met}$ in the P site (Pisareva and Pisarev, 2014).

v) Assembly of the 80S ribosome. After 48S initiation complex is formed, several events take place so that 60S subunit may join it and, thus, form the 80S ribosome. This reaction requires eIF5B, which hydrolyses the eIF2-GTP, thereby releasing the initiation factors, including eIF2-GDP, from the small ribosomal subunit, leaving the initiator $tRNA_i$ bound to the start codon (Hinnebusch, 2011; Kuhle and Ficner, 2014). Following eIF2-GDP dissociation, eIF5B-GTP binds to the 40S subunit and accelerates the rate of 60S subunit joining. Alterations in the C-terminal sequence of eIF1A reduce both the GTP hydrolysis and subunit joining activities of

[1] DHX29, a protein that binds directly to the 40S subunit and eIF1A (Pisareva and Pisarev, 2016)

eIF5B without significantly affecting earlier steps of translation initiation. On the other hand, disruption of the eIF5B C-terminal binding domain for eIF1A results in similar decrease in GTPase and subunit joining activities. Altogether, these data indicate that the eIF1A-eIF5B C-terminal interaction is decisive for efficient ribosomal subunit joining and subsequent hydrolysis by eIF5B (Acker *et al.*, 2006). Joining of the 60S subunit (figure 1.2.E) requires a second step of GTP hydrolysis in order to make the 80S ribosome set to polypeptide synthesis (Lee *et al.*, 2002; Shin *et al.*, 2002). GTPase activity of eIF5B is stimulated by 60S subunits and even more strongly by 80S ribosomes; so, GTP-bound eIF5B stimulates 60S subunit joining and GTP hydrolysis occurs after 80S subunit formation, promoting the release of the factor from the 80S complex once the subunit joining step has been completed. As a result, cells require hydrolysis of GTP by both eIF2 and eIF5B to complete translation initiation (Pestova *et al.*, 2000; Lee *et al.*, 2002; Shin *et al.*, 2002). At this stage, 80S ribosomal complex is assembled and ready to start decoding the sequence within the mRNA and eventually originate a polypeptide (figure 1.2.F).

vi) Recycling of eIF2-GDP. As soon as the initiation step is finished and the ribosome has entered the elongation phase, eIF2 released from the ribosome is bound to GDP. However, the latter must be replaced by GTP to enable ternary complex formation again for another round of translation. Provided eIF2 has a greater affinity for GDP, eIF2B works towards promoting guanine nucleotide exchange. The formed eIF2-GTP is not stable unless Met-tRNA$_i^{Met}$ joins to form the ternary complex (figure 1.2.G). This is one of the rate-limiting steps of translation initiation (Gomez *et al.*, 2002).

The scanning model for translation initiation states that both position (proximity to the 5' end) and context contribute to the selection of the initiation site. However, the first AUG rule is not always fulfilled, leading to an additional layer on gene expression control. According to the scanning model, the presence of upstream open reading frames (uORFs) can prevent translation of the major coding region by diverting ribosomal subunits from the authentic initiation codon, unless levels of eIF2-met-tRNA-GTP are low. Translation of downstream ORFs is possible by either leaky scanning or reinitiation (Kochetov *et al.*, 2008; Hinnebusch, 2011; Barbosa *et al.*, 2013). In leaky scanning, the 40S subunit can bypass AUG that are not in an optimal sequence context (a purine at position -3 and a guanine at position +4). This phenomenon can occur when nucleotides (nts) around the main AUG are far from the optimal context, when another AUG triplet is located closely after it, when a stop codon in the same reading frame is located closely after AUG, or if AUG is too close to the cap structure (Kozak, 2002). As far as reinitiation is concerned, following translation of a short ORF, the 40S ribosomal subunit remains connected to the mRNA after termination at the uORF stop codon and resumes scanning down the mRNA until it acquires another eIF2-met-tRNA-GTP in order to start protein synthesis at a downstream AUG (Kochetov *et al.*, 2008). Usually, the presence of a uORF inhibits initiation at downstream AUG and it often appears that the sole function of the uORF is to regulate the expression of the main ORF of the mRNA — as is the case of human erythropoietin in response to hypoxia (Barbosa and Romão, 2014) or the case of human

hemojuvelin, whose expression is tightly regulated by two uORFs that respond to iron overload in hepatic cells (Onofre *et al.*, 2015).

1.2.2. Elongation

After initiation, the 80S ribosome consists of large and small ribosomal subunits, mRNA, and Met-tRNA$_i^{Met}$ in the P site. The next codon to be translated is in an open ribosomal position called the Acceptor (A) site. A number of soluble protein synthesis factors engage the ribosome during the eukaryotic translation elongation cycle (figure 1.3). The latter is mediated by the concerted actions of: eukaryotic elongation factor (eEF) 1A, a G-protein that binds and delivers aminoacyl-tRNA to the A-site of an elongation ribosome harbouring a growing nascent peptide chain; eEF1B, a multi-subunit GEF composed of subunits α, β, and γ, that catalyses the exchange of GDP for GTP on eEF1A; and eEF2, which facilitates ribosomal translocation following each round of peptide bond formation (Kapp and Lorsch, 2004b; Taylor *et al. in* (Mathews *et al.,* 2007).

Once the 60S ribosomal subunit is properly assembled, an 80S ribosome is placed on an mRNA with the anticodon of Met-tRNA$_i$ base-paired with the start codon in the P site. The second codon of the ORF is present in the A site of the ribosome awaiting to be bound to the cognate aminoacyl-tRNA. The eEF1A-GTP binds and recruits aminoacyl-tRNA to the A site of the ribosome in a GTP-dependent manner. Codon recognition by tRNA triggers GTP hydrolysis by eEF1A, releasing the factor and enabling the aminoacyl-tRNA to be accommodated into the A site (Carvalho *et al.*, 1984; Gromadski *et al.*, 2007; Dever and Green, 2012a). The small ribosomal subunit decodes the incoming anticodon of the eEF1A-GTP-aminoacyl-tRNA ternary complex and ensures the formation of a proper codon-anticodon match. Only the correct codon-anticodon match results in a conformational change in the head of the small subunit, leading to a closed conformation. This conformational change in the small subunit, induced by the delivery of the cognate aminoacyl-tRNA, leads to GTP hydrolysis of the ternary complex. Hydrolysis is stimulated by a region of the large subunit named *GTP-associated centre*. The eEF1A-GDP complex is released from the ribosome, leaving the aminoacyl-tRNA in the A site. Spontaneous GDP dissociation from eEF1A is slow, and the eEF1Bαβγ complex stimulates the exchange of GDP for GTP, maintaining the level of active eEF1A-GTP (Taylor *et al. in* Mathews *et al.*, 2007; Agirrezabala and Frank, 2009). The large subunit catalyses peptide bond formation between the P-site tRNA and the incoming aminoacyl moiety of the A-site tRNA. The Exit (E)-site tRNA leaves the ribosome, with the assistance of L1 ribosomal protein during each round of elongation (Taylor *et al. in* Mathews *et al.*, 2007).

Following peptide bond formation, a ratchet-like motion of the ribosomal subunits triggers movement of the tRNA into the so-called hybrid P/E and A/P states with the acceptor ends of the tRNA in the E and P sites and the anticodon loops remaining in the P and A sites, respectively (Agirrezabala and Frank, 2010; Dever and Green, 2012a).

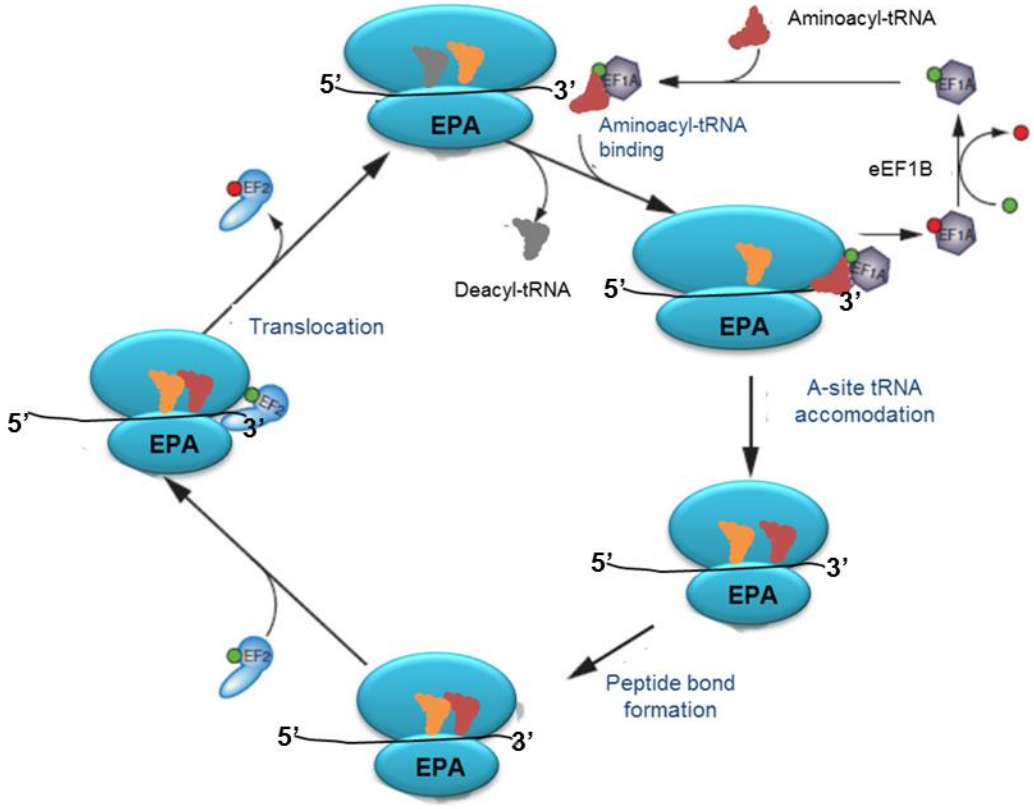


Figure 1.3 — Model of the eukaryotic translation elongation pathway. Starting at the top, a eukaryotic elongation factor (eEF) 1A-GTP-aminoacyl-tRNA ternary complex binds the aminoacyl-tRNA to the 80S ribosome with the anticodon loop of the tRNA in contact with the mRNA in the aminoacyl (A) site of the small subunit. Following the release of eEF1A-GDP, the aminoacyl-tRNA is accommodated into the A site, and the eEF1A-GDP is recycled into eEF1A-GTP by the exchange factor eEF1B. Peptide bond formation is accompanied by transition of the A- and peptidyl (P)-site tRNA into hybrid states with the acceptor ends of the tRNA moving to the P and exit (E) sites, respectively. Binding of eEF2-GTP promotes translocation of the tRNA into the canonical P and E sites, and is followed by the release of eEF2-GDP, which, unlike eEF1A, does not require an exchange factor. The ribosome is now ready for the next cycle of elongation with release of the deacylated tRNA from the E site and binding of the appropriate eEF1A-GTP-amynoacyl-tRNA to the A site. Adapted from Dever and Green, 2012a.

Translocation of the tRNA to the canonical E and P sites is mediated by eEF2. After the addition of an amino acid to the nascent peptide chain, the tRNA bearing that polypeptide moves from the A site into the P site on the ribosome as it moves one codon along the mRNA (Taylor *et al. in* Mathews *et al.*, 2007). Binding of eEF2 in complex with GTP is thought to stabilise the hybrid state and promote rapid hydrolysis of GTP. Conformational changes in eEF2 accompanying GTP hydrolysis and Pi release are thought to alternatively unlock the ribosome allowing tRNA and mRNA movement and then lock the subunits in the post-translocation state (Agirrezabala and Frank, 2010; Dever and Green, 2012a).

1.2.3. Termination and recycling

Termination in eukaryotes is catalysed by two protein factors, eRF1 and eRF3, that appear to collaborate in the process by binding to the A site as an eRF1-eRF3-GTP complex (figure 1.4) (Alkalaeva *et al.*, 2006). At the end of every message there is a stop [nonsense] codon that is not read by tRNA. Following completion of the elongation phase of protein synthesis, the ribosome is brought into its pretermination complex when such a codon is translocated into its A site. These codons are instead read by class I eukaryotic release factors (eRF). In eukaryotes, this class consists only of eRF1, which is responsible for high-fidelity recognition of all three universally conserved stop codons — UAA, UAG and UGA — and peptidyl-tRNA hydrolysis (Dever and Green, 2012a). The modes of action of tRNA during peptide elongation and of eRF1 during termination are similar. In both cases, codons in the ribosomal A site are recognised with high precision, which results in a distal chemical event: either the transfer of a peptide from the P-site to the A-site tRNA or the disruption of the bond between a finished peptide and the P-site tRNA. Accordingly, eRF1 can be viewed as a functional mimic of tRNA (Moffat and Tate, 1994). It can, therefore, be considered a tRNA-shaped protein factor composed of three domains (Song *et al.*, 2000). The amino-terminal domain is responsible for codon recognition and contains a distal loop with a highly conserved NIKS motif (positions 61-64 in human eRF1) (Chavatte *et al.*, 2002; Bulygin *et al.*, 2010) that has been proposed to decode stop codons through codon:anticodon-like interactions (Song *et al.*, 2000; Dever and Green, 2012a). Yet, other regions of eRF1 also appear to contribute to stop codon recognition including the YxCxxxF motif (positions 125-131). Moreover, two more invariant residues, Glu-55 and Tyr-125 (human eRF1 numbering), potentially involved in codon recognition, have been identified — this suggests that a three-dimensional network of amino acids may be responsible for stop-codon reading by eRF1 (Kolosov *et al.*, 2005). Recent cryo-electron microscopy (cryo-EM) data (Brown *et al.*, 2015) revealed structures at 3.5-3.8Å resolution of mammalian ribosomal complexes containing eRF1 interacting with each of the three stop codons in the A site. Binding of eRF1 flips nucleotide A1825 of 18S rRNA so that it stacks onto the second and third stop codon bases. This configuration pulls the fourth position base into the A site, where it is stabilised by stacking against G6 of 18S rRNA. Thus, eRF1 exploits two rRNA nucleotides also used during tRNA selection to drive mRNA compaction. In this compact mRNA conformation, stop codons are favoured by a hydrogen-bonding network, formed between rRNA and essential eRF1 residues, that constrains the identity of the bases. In conclusion, these structures show how stop codons are specifically selected by eRF1. At the +1 position, only uridine can form the network of interactions with the NIKS motif. The flipping of A1825 results in its stacking onto the +2 and +3 bases of a distorted mRNA so that they are decoded as a single unit. This solves the puzzle of how guanosine can occur at either the +2 or +3 position, but not at both: two successive guanosines would lead to repulsion between their O6 atoms and between them and Glu55. Logically, two consecutive purines can occur, since this premise specifically excludes consecutive guanosines (Brown *et al.*, 2015).

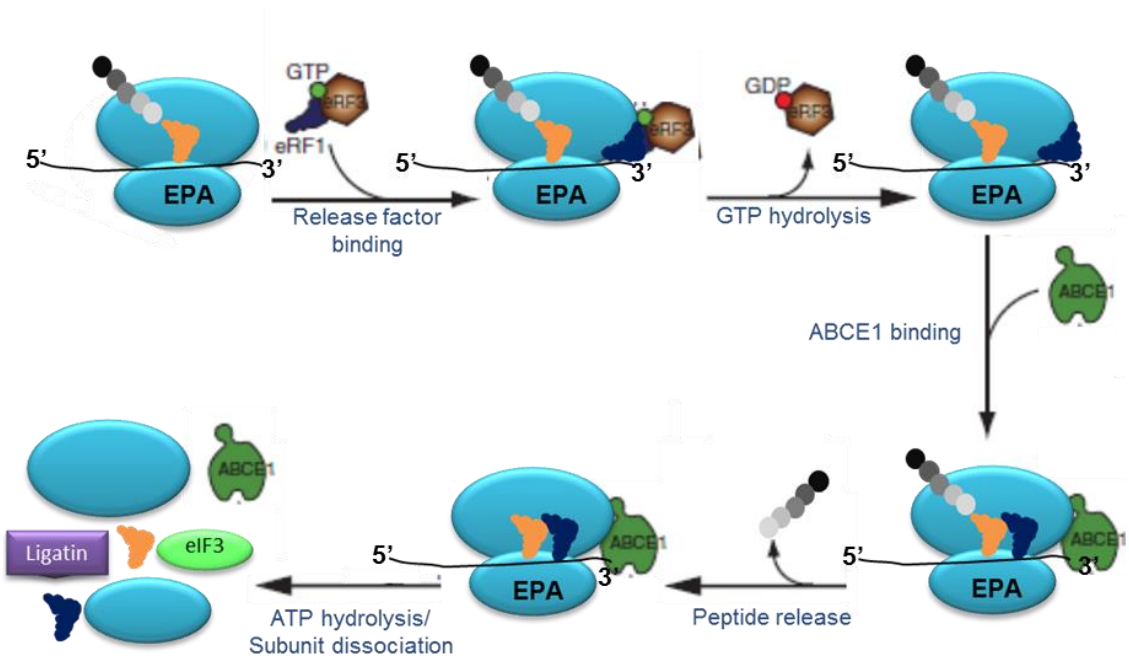


Figure 1.4 — Model of the eukaryotic translation termination and recycling pathways. On the recognition of a stop codon, the eRF1-eRF3-GTP ternary complex binds to the A site of the ribosome in a pre-accommodated state, GTP hydrolysis occurs, and eRF3 is released. ABCE1 binds and facilitates the accommodation of eRF1 into an optimally active configuration. Peptide release is catalysed by an ATP-independent activity of ABCE1. ATP hydrolysis of ABCE1 is coupled to subunit dissociation, and deacylated tRNA and mRNA dissociate from the isolated small subunits following *recycling* — an event enhanced by ligatin. Separated subunits are ready to bind again to initiation factors for subsequent rounds of initiation or reinitiation. Adapted from (Dever and Green, 2012a).

At the tip of eRF1 central domain there is an evolutionarily conserved GGQ triplet motif, similar to those that occur in bacterial RF1 and RF2 (class I release factors, as is eRF1) (Frolova *et al.*, 1999; Mora *et al.*, 2003). It induces hydrolysis of the ester bond in peptidyl-tRNA (Frolova *et al.*, 1999), and may mimic the CCA end of the tRNA. Furthermore, mutations in the GGQ motif greatly reduce termination efficiency and cell viability. (Song *et al.*, 2000; Mora *et al.*, 2003; Kong *et al.*, 2004). Considering this line of thought, GGQ is a successful chemical solution for catalysing peptidyl-tRNA hydrolysis in the highly conserved, RNA-rich peptidyl transferase centre (Dever and Green, 2012a).

As for the C-terminal domain of eRF1, it is involved in facilitating interactions with the class II release factor eRF3 (Merkulova *et al.*, 1999). eRF3 is a member of the GTPase family. Such enzymes display a transition from their GDP-bound state to their GTP-bound state in which they accomplish their task. This is followed by GTP hydrolysis and return to their GDP-bound form. The GDP-to-GTP exchange is often aided by a GEF, and GTP hydrolysis is sometimes triggered by a GTPase-activating protein (Ehrenberg *et al.* in Mathews *et al.*, 2007). eRF3 is essential for viability and its GTPase activity depends strictly on the presence of both the ribosome and eRF1. Free eRF3 forms a complex with eRF1, which is stabilised by the presence of GTP but not GDP (Ehrenberg *et al.* in Mathews *et al.*, 2007). eRF3 has a negligible, if any,

intrinsic GTPase activity, inasmuch as it is profoundly stimulated by the joint action of eRF1 and the ribosome. Separately, neither eRF1 nor the ribosome displays this effect, thus functioning eRF3 as a GTPase only in the context of a quaternary complex — the aforementioned eRF1-eRF3-GTP ternary complex of translation termination together with the ribosome (Frolova *et al.*, 1996). When the ternary complex joins the ribosome, it triggers GTP hydrolysis (Frolova *et al.*, 1996), and eventually leads to the deposition of the central domain of eRF1 in the peptidyl transferase centre. In this scenario, eRF3 plays a role in controlling delivery of a tRNA-like molecule into the peptidyl transferase centre (Dever and Green, 2012). In this regard, the GTPase activity of eRF3 is required to couple the recognition of translation termination signals by eRF1 to efficient polypeptide chain release (Salas-Marco and Bedwell, 2004)

Once the polypeptide chain has been released, the recycling process takes place. At this stage, the 80S ribosome is still bound to the mRNA, the now deacylated tRNA, and likely eRF1, which means the ribosomal subunits must be dissociated and the mRNA and deacylated tRNA released to regenerate the necessary components for subsequent rounds of translation (Dever and Green, 2012). Unlike in bacteria, eRF3 does not appear to promote the departure of the class I release factor eRF1, so, the latter remains associated with the ribosomal complex following termination (Pisarev *et al.*, 2007). This post-termination complex containing bound eRF1 and a deacylated tRNA (potentially in an unratcheted state) is what must be targeted by the recycling machinery in eukaryotes. Initial reports argued that eRF3 might play an active role in recycling in higher eukaryotes (Pisarev *et al.*, 2007), instead of functioning merely to stabilise dissociated subunits by directly binding to the subunit interface (Dever and Green, 2012). However, subsequent studies identified the multifunctional ABC-family protein ABCE1 (Pisarev *et al.*, 2010; Barthelme *et al.*, 2011), a highly conserved cytosolic ATPase essential to life (Dong *et al.*, 2004), as a likely candidate for promoting ribosomal recycling. It is proposed to somehow convert the chemical energy from ATP hydrolysis into mechanical motions that can separate subunits, thus aiding the intrinsic ribosome recycling activity of the canonical release factors (Dever and Green, 2012). ABCE1 has also been shown to directly promote the rate of peptide release by eRF1-eRF3, in an ATP hydrolysis-independent manner (Shoemaker and Green, 2011).

Summarising the events taking place during translation termination and recycling under normal conditions, the eRF1-eRF3 complex recognises stop codons, and GTP hydrolysis by eRF3 allows separation of the GDP form from the factor. A certain kind of accommodation takes place when the GGQ end of the release factor swings into the catalytic centre of the large subunit. Peptide release is, then, catalysed, stimulated by an ATP-independent activity of ABCE1. Finally, ATP hydrolysis on ABCE1 is coupled to subunit dissociation. Deacylated tRNA and mRNA are likely dissociated from the isolated small subunits following recycling, an event enhanced by ligatin[2], a factor that, together with the pair of proteins MCT-1 and DENR[3], can promote release of deacylated tRNA and mRNA from recycled 40S subunits after ABCE1-

[2]Ligatin is a member of the eIF2D family of initiation factors that is able to deliver tRNA to the P site of the eukaryotic ribosome in a GTP-independent manner (Skabkin *et al.*, 2010).

[3]These proteins are homologous to N-terminal and C-terminal regions of Ligatin, respectively (Skabkin *et al.*, 2010).

mediated dissociation of post-termination ribosomes (Skabkin *et al.*, 2010). Separated subunits are then ready to bind again to available initiation factors that prepare themselves for subsequent rounds of initiation or reinitiation (Pisarev et al 2007).

1.2.3.1. Premature termination and triggering of nonsense-mediated mRNA decay

Occurrence of mutations in the DNA sequence of a specific gene eventually results in the loss of production of the corresponding protein, and is among the major causes of inherited diseases. One of the most common types of mutation inactivates gene function by promoting premature translation termination (Jacobson and Izaurralde in Mathews *et al.*, 2007). Nonsense mutations result in stop codons (UAA, UAG, or UGA), leading to the termination of polypeptide elongation and, generally, to the triggering of a cellular surveillance mechanism known as nonsense-mediated mRNA decay (NMD, figure 1.5) (Jacobson and Izaurralde *in* Mathews *et al.*, 2007). NMD is tightly coupled to translation, since NMD machinery should recognise the translation termination codon on the mRNA as a premature termination codon (PTC) before mRNA degradation. Nevertheless, premature termination is a mechanistically different event from normal termination as it appears to be less efficient, thus reflecting different messenger ribonucleoprotein (mRNP) complexes at PTCs and normal termination codons (NTCs). Such inefficiency of premature termination is triggered by the mRNP structure downstream of a PTC which, in turn, may lead to poor release factor binding at the A site or to slow dissociation of the release factors after peptide hydrolysis (Celik *et al.*, 2015). These deficiencies are thought to be fixed by the recruitment of UPF1 to the premature termination complex. Activation of its ATPase and helicase activities promote ribosome reutilization and trigger NMD and the subsequent nascent polypeptide degradation (Kuroha *et al.*, 2009). More specifically, NMD is tightly coupled to the pioneer round of translation that is dependent on CBP80/20. Newly synthesized mRNA is exported from the nucleus to the cytoplasm with CBP80/20 bound to the cap structure at the 5' end of the mRNA. During the export of newly synthesized mRNA, CBP80/20 at the 5' end of the mRNA exposed to the cytoplasm recruits ribosomes to direct the first round of translation. All types of mRNA are believed to be subject to this mode of translation, because all mRNA that are completely processed in the nucleus contain a cap structure bound by CBP80/20 (Hwang and Kim, 2013). In most normal mRNA, the translation termination codon resides in the last exon of the gene. Consequently, all deposited exon junction complexes (EJCs)[4] are dissociated from the mRNA during the elongation step of the pioneer round of translation. In such cases, the mRNA is stable due to the lack of EJCs downstream of the translation termination codon.

However, in the case of mRNA harbouring PTCs more than 50-55 nucleotides upstream of

[4]Exon junction complexes are protein complexes that deposit at the exon-exon junctions formed during splicing of the pre-mRNA molecule. They consist of a stable heterotetramer core containing eIF4A-III bound to an ATP analogue, as well as the additional proteins Magoh and Y14 (Andersen, 2006). This core serves as a binding platform for other factors necessary for mRNA biogenesis (Tange *et al.*, 2004).

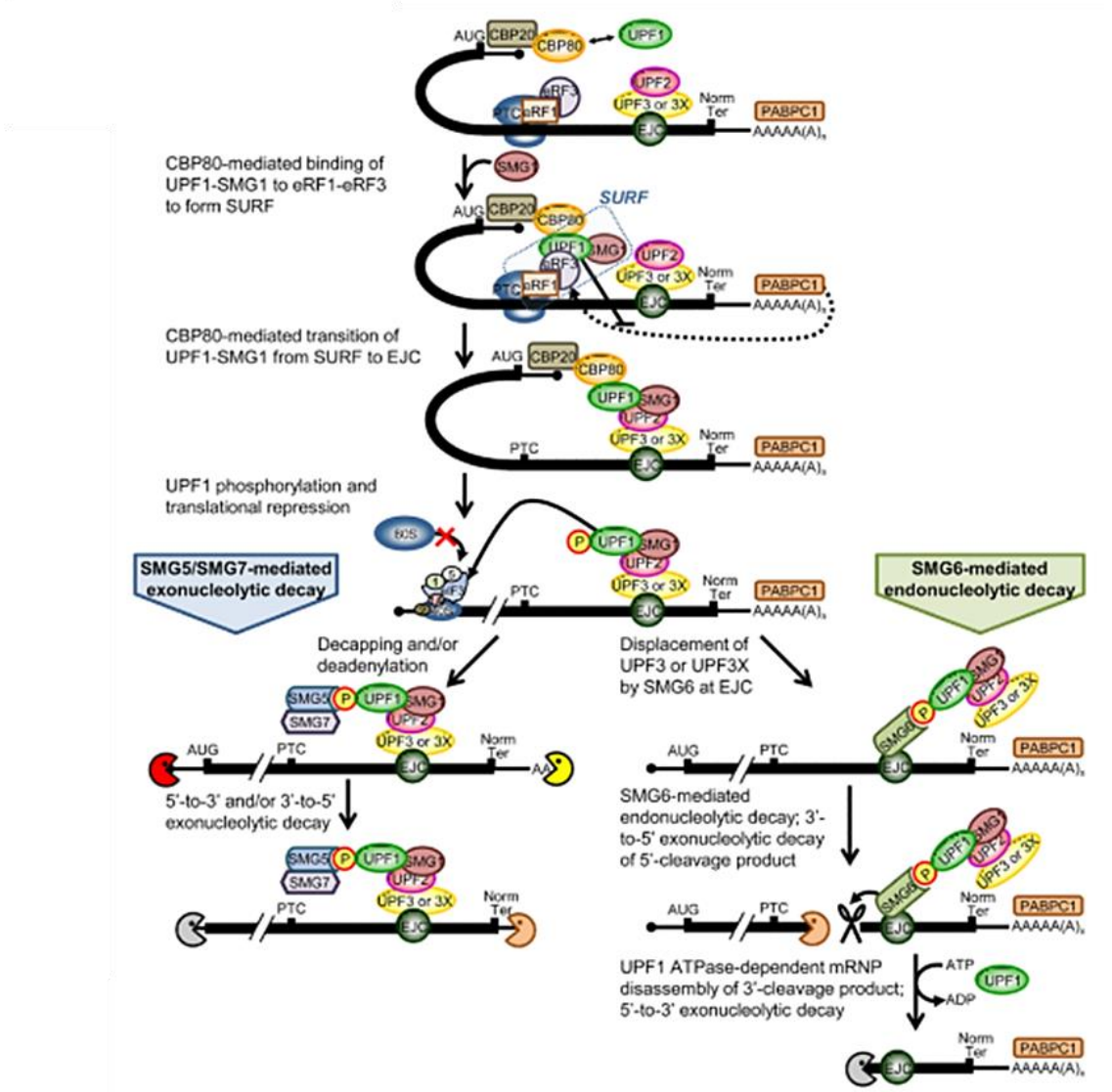


Figure 1.5 — The exon-junction complex (EJC)-dependent nonsense-mediated mRNA decay (NMD) model. NMD is a consequence of premature termination codon (PTC) recognition during the pioneer round of translation. This round utilizes newly synthesized mRNA bound by the cap-binding protein heterodimer cap-binding protein (CBP) 80-CBP20 and, providing the mRNA is derived from splicing, at least one exon-junction complex (EJC) situated ~ 20–24 nucleotides upstream of such a junction. The direct but weak, or transient, interaction of CBP80 with the central NMD factor UPF1 promotes at least two steps during NMD. The first step is the joining of UPF1 and its kinase SMG1 to eRF1 and eRF3, at a PTC, to form the SURF complex. During NMD, this step is thought to compete effectively with joining of the PABPC1 to eRF3, the latter of which is specified as a dotted line. The second is the joining of UPF1 and SMG1, presumably from SURF, to a downstream EJC, which leads to UPF1 phosphorylation by SMG1. SMG5 and SMG7 form a complex with phosphorylated UPF1, as does SMG6. It is unclear whether SMG5/SMG7 and SMG6 bind multiple phosphates on the same UPF1 molecule or, as shown, different phosphorylated UPF1 molecules. In favour of the first possibility, SMG6 co-immunoprecipitates with SMG5 and SMG7 in an RNAe A-resistant manner. Since SMG7-mediated mRNA decay occurs independently of SMG6, it is plausible that SMG5/SMG7-mediated NMD leads to deadenylation and/or decapping followed, respectively, by exosome-mediated 3′–5′ and XRN1-mediated 5′–3’ exonucleolytic activities. An alternative or additional mRNA degradation pathway involves SMG6, whose binding to hyperphosphorylated UPF1 competes with UPF3X and may replace the interaction of UPF3X with Y14-MAGOH EJC constituents. The endonuclease activity of SMG6 cleaves the NMD substrate into 5′- and 3′-cleavage products. Activation of the RNA-dependent ATPase activity of UPF1 subsequently results in the XRN1-mediated 5′–3′ decay of the 3′ fragment, which presumably depends on UPF1 helicase activity. PAPBC1, poly(A) binding protein C. Adapted from (Hwang and Maquat, 2011).

the last exon-exon junction, EJCs will remain downstream of the PTC, which serves as a molecular marker to induce NMD (Maquat, 2005; Hwang and Kim, 2013). The terminating ribosome at a PTC during the pioneer round of translation recruits the SURF[5] complex, which communicates with the EJC downstream that PTC via an interaction between UPF1 in the SURF complex and UPF2 in the downstream EJC (Kashima *et al.*, 2006).

The NMD pathway in human cells comprises several factors, such as the UPF proteins. These constitute the core NMD machinery; functionally, UPF1 is the most important factor for NMD (Perlick *et al.*, 1996; Culbertson and Leeds, 2003). UPF1 regulates the degradation of NMD-sensitive mRNA and the remodelling of the mRNA surveillance complex through phosphorylation/dephosporylation cycles. In detail, UPF1 is phosphorylated by SMG1[6] at specific serine residues in its C-terminus serine/glutamine motifs (Denning *et al.*, 2001; Yamashita *et al.*, 2001), which facilitates the assembly of degradation factors and, consequently, triggers the degradation of NMD-sensitive mRNA (Cho *et al.*, 2009). UPF1 phosphorylation triggers eIF3-dependent translational repression during the process of NMD. Phosphorylated but not hypophosphorylated UPF1 directly interacts with eIF3 in order to prevent the joining of 60S ribosomal subunit, thus inducing translational repression (Isken *et al.*, 2008). Moreover, phosphorylated UPF1 also interacts with SMG5, SMG6, SMG7, and human proline-rich nuclear receptor coregulatory protein 2 (PNRC2) and then triggers the degradation of NMD-sensitive mRNA. The association of SMG6 with phosphorylated UPF1 triggers RNA degradation by SMG6 endonuclease (SMG6-mediated endonucleolytic decay). In contrast, the association of heterodimer SMG5/SMG7 with phosphorylated UPF1 triggers RNA degradation by deadenylase and decapping enzyme (SMG5/SMG7-mediated exonucleolytic decay). On the other hand, PNRC2 interacts with UPF1 and decapping mRNA 1a (DCP1a), a component of the decapping complex, and triggers 5'–3' exonucleolytic decay (Cho *et al.*, 2009; Mühlemann and Lykke-Andersen, 2010; Okada-Katsuhata *et al.*, 2012). A study by Franks *et al.* (2010) revealed that ATP hydrolysis by UPF1 leads to the disassembly of mRNP complex targeted to NMD, which is critical in the final step of RNA degradation, and is involved in the recycling of NMD factors and other RNA-binding proteins derived from NMD substrates, and UPF1 ATPase activity plays an important role in ATPase-dependent mRNP disassembly in NMD (Imamachi, 2012). Furthermore, the ATPase cycle of the superfamily 1 (SF1) helicase UPF1 is required for mRNA discrimination during NMD. Mutations affecting the UPF1 ATPase cycle disrupt the mRNA selectivity of UPF1, leading to indiscriminate accumulation of NMD complexes on both NMD target and non-target mRNA (Lee *et al.*, 2015). In addition, two modulators of NMD — translation and termination codon-proximal poly(A) binding protein — depend on the ATPase activity of UPF1 to limit UPF1-non-target association. Preferential ATPase-dependent dissociation of UPF1 from non-target mRNA *in vitro* suggests that selective release of UPF1 contributes to the ATPase dependence of UPF1 target discrimination. Given the prevalence of helicases in RNA regulation, ATP hydrolysis may be a widely used activity in target RNA

[5]SURF complex is composed of: Suppressor with Morphological effect on Genitalia (SMG) 1; up-frameshift (UPF) 1; eRF1; and eRF3 (Kashima *et al.*, 2006).

[6]SMG1 is a phosphatidylinositol 3 kinase-related protein kinase (PIKK).

discrimination (Lee *et al.*, 2015). Thus, UPF1 is a highly processive RNA helicase and translocase with RNP remodelling activities, as Fiorini *et al.* (2015) demonstrated. UPF1 efficiently translocates through double-stranded structures and protein-bound sequences, demonstrating that it is an efficient RNP complex remodeler. Hence, UPF1, once recruited onto NMD mRNA targets, can scan the entire transcript to irreversibly remodel the mRNP, facilitating its degradation by the NMD machinery (Fiorini *et al.*, 2015). The remodelling activity of UPF1, combined with its remarkable processivity, may also serve to rearrange the mRNP far downstream the stop codon, paving the way for RNA degradation. Each mRNA is packed in a specific particle made of a complex set of ribonucleoproteins (RBP) essential for fine-tuning mRNA localisation, translation and decay (Müller-McNicoll and Neugebauer, 2013). Long-range remodelling by UPF1 may irreversibly affect the fragile equilibrium of mRNP (Mühlemann and Jensen, 2012) and push the mRNA towards degradation (Fiorini *et al.*, 2015).

The existence of a pathway that promotes rapid decay of nonsense-containing mRNA is not restricted to those derived from genes in which a mutation or an error in transcription or processing has given rise to a PTC. Instead, there are several classes of NMD substrates, including: inefficiently spliced pre-mRNA that enter the cytoplasm with their introns intact; mRNA in which a leaky scanning ribosome bypasses the initiator AUG and begins translation further downstream; some mRNA containing uORFs; transcripts with extended 3'UTR; mRNA subject to +1 frameshifting, bicistronic mRNA and some non-coding RNA. Thus, these substrates can all be considered targets of a quality control system that eliminates RNA capable of giving rise to potentially deleterious translation products (Jacobson and Izaurralde in Mathews *et al.*, 2007). Unfortunately, many nonsense mutations have still been implicated in hundreds of inherited diseases, including haemoglobinopathies, cystic fibrosis, Duchenne muscular dystrophy, lysosomal storage disorders, skin disorders, and various cancers. Potential therapeutic approaches to promote read-through of the nonsense codon have been investigated. Some of these therapies include aminoglycoside antibiotics, because they can promote the read-through of the PTC and therefore increase the expression levels of some mRNA. A caveat of these substances is the fact that they promote general inhibition of NMD and not of a specific PTC. A more sensitive approach includes the use of small molecules like Ataluren[7] (Welch *et al.*, 2007). These can minimise undesirable side-effects and, furthermore, highlight the notions that the termination and mRNA decay functions of NMD are separable, and that premature termination is not the same biochemical event as normal termination (Jacobson and Izaurralde *in* Mathews *et al.*, 2007).

[7]Ataluren, formerly known as PTC124, is a pharmaceutical drug approved for the treatment of Duchenne muscular dystrophy in the European Union. It can potentially be used to treat other genetic disorders caused by nonsense mutations, such as cystic fibrosis. Ataluren appears to be most effective for the read-through of the stop codon UGA (Welch *et al.*, 2007).

1.3. Non-canonical translation initiation mechanisms

The scanning model of translation is widely accepted as the most frequent mechanism of translation initiation in eukaryotes. However, in conditions impairing this mechanism, several proteins are able to maintain their expression via non-canonical mechanisms of translation initiation that can occur under stress conditions. These mechanisms can be either cap-dependent or cap-independent. Several proteins are able to maintain their expression levels under conditions that impair the recognition of the cap structure or, to a lesser extent, the proper scanning of the 5'UTR.

Below, we provide an overview of some non-canonical mechanisms of translation initiation that are alternatives to the scanning model of translation initiation.

1.3.1. Scanning-independent mechanisms of translation initiation

Although most eukaryotic mRNA are translated via the canonical scanning mechanism, there are mRNA that are dependent on the m^7G cap but avoid scanning. These special mechanisms direct protein synthesis of mRNA with either an extremely short or a highly complex 5'UTR and are advantageous under specific physiological settings (Haimov *et al.*, 2015).

1.3.1.1. Ribosome shunting

Although in some mRNA the presence of uORFs and hairpins can inhibit translation, in others they may be bypassed. Ribosome shunting (figure 1.6) is an atypical mode of ribosomal movement in eukaryotic translation systems. It explains how obstacles in a 5'UTR can be bypassed in mRNA containing elements that function as shunt sites (Ogawa, 2013). According to this model, ribosomal subunits are recruited to the mRNA either via the eIF4F complex at the cap structure or through internal mRNA elements; these subunits then recruit the translation machinery through direct interactions — base-pairing between rRNA and mRNA or binding to ribosomal proteins — or indirectly — binding to initiation factors or other proteins that can interact with the translation machinery (Chappell *et al.*, 2006). Such recruitment sites would effectively increase the local concentration of 40S subunits and associated factors. This would enhance shunting by increasing the likelihood of interactions between ribosomal subunits and other accessible recruitment sites in the mRNA and might also increase the likelihood of interactions between the initiator tRNA-Met and the initiation codon itself (Chappell *et al.*, 2006). Although most examples of ribosome shunting are found in virus mRNA, there are several cellular mRNA that use this mechanism. For instance, cellular inhibitor of apoptosis protein (cIAP2) mRNA is exclusively translated through a mechanism of ribosome shunting. The 43S scans only a short distance of the 5'UTR and, then, is shunted across the base of a highly stable RNA stem. This allows the ribosome to bypass 62 of the 64 uAUG present in the cIAP2 5'UTR and places the shunted ribosome just upstream the cIAP2 start codon.

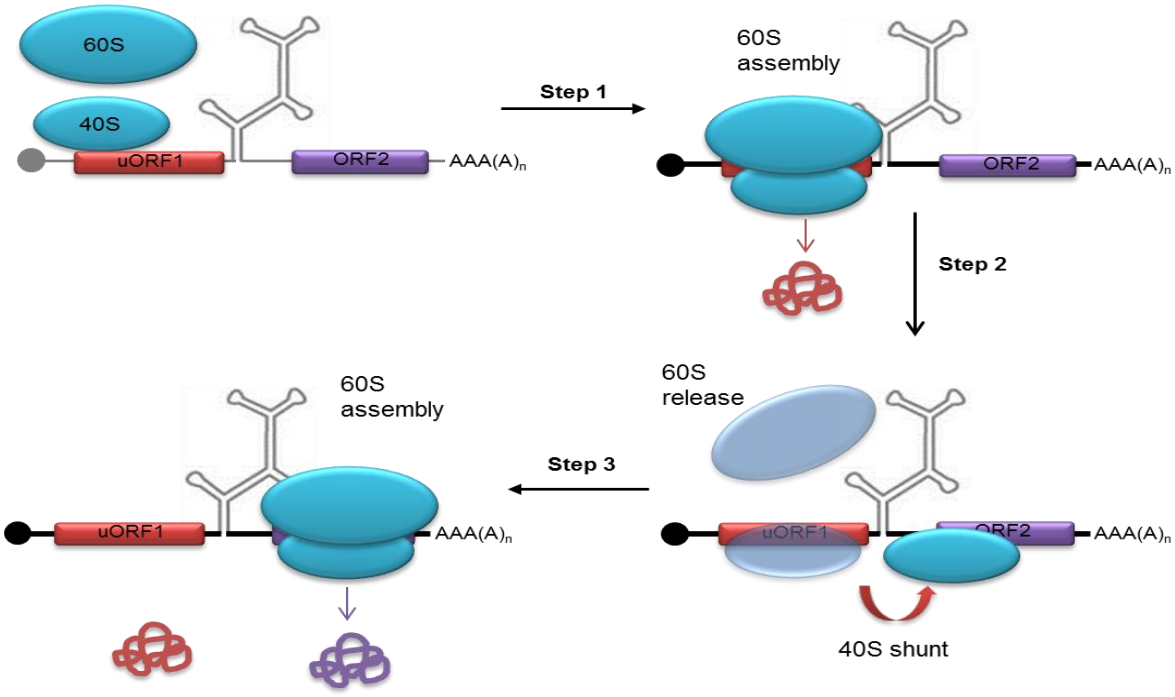


Figure 1.6 — Ribosome shunting. The AUG of the first open reading frame (ORF) is recognised by the 40S subunit, and then, after proper assembly of the 60S subunit, the corresponding peptide is produced (step 1). The 60S subunit is released and the 40S subunit is shunted to the next AUG by means of interactions between the mRNA that is about to be translated and the rRNA of the 40S (step 2). Again, the 60S subunit is assembled and the translation of the downstream ORF occurs (step 3).

This shunting mechanism ensures cIAP2 translation during stress conditions that block canonical scanning-dependent translation initiation (Sherrill and Lloyd, 2008).

Another of the few examples of this mechanism in cellular mRNA is the β-secretase enzyme (BACE1), which is involved in the formation of Aβ-amyloid plaques in Alzheimer's disease patients (Rogers *et al.*, 2004). Although BACE1 5'UTR has four uORFs and highly stable secondary structures, it efficiently regulates the cap-dependent translation of a luciferase mRNA (Rogers *et al.*, 2004). These findings were perceived as evidence of ribosomal shunting. However, other studies demonstrated a substantial inhibition by BACE uAUG (De Pietri Tonelli, 2004; Mihailovich *et al.*, 2007). These contradictions have been interpreted by Koh and Mauro (2009) as resulting from the involvement of different expression systems, i.e., when transcription occurs at the nucleus through CMV reporter plasmid, the inhibitory effect of the uAUG is small, while *in vitro* or cytoplasmic transcription rendered the uAUG highly inhibitory, and the folding of the BACE1 5'UTR is therefore dependent on the site of transcription.

1.3.1.2. Translation of mRNA with short 5'UTR

Translation of eukaryotic mRNA with a short 5'UTR is a poorly understood field and is usually considered a non-efficent process, leading to leaky scanning (Kozak, 1991). A non-canonical mechanism driven by a translation initiator of short 5'UTR (TISU) element

(SAASATGGCGGC) operates in higher eukaryotes in mRNA bearing extremely short 5'UTR (Dikstein, 2012). TISU is strictly located downstream of the transcription start site (TSS) from position +5 up to position +30 relative to the TSS, being quite close to the m^7G. Thus, the TISU mRNA have an unusually short 5'UTR with a median length of 12 nts (Elfakess and Dikstein, 2008). Detailed comparison of TISU to the well-characterised and strong Kozak element established it as an element optimised to direct efficient translation initiation from mRNA with an extremely short 5'UTR (Elfakess and Dikstein, 2008; Elfakess *et al.*, 2011; Dikstein, 2012).

Depletion of eIF1, whether *in vitro* or in cell, substantially diminishes translation directed by the TISU element but not by an AUG in a strong context, which indicates eIF1 facilitates TISU-mediated translation and is a major player in its potency (Sinvani *et al.*, 2015). Analysis of additional factors revealed unexpectedly that eIF4GI — but not eIF3 — acts similarly to eIF1, suggesting they cooperate in TISU-mediated translation (Sinvani *et al.*, 2015).

The TISU element is highly prevalent among genes associated with mitochondrial activities and energy metabolism, including the two catalytic subunits of AMP-activated protein kinase (AMPK), a highly conserved sensor of the cellular energy status (Zhang *et al.*, 2009). Examination of TISU activity under conditions of low energy availability showed that AMPK remains translationally active. The resistance to energy stress is granted by the TISU sequence (Sinvani *et al.*, 2015). Thus, TISU-mediated initiation enables continuous translation of proteins under conditions of energy shortage, allowing cells to cope with the stress.

1.3.2. Cap-independent mechanisms of translation initiation

Sometimes, cap structure recognition does not occur. This can be due to external cellular stimuli that lead to i) hypophosphorylation of 4E-BP — due to serum starvation or picornavirus infection —, which allows it to compete with eIF4G for the binding to eIF4E, hence making eIF4F levels become limiting; or ii) phosphorylation of the α subunit of eIF2 by kinases such as PKR[8], which causes it to bind with stronger affinity to its GEF eIF2B, resulting in low levels of ternary complex (King *et al.*, 2010).

1.3.2.1. Internal ribosome entry site-mediated translation

In 1988, Pelletier and Sonenberg discovered that some viral mRNA from polioviruses are translated by a mechanism that enables ribosomes to initiate translation effectively on highly structured regions located within the 5'UTR. Up until then, the only known mechanisms of translation initiation were dependent on the binding of eIF4E to the 5' cap of mRNA, but these authors have shown that some mRNA have a mechanism to bypass that need. This mechanism was called *internal ribosome entry site* (IRES) *-mediated translation* (figure 1.7) (Pelletier and Sonenberg, 1988; Jang *et al.*, 1988). This mechanism of translation initiation is generally independent of mRNA 5′ cap structure recognition, but may either involve scanning — in search for an initiation codon —, or direct recruitment of the 40S ribosomal subunit to the vicinity of the initiation codon. Ribosomal 40S subunit recruitment can occur both in the complete absence of

[8]Double-stranded-RNA-dependent protein kinase

any other protein factors (dicistrovirus intergenic IRES) and with the aid of various combinations of canonical initiation factors (such as eIF4G and eIF3) and auxiliary proteins (reviewed in Lozano and Martínez-Salas, 2015). Since these discoveries, it has been found that many viruses contain IRES sequences in the 5'UTR of their mRNA that direct translation of viral proteins without the need of all translation initiation factors. These viruses are able to usurp the host eukaryotic translation machinery by cleaving factors necessary for canonical cap-dependent translation initiation, but dispensable for IRES-mediated translation. In this way, viral mRNA are able to outrun eukaryotic mRNA for ribosome binding, becoming, in many cases, the most abundant transcript being translated. The majority of viral IRES possess defined secondary and tertiary structures that allow their efficient interaction with the 40S ribosome. This interaction may be direct or partially indirect, requiring the assistance of both some canonical initiation factors and IRES *trans*-acting factors (ITAFs). ITAFs are known to assist in the recruiting of the 40S ribosomal subunit to the mRNA through specific interactions or stabilisation of specific active conformations of the IRES (figure 1.7) (Balvay *et al.*, 2009; Hellen, 2009; Jackson *et al.*, 2010; Jackson, 2013; Lozano and Martínez-Salas, 2015).

Several eukaryotic cellular mRNA can also be translated in an IRES-dependent way, in which there is cap-independent binding of the 40S ribosomal subunits (figure 1.7). The first cellular IRES in eukaryotes was discovered by Macejak and Sarnow (1991) in the mRNA encoding the immunoglobulin heavy-chain binding protein (*BiP*). Since this discovery, many transcripts containing IRES structures within their 5'UTR have been described, and it has been estimated that 10–15% of the cellular mRNA can be translated by an IRES-dependent mechanism (Spriggs *et al.*, 2008). Accordingly, recent data from a systematic screen for IRES-mediated translation activity have shown that about 10% of human 5'UTR have the potential to be translated by this cap-independent mechanism (Weingarten-Gabbay *et al.*, 2016). Apart from the most recent discoveries, they are included in the IRESite, which presents carefully curated experimental evidence of many viral and cellular IRES elements (Mokrejš *et al.*, 2010). Like viral IRES-containing mRNA, cellular mRNA containing IRES elements were found to be preferentially translated under conditions inhibiting cap-dependent initiation, such as endoplasmic reticulum stress, hypoxia, nutrient limitation, mitosis, and cellular differentiation. Also, the little requirement for canonical initiation factors and/or the need for specific ITAFs (often shared between viral and cellular IRES), appear to be quite similar in viruses and eukaryotic cells (Komar and Hatzoglou, 2005; Komar and Hatzoglou, 2011). However, cellular IRES elements may differ from their viral counterparts in several characteristic features in that they appear to be less structured (Komar and Hatzoglou, 2005; Komar and Hatzoglou, 2011). Among the cellular IRES-containing 5'UTR, there are some common features shared by the majority, such as being long and rich in guanine (G) and cytosine (C), which confers great stability to the RNA secondary structure (Baird *et al.*, 2006).

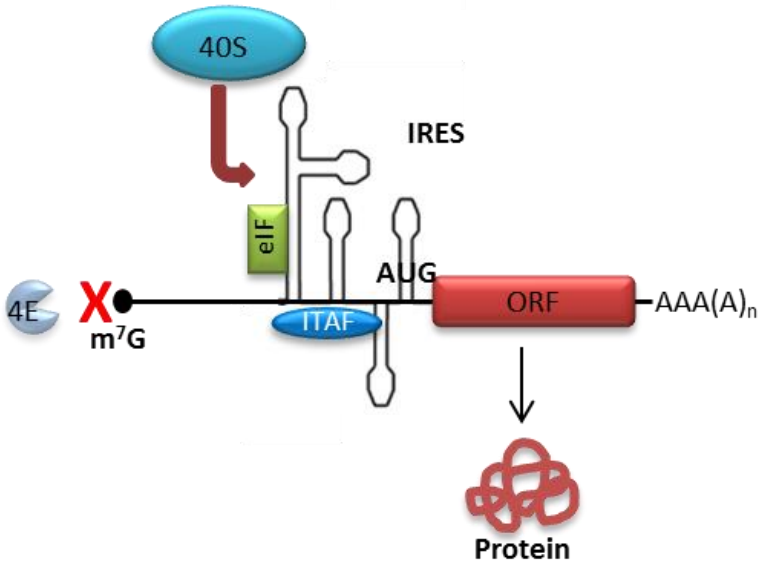


Figure 1.7 — Model of internal ribosome entry site (IRES) -dependent translation initiation. Strong mRNA secondary structures (represented by stem loops) can directly recruit the 40S ribosomal subunit to the initiation codon (AUG) of the open reading frame (ORF) or its vicinity, skipping, or not, the scanning process. This interaction may be direct or partially indirect, requiring the assistance of some canonical initiation factors (eIFs) and/or IRES *trans*-acting factors (ITAFs).

However, when comparing these characteristics on a set of human IRES sequences published at the UTRdb (http://utrdb.ba.itb.cnr.it) and/or RefSeq (http://www.ncbi.nlm.nih.gov/refseq) databases, the conclusion is that none of these criteria is specific enough to be used in further identification of putative IRES sequences. Moreover, a common Y-shaped structure has been predicted for cellular IRES (Le and Maizel Jr., 1997). This pattern has been adapted for the PATSEARCH (Grillo, 2003) to annotate the UTRdb entries as putative IRES motifs and is used by the UTRcan web server (http://itbtools.ba.itb.cnr.it/UTRcan). Unfortunately, this pattern is no more common in known IRES-containing UTR than in all UTR, meaning that it is very difficult to identify *in silico* genes whose transcripts can be translated via IRES, based only on such unspecific characteristics of their 5'UTR. By using a high-throughput bicistronic assay, a recent systematic analysis of sequences mediating IRES-dependent translation in human and viral genomes revealed that the fraction of sequences that mediate IRES-dependent translation is higher in viruses than in the human genome and that, in general, viral IRES are more active than human counterparts, as previously predicted (Jackson, 1991; Weingarten-Gabbay *et al.*, 2016). Furthermore, this recent analysis also revealed that viral 5'UTR with IRES activity have lower GC content and higher minimal free energy in comparison to their human counterparts. On the other hand, the comparison between the GC content and minimal free energy for all active and inactive 5'UTR, from both human and viral origins, revealed that the active 5'UTR have lower GC content and higher minimal free energy (Weingarten-Gabbay *et al.*, 2016). This study also showed that there are two functional classes

of IRES: (i) IRES for which expression is reduced only when a specific position is mutated and (ii) IRES for which mutation in most positions greatly reduces expression (Weingarten-Gabbay *et al.*, 2016). These two classes may represent differences in the underlying mechanism of IRES activity. IRES can either act through a short sequence motif, such as ITAF binding sites — in which only mutations in a specific motif reduce activity (local sensitivity) —, or involve the formation of a secondary structure in which mutations at various positions can disrupt the overall structure and result in reduced activity (global sensitivity) (Weingarten-Gabbay *et al.*, 2016). The mechanism of IRES-mediated translation has been further investigated in detail using the X-linked inhibitor of apoptosis (XIAP) mRNA as a model (Jackson, 1991). XIAP protein is encoded by two mRNA splice variants that differ only in their 5'UTR regions. The more abundant, shorter transcript produces the majority of XIAP protein under normal growth conditions by cap-dependent translation. However, during cellular stress, the longer transcript, containing the IRES element, directs efficient translation despite attenuation of global, cap-dependent translation (Le and Maizel Jr., 1997). During normal proliferative conditions, when the ternary complex is available in abundance, XIAP translation continues in a cap- and eIF2α-dependent mode, similar to other cellular mRNA. Upon serum deprivation, the XIAP IRES-dependent translation switches to an alternative, eIF5B-dependent mode to circumvent attenuation due to eIF2α phosphorylation (Thakor and Holcik, 2012). The cell's ability to evade ternary complex requirement suggests that cells have developed an alternative, eIF2α-independent mechanism of tRNA delivery to support a "rescue" mechanism of translation of critical survival proteins under conditions when the "normal" mechanism is not available (Thakor and Holcik, 2012). Interestingly, a limited investigation of other cellular IRES-containing mRNA (Bcl-xL, cIAP1, Apaf-1, and p97/DAP5) suggests that not all cellular IRES utilise eIF5B-dependent mode of tRNA delivery during serum deprivation (Holcik, 2015). However, the full spectrum of eIF5B-dependent cellular mRNA transcripts still needs to be determined. Still, aiming to better understand how IRES allow direct association of the mRNA with the ribosome without the need for eIF4E, a different study revealed that BCL2 IRES-translation involves the association of DAP5 protein (an eIF4G homolog) with eIF2β and eIF4AI (Liberman *et al.*, 2015). Accordingly, a previous study stated that eIF4A elicits potent activity on the lymphoid enhancer factor-1 (LEF-1) IRES, and, on the contrary, hippuristanol inhibition of eIF4A stalls LEF-1 IRES-mediated translation (Tsai *et al.*, 2014). Recent discoveries revealed that a eukaryotic viral IRES can initiate translation in live bacteria (Colussi *et al.*, 2015). Using crystal structure-solving data, these authors showed that despite differences between bacterial and eukaryotic ribosomes, this IRES binds directly to them and occupies the space normally used by tRNA (Colussi *et al.*, 2015). Initiation in both bacteria and eukaryotes depends on the structure of the IRES RNA; in bacteria, this RNA uses a different mechanism that includes a form of ribosome repositioning after initial recruitment. They propose that the structured IRES RNA forms interactions with bacterial ribosomes that are transient and weaker than the highly-tuned interactions that occur in eukaryotes, but allow internal entry of the ribosome to the message. Recruited subunits or ribosomes are repositioned to a downstream start codon where protein

synthesis starts (Colussi *et al.*, 2015). This primitive mechanism suggests that RNA structure-driven or -assisted initiation may potentially be used in all domains of life, driven by diverse RNA, perhaps possessing tRNA-like character or decoding groove-binding capability, thus bridging billions of years of evolutionary divergence.

The existence of IRES in capped cellular mRNA raised the question of their pathophysiological function and of the advantage of a cap-independent translation (Jackson, 1991). Actually, several reports have demonstrated that cellular IRES function in various physiological processes including spermatogenesis, neuron plasticity, and cell differentiation (Gonzalez-Herrera *et al.*, 2006; Sonenberg and Hinnebusch, 2007; Audigier *et al.*, 2008; Conte *et al.*, 2009). Still, some reports have shown that several IRES are also activated during cell cycle mitosis (Cornelis *et al.*, 2000; Pyronnet *et al.*, 2000) and apoptosis (Holcik and Sonenberg, 2005; Hsu *et al.*, 2016), or are aberrantly activated in tumour cells, and are thus involved in deregulation of gene expression in cancer (Sonenberg and Hinnebusch, 2007; Silvera *et al.*, 2010; Topisirovic and Sonenberg, 2011; Marcel *et al.*, 2013; Leprivier *et al.*, 2015). Furthermore, cellular IRES activity is stimulated during various cellular stresses when cap-dependent translation is blocked (Jackson, 1991; Holcik and Sonenberg, 2005; Bornes *et al.*, 2007; Sonenberg and Hinnebusch, 2007; Piccirillo *et al.*, 2014; Morfoisse *et al.*, 2014; Ozretić *et al.*, 2015). Hence, we may assume that cellular IRES exist to play a crucial role at some critical moments of cell life when cap-dependent translation initiation is compromised, in order for the cell to cope with environmental changes affecting its viability. As IRES-containing transcripts occur throughout every functional class of protein-encoding genes, we decided to cluster them according to the function of the encoded protein, in order to understand which proteins are more prone to be translated via an IRES-dependent mechanism (figure 1.8). Data show that most IRES described so far are found in transcription factor mRNA (21%), in messages coding for growth factors (15%), and in mRNA encoding transporters, receptors and channels (22%). Transcription factors like c-MYC and HIF1α, for instance, are key players in gene expression regulation, since they respond to quick changes in the environment and adapt their transcription levels to the cells' needs in a specific context (Brocato *et al.*, 2014; Kress *et al.*, 2015). As for growth factors (e.g., FGF and VEGF families of proteins), they are of utmost importance to the growth of specific tissues and play a relevant part in promoting cell proliferation and differentiation, and in regulating cell survival (Nakayama, 2009; Brocato *et al.*, 2014; Kress *et al.*, 2015; Masoud and Li, 2015; Rohban and Campaner, 2015). Transporters, receptors and channels (CAT-1, voltage-gated potassium channel, estrogen receptor α, among others) are the main vehicles for cell-cell communication and play a critical role in signal transduction; this makes them key elements in cellular homeostasis in responding to extracellular environment alterations. Thus, perturbations in their function and expression are associated with profound changes in cellular function and significantly contribute to the development and progression of disease (Nakayama, 2009).

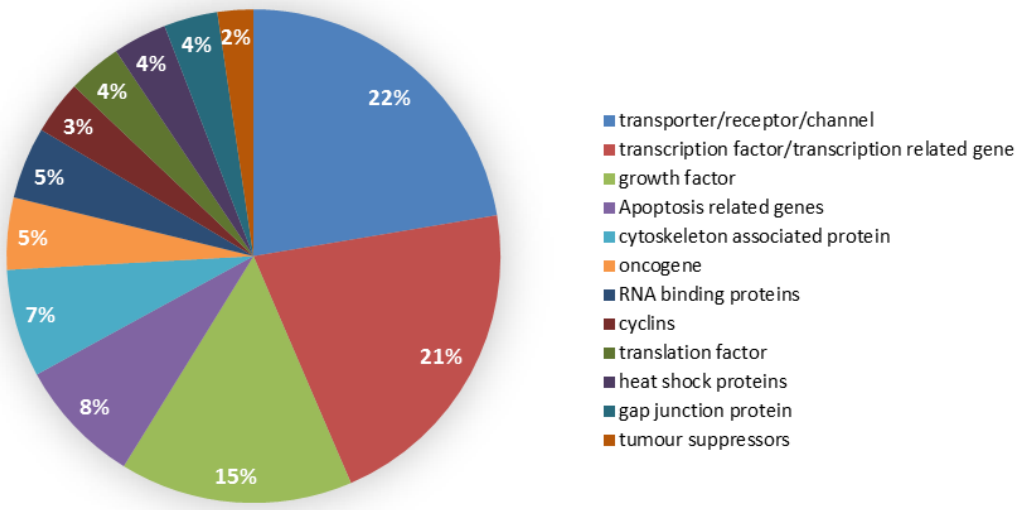


Figure 1.8 — Distribution of IRES-containing transcripts by functional gene families. According to what has been described in the literature, the majority of IRES-containing transcripts encode transcription factors or transcription-related genes, transporters, receptors or channels, and growth factors; nevertheless, several other classes of proteins have been described as synthesized via an IRES-dependent mechanism of translation initiation. The latter include apoptosis-related genes, heat-shock proteins, tumour suppressors, cytoskeleton-associated proteins, gap junction proteins, oncogenes, RNA-binding proteins, cyclins, and translation factors. All these proteins need a fine-tuned regulation of their synthesis, inasmuch as they are somehow involved in crucial cell fitness and survival processes.

Hence, it is logical that transcripts that encode the aforementioned classes of proteins might be translated via IRES as a back-up mechanism when cap-dependent translation is impaired or reduced by some environmental stresses. All other gene families present in this graph (figure 1.8) include proteins with crucial roles in cellular processes that require a fine-tuned regulation and whose expression levels need to be adjusted in response to external cues that interfere with regular mechanisms of translation initiation and concomitant protein synthesis. Furthermore, alterations on their expression levels may account for many of the types of cancer that arise in human population.

1.3.2.1.1. IRES *trans*-acting factors

Although it is still unclear how the actual mechanism of IRES-mediated translation initiation occurs and is regulated, it is already known that most cellular elements are seldom capable of recruiting the 40S ribosomal subunits per se, and may therefore require not only binding of some canonical initiation factors (Spriggs *et al.*, 2009a), but also interaction with other protein factors — the IRES *trans*-acting factors (ITAFs) (Holcik and Sonenberg, 2005). ITAFs are RNA-binding proteins that act to facilitate or block ribosome recruitment to the IRES, thus enhancing or inhibiting translation of these mRNA (Spriggs *et al.*, 2005; Fitzgerald and Semler, 2009). Interestingly, apart from their role in translation regulation, many ITAFs are involved in other

aspects of RNA metabolism such as mRNA splicing, export and stability (Faye and Holcik, 2015). In addition, it has been shown that different cellular IRES reveal different responses to various stress conditions that inhibit cap-dependent translation. For instance, during mitosis, the Upstream of N-ras (Unr), the v-myc avian myelocytomatosis viral oncogene cellular homolog (c-myc), and the cyclin-dependent kinase 11B (PITSLREp58 kinase) IRES become more active, while others do not (Schepens *et al.*, 2007). Furthermore, during apoptosis, the Apaf-1 IRES-dependent translation is active (Holcik and Sonenberg, 2005), whereas XIAP IRES-mediated translation is inhibited (Ungureanu *et al.*, 2006). A striking feature of many ITAFs is that they belong to the group of heterogeneous nuclear ribonucleoproteins (HnRNP A1, C1/C2, I, E1/E2, K and L) known to shuttle between the nucleus and the cytoplasm (Komar and Hatzoglou, 2005; Spriggs *et al.*, 2005; Lewis and Holcik, 2008). In addition, overexpression or depletion of specific ITAFs in normal cells can affect the activity of the cellular IRES that normally uses those ITAFs without altering cap-dependent translation (Lewis and Holcik, 2008), which clearly means that the intracellular concentration of ITAFs plays an important role in modulating the activity of IRES; yet, the exact mechanism(s) underlying ITAF function and that are responsible for regulating ITAF concentration are not fully defined. Here are some hypotheses: i) they remodel IRES special structures to produce conformations with higher or lower affinity for components of the translation apparatus; ii) they build or abolish bridges between the mRNA and the ribosome, in addition to those provided by canonical initiation factors; iii) they take the place of canonical factors in building bridges between the mRNA and the ribosome (Komar and Hatzoglou, 2011). Moreover, two alternative mechanisms have been proposed to explain the effect of ITAF compartmentalisation: nuclear-localised ITAFs associate with their target IRES-containing mRNA and sequester them in the nucleus away from the translational machinery (Semler and Waterman, 2008); ITAFs in the nucleus are primarily in an mRNA-unbound form, separated from their target IRES-containing messages residing in the cytoplasm. Upon appropriate signals, caused by stress or other physiological conditions, either the ITAF-bound mRNA or the unbound ITAFs themselves translocate from the nucleus to the cytoplasm, allowing translation of the mRNA to proceed (Lewis and Holcik, 2008).

Many proteins have been identified as ITAFs that can play decisive roles in regulating IRES-mediated translation, especially in processes such as cancerigenesis or other disease-related processes. These ITAFs include polypyrimidine tract binding protein (PTB), hnRNPC1/C2, human antigen R (HuR), Unr, Poly(rC)binding protein 1 (PCBP1), La autoantigen (La), or death associated protein 5 (DAP5), etc. For instance, PTB interacts with and controls the expression of Unr (Schepens *et al.*, 2007), tumour protein 53 (p53) (Grover *et al.*, 2008), human insulin receptor (hiR) (Spriggs *et al.*, 2009b), or c-myc (Cobbold *et al.*, 2008) IRES. Another ITAF, hnRNPC1/C2 has been shown to interact with Unr, XIAP (Schepens *et al.*, 2007), and p53 IRES (Grover *et al.*, 2008). Concerning Bag-I IRES, PCBP1 and PTB proteins bind to IRES RNA and unwind a specific region via RNA chaperone activity — changes that eventually facilitate the recruitment of the ribosome (Pickering *et al.*, 2004).

IRES-mediated translation of cIAP1 transcript, which contains a stress-inducible IRES

governing cIAP1 protein expression, is also aided by several proteins that bind to the cIAP1 IRES, such as the insulin-like growth factor 2 mRNA binding protein 1 (IGF2BP1), a critical translational regulator of cIAP1-mediated apoptotic resistance (Faye *et al.*, 2014), and the nuclear factor 45 (NF45), which is required for IRES-mediated induction of cIAP1 protein during the unfolded protein response (Graber *et al.*, 2010; Faye *et al.*, 2013). Further characterisation of NF45 as an ITAF uncovered that it preferentially binds to AU-rich 5'UTR regions (Graber *et al.*, 2010). It is predictable that an IRES containing more than 60% of AU will be NF45-dependent (Faye *et al.*, 2013). Additionally, these authors provide evidence that cells deficient in NF45 ITAF activity exhibit reduced IRES-mediated translation of XIAP and cellular inhibitor of cIAP1 mRNA. NF45 is usually found to be in complex with other members of the nuclear factor associated with double-stranded RNA family, in particular NF90 and its isoforms, which mutually safeguard their protein stability (Guan *et al.*, 2008). Both NF45 and NF90 can bind double-stranded, as well as structured single-stranded RNA (Faye and Holcik, 2015). NF90 was also shown to bind to the hypoxia stability region of the VEGF 3'UTR and to modulate its mRNA stability and translation during conditions of hypoxia, which means NF90 knock-down — and consequent NF45 depletion — limits the induction of VEGF mRNA and protein expression during hypoxia, resulting in growth reduction and angiogenic potential in a xenograft tumour model (Vumbaca *et al.*, 2008).

Another example of the importance of ITAFs on regulating IRES-mediated translation is p53. Translation regulation of this mRNA is controlled by *cis*-acting elements and *trans*-acting factors. Several cellular proteins have been identified as ITAFs for p53 mRNA translation. These — such as PTB (Grover *et al.*, 2008), hnRNPC1/C2 (Grover *et al.*, 2011), MDM2 (Yin *et al.*, 2002), and RPL26 (Takagi *et al.*, 2005) — bind to p53 mRNA and positively regulate translation, whereas RNPC1 (Zhang *et al.*, 2011) and nucleolin (Takagi *et al.*, 2005) negatively regulate p53 translation, inducing G1 cell cycle arrest. As far as PTB is concerned, it was reported that, after doxorubicin treatment, this ITAF relocalises from nucleus to cytoplasm; consequently, there is an increase in p53 IRES activity (Grover *et al.*, 2008; Grover *et al.*, 2009). Several reports have added more proteins to the list of ITAFs regulating p53 IRES-mediated translation. Sharathchandra and colleagues (2012), using RNA affinity approach, have identified Annexin A2 and PTB-associated Splicing Factor (PSF/SFPQ) as novel ITAFs for two p53 isoforms — full-length p53, FL-p53, and a truncated isoform, ΔN-p53, that modulates the functions of FL-p53 and also has independent functions (Ray *et al.*, 2006; Candeias *et al.*, 2006). They have shown that the purified Annexin A2 and PSF proteins specifically bind to p53 IRES elements. In the presence of calcium ions, Annexin A2 showed increased binding to p53 IRES, and immunopulldown experiments suggest that both Annexin A2 and PSF associate with p53 mRNA *ex vivo*, as well. Furthermore, partial knock-down of these two proteins showed a decrease in p53 IRES activity and reduced levels of both the p53 isoforms. Additionally, the interplay among Annexin A2, PSF and PTB proteins for binding to p53 mRNA appears to play a crucial role in IRES function, suggesting the importance of the two new *trans*-acting factors in regulating p53-IRES function, which, in turn, influences the synthesis of p53 isoforms. Similarly,

Malbert-Colas *et al.* (2014) have demonstrated that following phosphorylation by the ataxia telangiectasia mutated (ATM) kinase at serine 403, the C-terminal RING domain of HDMX binds the nascent p53 mRNA to promote a conformation that supports the p53 mRNA-HDM2 interaction and the induction of p53 synthesis. HDMX and its homolog HDM2 bind the same p53 IRES sequence structure but with different specificity and function. These results show how HDMX and HDM2 act as non-redundant ITAFs to bring a positive synergistic effect on p53 expression during genotoxic stress by first altering the structure of the newly synthesized p53 mRNA, followed by stimulation of translation. Finally, a 2015 study reveals two novel p53 ITAFs, translational control protein 80 (TCP80) and RNA helicase A (RHA), which positively regulate p53 IRES activity (Halaby *et al.*, 2015a). According to the authors, overexpression of TCP80 and RHA leads to increased expression and synthesis of p53. Furthermore, they discovered two breast cancer cell lines that retain wild-type p53 but exhibit defective p53 induction and synthesis following DNA damage. This occurs due to the extremely low levels of TCP80 and RHA in both cell lines, meaning expression of both proteins is required to significantly increase p53 IRES activity in these cells. Moreover, they found that cancer cells transfected with a short hairpin RNA (shRNA) against TCP80 not only exhibit decreased expression of TCP80 and RHA but also display defective p53 induction and diminished ability to induce senescence following DNA damage. Thus, these data reveal a novel mechanism of p53 inactivation that links deregulation of IRES-mediated p53 translation to tumourigenesis.

Altogether, the examples provided above give us a clear view on the importance of identifying proteins serving IRES-mediated translation as ITAFs and find innovative therapeutic approaches able to target them.

1.3.2.2. Cap-independent translation enhancer-mediated translation

Some cellular mRNA that have been considered to contain IRES fail to pass stringent control tests for internal initiation, thus raising the question of how they are translated under stress conditions. Terenin and co-workers (2013) showed that the insertion of an eIF4G-binding element from a viral IRES into 5'UTR of strongly cap-dependent mRNA dramatically reduces their cap requirement in mammalian cells. This mechanism has been proven to be different from the internal entrance because these mRNA fail the bicistronic test, meaning they need a free 5' end for the preinitiation complex to bind. Thus, although this is a cap-independent mechanism, it is 5' end-dependent and involves special elements, the so-called *cap-independent translation enhancers* — CITE (figure 1.9) (Shatsky *et al.*, 2010). CITE are located within the untranslated regions of the mRNA that attract key initiation factors that promote the assembly of translation initiation complexes (Shatsky *et al.*, 2010; Andreev *et al.*, 2013).

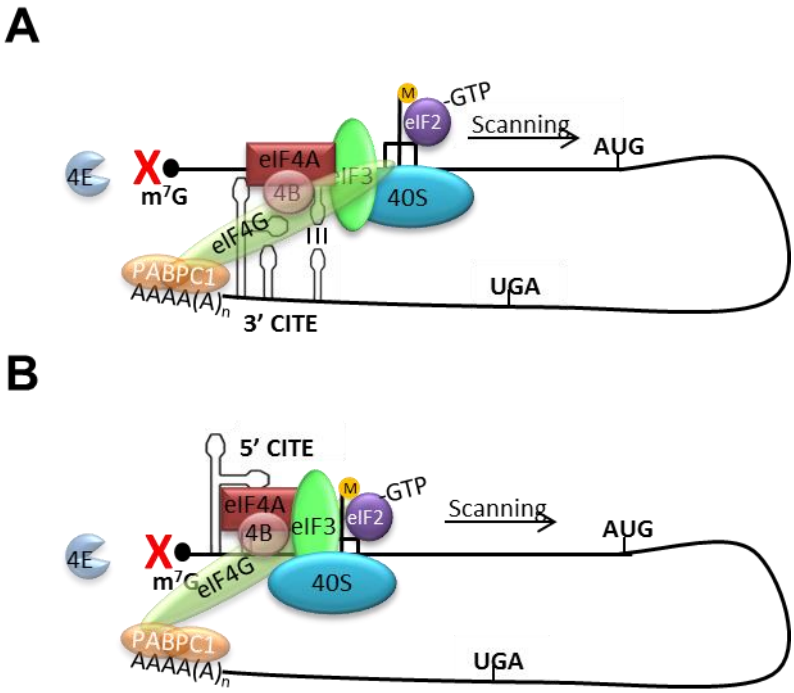


Figure 1.9 — Model of cap-independent translation enhancer (CITE) -dependent translation initiation. CITE are *cis*-acting elements located within the untranslated regions (UTR) of a mRNA that attract key initiation factors (eIFs), such as eIF4G, in order to promote the assembly of translation initiation complexes. Then, the initiation codon is found by ribosomal scanning. (A) The 3'CITE (CITE is located within the 3'UTR, represented by stem loops) is thought to recruit components of translational apparatus to deliver them to the 5' end of mRNA through long-distance base pairing between 5' and 3'UTR. (B) The 5'CITE (CITE is located within the 5'UTR) is capable of presumably establishing rather weak interactions with initiation factors of the scanning machinery to initiate translation.

The majority of CITE have been described within the 3'UTR (3' CITE) of plant viral mRNA and are believed to recruit the 80S ribosome to bring it into close proximity with the initiation codon through long-distance base pairing between 3' and 5'UTR (figure 1.9.A) (Fabian and White, 2004; Karetnikov and Lehto, 2008; Simon and Miller, 2013; Simon, 2015; Blanco-Pérez *et al.*, 2016). In the case of 5' CITE, a CITE is located within the 5'UTR and is presumably capable of additional, rather weak interactions with initiation factors of the scanning machinery (figure 1.9.B) (Shatsky *et al.*, 2010; Andreev *et al.*, 2013). Although cap recognition by eIF4E still plays a major role in the mRNA recruitment, the primary binding of the mRNA is still possible in the absence of this interaction, solely due to some interaction between key initiation factors (or the 40S ribosomal subunit itself) with 5' CITE. Some components of the translation apparatus — for example, eIF4G and eIF3 — are able to be directly or indirectly recruited into the 5'UTR via RNA-protein interactions with concomitant recruitment of other components of the scanning apparatus (Andreev *et al.*, 2013; Roberts *et al.*, 2015). In this way, the 5'UTR of an mRNA creates, in its vicinity, a high concentration of translational components. This also helps overcome the competition for factors from other cellular mRNA. Human *Apaf-1* mRNA initiates

translation via this mechanism, under conditions that suppress the cap-binding factor eIF4E (Andreev *et al.*, 2013). Indeed, it has been shown that *Apaf-1* 5'UTR can mediate an m^7G cap-independent — but dependent on a free 5' end — translation initiation, even under apoptosis (Andreev *et al.*, 2013). As a consequence, this leads to the relatively preferential translation of *Apaf-1* mRNA under stress conditions. *Apaf-1* plays a central role in DNA damage-induced apoptosis and its depletion therefore contributes to malignant transformation. Inactivation of the *Apaf-1* gene is implicated in disease progression and chemoresistance of some malignancies, such as metastatic melanomas (Soengas *et al.*, 2001). In this regard, CITE-mediated translation under apoptosis may contribute extensively to the maintenance of Apaf-1 protein levels, and to its tumour suppressor activity under stress conditions. It has been previously shown that *Apaf-1* 5'UTR also has IRES-activity that is triggered by UV-induced apoptosis (Ungureanu *et al.*, 2006). How this whole set of data can be articulated remains to be assessed.

1.3.2.3. m^6A-mediated translation

A feature of many eukaryotic mRNA is N^6-methyladenosine (m^6A), a reversible base modification seen in the 3'UTR, the coding region, and the 5'UTR (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). Although the biological function of the m^6A in 3'UTR had already been explored (Meyer *et al.*, 2012; Wang *et al.*, 2013; Wang *et al.*, 2014; Wang *et al.*, 2015), the role of m^6A in the 5'UTR has only recently been unveiled (Meyer *et al.*, 2015). Indeed, it has been recently shown that the m^6A in the 5'UTR works as an alternative to the 5' cap to stimulate translation initiation; m^6A residues within the 5'UTR act as m^6A-induced ribosome engagement sites (MIRES, figure 1.10) (Meyer *et al.*, 2015). In addition, data have shown that the m^6A in the 5'UTR can bind eIF3, which is sufficient to recruit the 40S ribosomal subunit to initiate translation in the absence of the cap-binding factor eIF4E (Meyer *et al.*, 2015). Still, it appears that the m^6A-mediated translation initiation involves a 5' end-dependent 5'UTR scanning mechanism (Meyer *et al.*, 2015), contrary to internal ribosomal entry (Jackson, 2013). As m^6A-mediated cap-independent translation initiation requires 5'UTR scanning, it seems to perform similarly to what has been previously described for mRNA containing an eIF4G-binding viral IRES-domain in its 5'UTR (Andreev *et al.*, 2012; Terenin *et al.*, 2013; Meyer *et al.*, 2015). How m^6A is recognised by the translation machinery and facilitates cap-independent initiation still needs further study. However, the significance of 5'UTR m^6A residues has been observed in both ribosome profiling datasets and individual cellular mRNA analyses, such as the heat-shock protein 70 (HSP70) (Meyer *et al.*, 2015; Zhou *et al.*, 2015). Data from *HSP70* mRNA experiments revealed that a single m^6A modification site in the 5'UTR enables translation initiation independent of the 5' end N^7-methylguanosine cap, granting a mechanism for selective mRNA translation under heat-shock stress (Meyer *et al.*, 2015; Zhou *et al.*, 2015). Since it has been previously shown that *HSP70* 5'UTR also possesses IRES activity (Rubtsova *et al.*, 2003; Hernández *et al.*, 2004; Sun *et al.*, 2011b), it remains to be determined whether both mechanisms cooperate to increase cap-independent translation in response to heat shock.

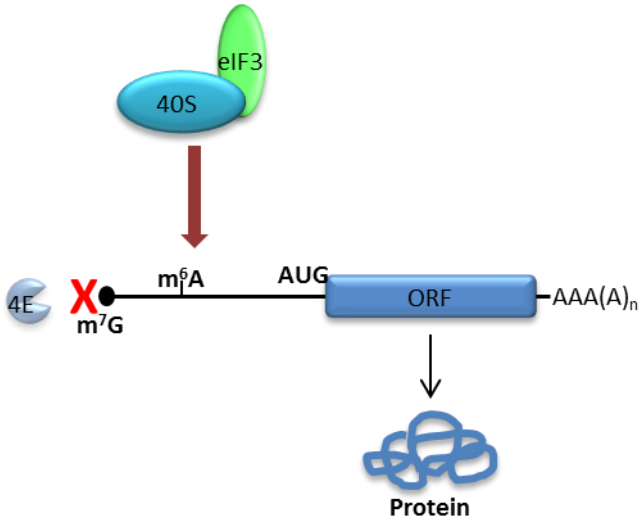


Figure 1.10 — Model of m^{6}A-dependent translation initiation. In m^{6}A-mediated translation, m^{6}A residues within the mRNA 5' untranslated region (5'UTR) act as an m^{6}A-induced ribosome engagement site to recruit initiation factor 3 (eIF3) and the 40S ribosomal subunit. Then, the formed preinitiation complex is able to scan the 5'UTR until it recognises the initiation codon.

Furthermore, it will be important to analyse whether other stress response pathways also induce m^{6}A modification in the 5'UTR, in order to use them to mediate cap-independent translation initiation, in response to stress. Also, it will be of great importance to know whether m^{6}A-mediated translation is involved in triggering disease states, such as carcinogenesis, and/or in responding to chemotherapeutics. Considering that putative cellular IRES often lack the complex structural elements seen in viral IRES (Hellen and Sarnow, 2001; Jackson 2013), and that there are flows inherent to many assays that test cellular IRES function, the utility of m^{6}A in the 5'UTR might be an additional or alternative mechanism that explains the occurrence of cap-independent translation. Noteworthy, 5'UTR methylation in the form of m^{6}A is dynamic, and stress-inducible (Meyer *et al.*, 2015; Zhou *et al.*, 2015; Zhou *et al.*, 2016). Thus, stress-inducible 5'UTR methylation, alongside cap-independent translation initiation stimulation, constitutes a new pattern of translational control.

Recent data have revealed that RNA cytosine hydroxymethylation can favour translation in *Drosophila* cells (Delatte *et al.*, 2016). It remains to be determined whether this RNA modification also occurs in mammalian cells and whether it mediates cap-independent translation.

1.4. Cooperation between IRES and other *cis*-acting RNA regulons

There are several structural motifs within the 5'UTR of a transcript that may influence the translation of the corresponding protein. All translation initiation mechanisms, either cap-dependent or -independent are influenced by the presence of those structural elements. Here,

we present some of the interaction that may contribute to regulate translation initiation of a transcript either in normal conditions or under stress stimuli.

1.4.1. IRES and G-quadruplex structures

A G-quadruplex is a 3-D structure that arises when several G-quartets can form proximally within a single strand of nucleic acids and stack upon each other by means of π-π interactions. A G-quartet is formed by four G bases arranged in a square planar cyclic hydrogen-bonding pattern, where each guanine is both the donor and acceptor of two hydrogen bonds, providing a central site where the oxygen lone pair of the carbonyl groups can coordinate with metal cations. G-quartets can arise intermolecularly between G-rich strands or intramolecularly within some G-rich nucleic acid sequences (Bugaut and Balasubramanian, 2012). Several examples in the literature have described that 5'UTR RNA putative quadruplex sequences (PQS) inhibit translation, leading to the proposal that 5'UTR RNA G-quadruplexes are "predictable" inhibitory elements of gene expression. There are several examples in the literature that show a decrease in cap-dependent translation initiation when a G-quadruplex is formed. This repression ranges from 35% in proteins like neural cell adhesion molecule 2 and thyroid hormone receptor α (Beaudoin and Perreault, 2010), up to 85% for estrogen receptor α (Balkwill, G.D. *et al.*, 2009). However, there are cases when RNA G-quadruplex formation has been shown to actually promote translation.

The human *FGF-2* mRNA has a G-quadruplex motif within its 5'UTR that is a structural determinant of IRES activity. The 176-nucleotide long *FGF-2* IRES is highly structured and contains two RNA stem-loops and a G-quartet motif, and each contributes to IRES activity (Bonnal, 2003). Another example in which the presence of a G-quadruplex promotes translation is the human *VEGF* mRNA. The 5'UTR of *VEGF* transcript is 1038 nucleotides long, GC-rich, and able to initiate translation via IRES. This untranslated region harbours two separate IRES. A 293-nucleotide portion, IRES-A, initiates translation at the canonical AUG and is known to maintain VEGF translation under hypoxia. This region also includes a sequence containing more than four guanines in a stretch (nts 774–790), which provides enough redundancy to ensure the formation of RNA G-quadruplex structures. These are critical to the IRES-dependent translation initiation. When the sequence is mutated in a way that disrupts the formation of the G-quadruplex, IRES activity is eliminated. This suggests a G-quadruplex structure must be formed in order to maintain the IRES function and hence promote translation (Morris *et al.*, 2010). A recent study, however, states that the G-quadruplex within the *VEGF* IRES is dispensable for cap-independent function and activation under stress conditions. Yet, stabilisation of the *VEGF* G-quadruplex by increasing the G-stretches length, or by replacing it with the one of NRAS[9], results in strong inhibition of IRES-mediated translation of VEGF (Cammas *et al.*, 2015). The authors have also shown that G-quadruplex ligands stabilise the *VEGF G*-quadruplex and inhibit cap-independent translation *in vitro*. Importantly, the amount of

[9]The *NRAS* G-quadruplex efficiently blocks mRNA translation when it is positioned close to the 5' end, within the first 50 nucleotides of the NRAS 5'UTR (Kumari *et al.*, 2007), but it loses its inhibitory activity when relocated farther away (Kumari *et al.*, 2008).

human *VEGF* mRNA associated with polysomes decreases in the presence of a highly selective stabilising G-quadruplex ligand, leading to reduced VEGF protein expression. These findings show that intrinsically stable or ligand-stabilised G-quadruplexes function as inhibitors of IRES-mediated translation and, therefore, uncover the existence of functionally silent G-quadruplex structures that are susceptible to conversion into efficient repressors of cap-independent mRNA translation. In view of the dynamic nature of IRES, and of the regulation of RNA interactions within them being a mechanism for regulating their activity, these data are consistent with a mechanism whereby stable G-quadruplex structures prevent the conformational changes necessary to recruit the ribosome.

Translation of angiogenic and growth factors like VEGF and FGF family members is crucial in cancer onset and development. The synthesis of these proteins allows the tumorigenic cells to grow and proliferate, since it creates the physiological conditions for their nourishing. Considering how critical VEGF expression in tumour angiogenesis is, the G-quadruplex at *VEGF* IRES-A may represent a potential therapeutic target to downregulate VEGF expression in tumours. As a result, G-quadruplex ligand-mediated down-regulation of transcription of *VEGF* (Sun *et al.*, 2008), *HIF* (Welsh *et al.*, 2013) and the *VEGFR-2* receptor (Salvati *et al.*, 2014) certainly corroborates the application of ligands in a cellular context to target G-quadruplexes acting on the VEGF axis and mediating tumour angiogenesis (Cammas *et al.*, 2015).

1.4.2. IRES and upstream open reading frames

Upstream open reading frames (uORFs) are another kind of *cis*-acting elements existing within the 5'UTR of transcripts, able to regulate translation. uORFs can modulate cap-dependent translation by repressing the main ORF's translation. In addition, some reports have shown that the presence of a uORF can regulate the IRES-dependent translation.

There are several pieces of evidence showing that many uORF- and IRES-containing genes are involved in cell growth and differentiation, such as platelet-derived growth factor (Gerlitz *et al.*, 2002), GATA-6 (Takeda, 2004), Cat-1 (Yaman *et al.*, 2003; Fernandez *et al.*, 2005), VEGF-A (Bastide *et al.*, 2008) and FGF9 (Chen *et al.*, 2014), so, it is logical to think that the interaction between IRES and uORFs co-existing within the same 5'UTR leads to alterations in the regular expression pattern of proteins.

It has been shown by Yaman and colleagues (2003) that the 5'UTR of *CAT-1* transcript has a uORF that modulates the activity of the IRES. These results suggest a model for regulation of the *CAT-1* IRES, which is dependent on translation of the uORF. In the absence of uORF translation, the mRNA leader exists in a structure that locks the IRES in a dormant state (figure 1.11). However, translation of the uORF disrupts this structure, allowing the leader to form the IRES that can be induced by amino acid starvation, during which an ITAF binds the inducible IRES, leading to increased translation initiation at the CAT-1 ORF. This model suggests that translation of the uORF plays different roles in fed and starved cells. In fed cells, uORF translation inhibits downstream translation initiation by preventing the ribosome from reaching

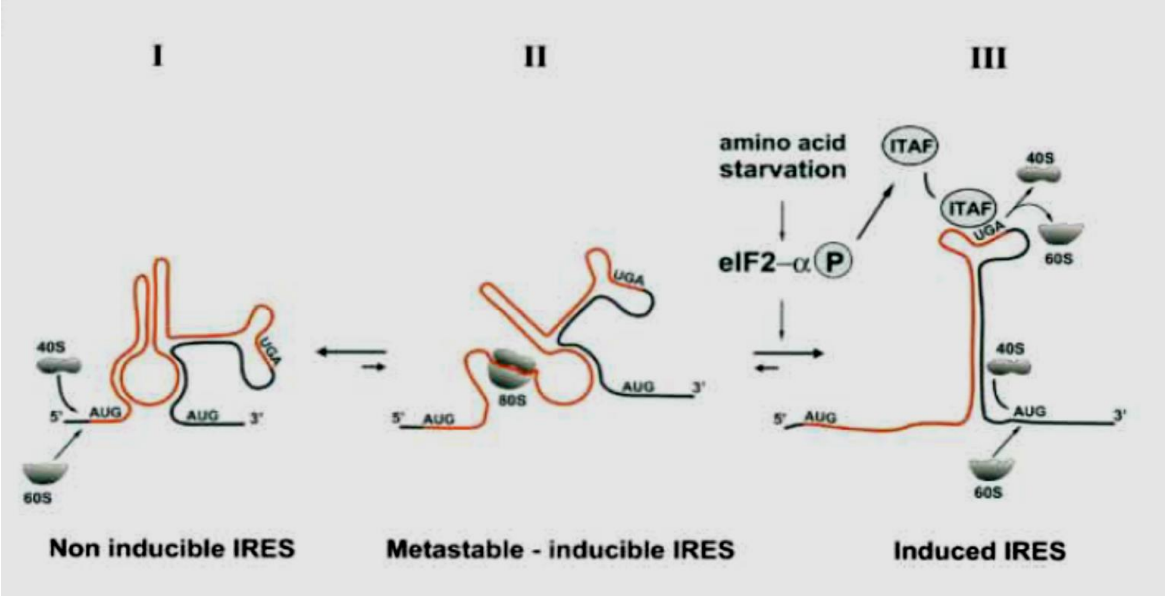


Figure 1.11 — The zipper model of translational control. According to this model, both uORF and IRES element co-exist within the 5'UTR of the transcript. Cap-dependent translation of the uORF induces a conformational change in the secondary structure and exposes the IRES element. The latter is therefore capable of mediating a cap-independent translation initiation mechanism. This kind of transition occurs mainly under stress conditions, such as nutrient deprivation or hypoxia. Adapted from Yaman *et al.* (2003).

the CAT-1 ORF. In starved cells, uORF translation unfolds the leader, allowing the ITAF that is synthesized in response to eIF2α phosphorylation to bind the IRES and initiate CAT-1 protein synthesis (Fernandez *et al.*, 2005). This model of *CAT-1* IRES proposes that the uORF plays the role of a zipper that opens and closes the IRES (Fernandez *et al.*, 2005). Likewise, there may be uORFs that are translated via an IRES-dependent mechanism. The expression of GATA-6 and different VEGF-A isoforms is regulated by a small uORF located within an IRES, and a cap-independent mechanism (Takeda, 2004; Bastide *et al.*, 2008). On the other hand, the uORF may be located upstream the IRES, as is the case of FGF9. Under normal conditions, the uORF is generally translated in order to repress the expression of the main ORF and keep a low level of protein synthesis. Under specific environmental conditions, such as hypoxia, the high levels of FGF9 expression are achieved by activating the FGF-IRES, and ribosomes are switched from the AUG of the uORF to the AUG of the main ORF. Thus, these two elements play opposite roles in FGF9 translational control in order to fine-tune its protein expression levels, either in normoxia or under hypoxia (Chen *et al.*, 2014).

A recent report by Ozretić *et al.* (2015) on the regulation of human *PTCH1b* expression revealed that the transcript — encoding a 12-pass transmembrane receptor with a negative regulatory role in the Hedgehog-Gli signalling pathway[10] — contains several *cis*-elements within its 5'UTR that account for the regulation of protein expression levels. These authors have

[10]The Hedgehog-Gli (Hh-Gli) pathway is a highly conserved cellular mechanism for transducing signals from the cell surface into the nucleus, stimulating expression of many genes, which results in an appropriate physiological response to changes in the cellular environment (Briscoe and Thérond, 2013).

shown that upstream AUG codons (uAUG) present only in longer 5'UTR could negatively regulate the amount of PTC1 isoform L (PTC1-L), whereas the existence of an IRES would counteract the effect of those uAUG and enable synthesis of PTC1-L under stress conditions, such as during hypoxia. These results highlight an exceptionally complex (and so far unexplored) role of 5'UTR *PTCH1b cis*-element features in the regulation of the Hh-Gli-signalling pathway (Ozretić *et al.*, 2015).

Such interplay between uORF and IRES allows a deeper control of protein synthesis and a quicker response to adverse conditions that impair the cap-dependent canonical mechanism of translation initiation.

1.5. Translational control in health and disease

Deregulation of gene expression, namely at translation initiation, can lead to the onset of several diseases, such as cancer, neurodegenerative or inflammatory conditions. Many proteins with a role in such diseases can be translated via a cap-independent mechanism, the most frequent being the IRES-mediated translation initiation. For instance, Sammons *et al.* (2010) identified ZNF9 (zinc finger protein 9) as a regulator of cap-independent translation, which indicates that its activity may contribute mechanistically to the myotonic dystrophy type 2 (DM2)[11] phenotype. They showed that ZNF9 is associated with actively translating ribosomes and hence functions as an activator of cap-independent translation of the human ornithine decarboxylase (ODC) mRNA. This activity is mediated by direct binding of ZNF9 to the IRES sequence within the 5'UTR of *ODC* mRNA. ZNF9 can activate IRES-mediated translation of ODC within primary human myoblasts; such activity is, however, reduced in myoblasts derived from a DM2 patient. On the other hand, Rubsamen *et al.* (2012) presented evidence of a new mechanism of EGR2[12] (early growth response 2) regulation via enhanced IRES-dependent translation under pro-inflammatory conditions. Using bicistronic reporter assays, these authors found that *EGR2* contains an IRES within its 5'UTR, which facilitates enhanced translation after treatment with a conditioned medium of activated monocyte-derived macrophages, and concluded that *EGR2*-IRES activity was induced by IL-1β and p38-MAPK signalling. Together, these data prove that EGR2 expression is translationally regulated via an IRES element, which is responsive to an inflammatory environment. Since EGR2 plays a crucial role in T-cell tolerance, this knowledge on EGR2 regulation will be of great interest for conditions where T-cell activation should be therapeutically altered, such as transplantations or tumour immunotherapies.

[11]Myotonic dystrophy types 1 and 2 (DM1 and DM2) are forms of muscular dystrophy that share similar clinical and molecular manifestations, such as myotonia, muscle weakness, cardiac anomalies, cataracts, and the presence of defined RNA-containing foci in muscle nuclei. DM2 is caused by an expansion of the tetranucleotide CCTG repeat within the first intron of ZNF9 (Sammons *et al.*, 2010).

[12]Defects in this gene are associated with Charcot-Marie-Tooth disease type 1D, Charcot-Marie-Tooth disease type 4E, and with Dejerine-Sottas syndrome (Bird, 1998).

1.5.1. Cap-independent translation and cancer

Cancer is a disease caused by oncogene activation and tumour suppressor gene inhibition. Deep-sequencing studies identified numerous tumour-specific mutations, not only in protein-coding, but also in non-coding sequences. The coding-independent mutations in regulatory elements, UTR, splice sites and non-coding RNA, and synonymous mutations, are able to affect gene expression from transcription to translation (reviewed in Diederichs *et al.*, 2016). In addition, the process of tumourigenesis involves back-up mechanisms that allow tumour cells to cope with stress, including those involved in the synthesis of proteins required for stress adaptation (Holcik and Sonenberg, 2005; Blais *et al.*, 2006; Gaccioli *et al.*, 2006; Braunstein *et al.*, 2007; Lewis *et al.*, 2007; Topisirovic and Sonenberg, 2011; Leprivier *et al.*, 2015). Many transcripts relevant to cancer — but with no specific tumour-associated mutations — are able to initiate translation through a cap-independent mechanism, namely through an IRES element. Accordingly, several oncogenes, growth factors and proteins involved in the regulation of programmed cell death are translated via IRES elements in their 5'UTR. Selective translation of these factors may contribute to the survival of cancer cells under stress situations induced within the tumour's microenvironment (such as nutrient deprivation, hypoxia, or therapy-induced DNA damage) and the establishment of cancer cells that resist conventional therapies.

It is known that 4E-BP activation in response to hypoxia and mTORC1 inhibition dictates a switch from cap-dependent to cap-independent translation to support tumour growth and angiogenesis (Blais *et al.*, 2006; Braunstein *et al.*, 2007). Indeed, Braunstein *et al.* (2007) demonstrated that the majority of large, advanced breast cancers overexpress the translation regulatory protein 4E-BP1 and the initiation factor eIF4G. Overexpression of these two proteins leads to cap-independent mRNA translation that promotes increased tumour angiogenesis and growth. This switch results in selective translation of IRES-containing mRNA. These include a number of mRNA that encode proteins involved in signal transduction pathways, gene expression and development, differentiation, apoptosis, angiogenesis, cell cycle, or stress response (Holcik and Sonenberg, 2005; Sonenberg and Hinnebusch, 2007; Topisirovic and Sonenberg, 2011), as is the case of *VEGF-A* (Stein *et al.*, 1998), *HIF1α* (Lang *et al.*, 2002) and *FGF2* (Conte *et al.*, 2008), among others. For example, hypoxia reduces vascular endothelial growth factor C (VEGF-C) cap-dependent translation via the up-regulation of hypophosphorylated 4E-BP, but induces its IRES-mediated translation initiation in an HIF1α signalling-independent way (Morfoisse *et al.*, 2014). Notably, the VEGF-C IRES activity is higher in metastasizing tumour cells in lymph nodes than in primary tumours, most likely because lymph vessels in these lymph nodes are severely hypoxic (Morfoisse *et al.*, 2014). Still, some studies assessing IRES activities of *HIF1α* and *VEGF* showed only very low translation activity from these elements, suggesting that cryptic promoter activity in constructs used for those studies may interfere therein (Bert *et al.*, 2006; Jackson, 2013). Of note, Young *et al.* (2008) confirmed that *VEGF* transcripts are selectively translated under hypoxia, even without significant IRES-mediated translation, suggesting that selective and alternative IRES-independent translation mechanisms might sustain VEGF synthesis under these conditions.

Silvera and Schneider (2009) have shown that inflammatory breast cancer cells have adapted so as to mimic a state of prolonged hypoxia during translation. This likely optimises the production of proteins required for tumour emboli survival and dissemination, a state promoted by high levels of eIF4GI protein coupled with a constitutively active 4E-BP1, leading to higher rates of translation of IRES-containing mRNA, namely VEGF and p120 catenin, which are responsible for maintaining high rates of angiogenesis and membrane associated E-cadherin, respectively. Fibroblast growth factors (FGFs) play a major role in the processes of proliferation and differentiation of wide variety of cells and tissues; thus, their translation has to be tightly regulated so that the expression levels are maintained within a range that promotes healthy growth and development. Some FGFs, such as *FGF1* and *FGF2*, contain IRES elements within their 5'UTR, which enable cap-independent translation initiation (Vagner *et al.*, 1995; Martineau *et al.*, 2004). These factors have been shown to be expressed at increased levels in prostate cancer (Kwabi-Addo *et al.*, 2004). Besides, the role of IRES-mediated regulation of FGF2 translation in tumourigenesis is considered a critical step not only in solid tumours but also in multiple myeloma, in a way that the *FGF2* IRES is the non-cytotoxic primary molecular target of thalidomide and should be considered the target for the development of immunomodulatory drugs in multiple myeloma (Lien *et al.*, 2014). FGF9 is another FGF family member, whose aberrant expression usually results in human malignancies (Huang *et al.*, 2015). Overexpression of FGF9 has transforming potential in fibroblasts and stimulates the invasion of epithelial and endothelial cells, suggesting it might result in uncontrolled cell proliferation and malignancy. Under normoxia, FGF9 protein levels are kept low due to the presence of a uORF that represses its expression. In response to hypoxia, a switch to IRES-dependent translational control up-regulates FGF9 protein expression, and becomes the likely mechanism underlying its expression in cancer cells, namely colon cancer cells (Chen *et al.*, 2014). Another case of a protein whose expression is up-regulated during tumourigenesis by activation of IRES-mediated translation is specificity protein-1 (Sp1). It is accumulated during hypoxia in an IRES-dependent manner and is strongly induced at protein, but not mRNA, level in lung tumour tissue, suggesting that translational regulation might contribute to the accumulation of Sp1 during tumourigenesis (Yeh *et al.*, 2011). Further studies have revealed that IRES-mediated translation of Sp1 occurs through the recruitment of nucleolin to the 5'UTR of *Sp1* mRNA (Hung *et al.*, 2014). CDKN2A/p16INK4a is an essential tumour suppressor gene that controls cell cycle progression and replicative senescence, and is the main melanoma susceptibility gene. Its mRNA is also subject to IRES-mediated translation. In fact, p16INK4a 5'UTR acts as a cellular IRES and Y-box binding protein 1 (YBX1) acts as its ITAF under hypoxic stress, both in cancer-derived cell lines and p16INK4a wild-type lymphoblastoid cells obtained from a melanoma patient (Bisio *et al.*, 2015). Interestingly, a germline sequence variant found in the p16INK4a 5'UTR (c.-42T>A) of a multiple primary melanoma patient results in local flexibility changes in RNA structure, impairing the binding of YBX1 and its stimulatory effect on IRES-dependent translation efficiency. This sequence variant appears to alter p16 protein expression levels. Impaired p16 translation under hypoxia could provide a mechanistic clue to explain

melanomagenesis associated with this germline variant (Bisio *et al.*, 2015). In a different study, data showed that in multiple myeloma cells under endoplasmic reticulum (ER) stress induced by thapsigargin, tunicamycin or the myeloma therapeutic bortezomib, the *c-Myc* IRES is also activated and requires proteins hnRNP A1 and RPS25 as ITAFs for c-Myc protein levels to be maintained (Shi *et al.*, 2016).

Translation of specific transcripts in response to nutrient deprivation also occurs through cap-independent mechanisms. Specifically, synthesis of two amino acid transporters, namely CAT-1 and sodium-coupled neutral amino acid transporter 2 (SNAT2), which are required to promote recovery of amino acid balance, are controlled by IRES under amino acid or glucose starvation (Fernandez *et al.*, 2001; Fernandez *et al.*, 2002). As stated above, under amino acid starvation, eIF2α phosphorylation by GCN2 kinase induces synthesis of an ITAF that binds the *CAT-1* IRES and initiates translation (Yaman *et al.*, 2003). In tumour cells under glucose deprivation, *CAT-1* IRES-dependent translation is also induced, but only through phosphorylation of eIF2α by the transmembrane endoplasmic reticulum kinase (PERK) (Fernandez *et al.*, 2002). Moreover, phosphorylation of eIF2α by GCN2 in response to amino acid deprivation also induces SNAT2 IRES-mediated translation (Gaccioli *et al.*, 2006). Growth factor deprivation conditions also induce IRES-mediated translation of specific transcripts. It is the case of the mRNA encoding XIAP and the sterol regulatory element-binding transcription factor 1 (SREBP-1), which are translated via an IRES in the absence of growth factors in tumour cells, thus protecting them from apoptosis (Damiano *et al.*, 2010; Riley *et al.*, 2010). IRES-mediated translation of these proteins is involved in cell survival under nutritional stress, and might constitute an advantage for cancer cell survival (Liu *et al.*, 2015). In addition, the anti-apoptotic protein XIAP is up-regulated under γ-irradiation via IRES-mediated translation, which makes tumour cells radiotherapy-resistant (Holcik *et al.*, 1999; Holcik *et al.*, 2000). Accordingly, it has been shown that inhibition of XIAP by RNA interference enhances chemotherapeutic drug sensitivity and decreases myeloma cell survival (Holcik *et al.*, 2000). In a different study, it was found that paclitaxel (PTX) — a drug commonly used in the chemotherapy of ovarian cancer — induces IRES-mediated translation of β-catenin in human ovarian cancer cell lines, which regulates the expression of downstream factors (c-Myc and cyclin D1), reducing PTX sensitivity (Fu *et al.*, 2015). Thus, the regulation of the IRES-dependent translation of β-catenin may be involved in the cancer cell response to PTX treatment (Fu *et al.*, 2015). Other anti-apoptotic proteins are also translationally controlled by IRES under oxidative and genotoxic stress. These include c-Myc cancer-associated transcription factor and Bcl-2-associated athanogene 1 (BAG-1) that strengthens tumour cells' resistance to DNA damage-inducing drugs (Yang *et al.*, 1999; Subkhankulova *et al.*, 2001; Dobbyn *et al.*, 2008; Ott *et al.*, 2013). In addition, synthesis of cIAP1 and Bcl-2 are enhanced by etoposide and arsenite treatments through IRES-mediated translation (Sherrill *et al.*, 2004; Van Eden, 2004). The transcriptional master regulator of the oxidative and genotoxic stress response p53 is also translated via IRES (Candeias *et al.*, 2006; Ray *et al.*, 2006; Khan *et al.*, 2015). Indeed, *p53* transcript has two IRES structures that control the translation of full-length p53 and an N-terminal-truncated isoform (Δ40p53) from the same

mRNA (Candeias *et al.*, 2006; Ray *et al.*, 2006). IRES-mediated translation of both isoforms is enhanced under different stress conditions that induce DNA damage, ionizing radiation and endoplasmic reticulum stress, oncogene-induced senescence and cancer. Polypyrimidine tract-binding protein (PTB), an ITAF, stimulates IRES-mediated translation of both p53 isoforms in response to doxorubicin, following PTB relocalisation from the nucleus to the cytoplasm (Grover *et al.*, 2008). This regulation is altered in the presence of melanoma-associated mutations in the *p53* 5'UTR (Khan *et al.*, 2013). In addition, human double minute 2 homolog (HDM2) and HDM4 act as other ITAFs that synergistically increase *p53* IRES activity under DNA damage following HDMX phosphorylation by ATM (Malbert-Colas *et al.*, 2014). On the other hand, it was shown that glucose depletion induces *p53* IRES activity of both isoforms through the involvement of the scaffold/matrix attachment region-binding protein 1 (SMAR1), a protein predominantly nuclear that becomes abundant in the cytoplasm under glucose deprivation, while PTB does not show nuclear-cytoplasmic relocalisation highlighting the novelty of SMAR1 functioning as an ITAF under stress (Khan *et al.*, 2015). Other ITAFs have been reported to control *p53* IRES activity, such as eIF4G2 (also known as DAP5), Annexin A2 and PTB-associated Splicing Factor (PSF) (Sharathchandra *et al.*, 2012; Weingarten-Gabbay *et al.*, 2014). Furthermore, a different mechanism of *p53* inactivation that links deregulation of IRES-mediated p53 translation with tumourigenesis was identified in two breast cancer cell lines. Here, the connection between IRES-mediated p53 translation and p53 tumour suppressive function was established through the identification of two new *p53* ITAFs — translational control protein 80 (TCP80) and RNA helicase A (RHA) — that positively regulate *p53* IRES activity. Indeed, these two cell lines proved to retain wild-type p53 but exhibit defective p53 induction and synthesis following DNA damage, as the levels of TCP80 and RHA are extremely low in both cell lines, and expression of both proteins is required to significantly increase *p53* IRES activity (Halaby *et al.*, 2015a; Halaby *et al.*, 2015b). NRF2 is another master regulator of the response to oxidative stress, which is translationally induced through an IRES under oxidative stress (Li *et al.*, 2010c; Shay *et al.*, 2012; Zhang *et al.*, 2012). While NRF2 synthesis is blocked under basal conditions due to the presence of a highly structured inhibitory hairpin element present in its 5'UTR, its synthesis is enhanced by oxidative stress through stimulation of an IRES element also present within its 5'UTR (Zhang *et al.*, 2012). IRES-mediated translation of NRF2 requires La autoantigen ITAF binding (Zhang *et al.*, 2012). Examples of other transcription factors induced by oxidative and genotoxic stress through IRES-mediated translation are the octamer-binding protein 4 (OCT4), which is synthesized upon H_2O_2 treatment in breast cancer and liver carcinoma cells (Wang *et al.*, 2009), and runt-related transcription factor 2 (RUNX2), whose translation is stimulated by mitomycin C (Xiao *et al.*, 2003). All these examples support a model whereby, under oxidative and genotoxic stress, IRES-mediated translation of key regulators and pro-survival factors provide tumour cells with mechanisms for attaining resistance to chemotherapy and radiation (Leprivier *et al.*, 2015).

On the other hand, the presence of IRES within transcripts coding tumour suppressor proteins can help the cell maintain the levels of these proteins and prevent cancer outbreak.

The oncogene-induced senescence (OIS), a critical cellular response that counteracts cellular transformation, is characterised by cell cycle arrest and induction of *p53*, thus restraining the proliferative potential of preneoplasic clones (Serrano *et al.*, 1997). Bellodi *et al.* (2010) have demonstrated that during OIS, there is a switch from cap-dependent translation to IRES-dependent translation, during which an IRES element positioned in the 5'UTR of *p53* is engaged to promote p53 translation and specialised translational control of mRNA, such as *p53*, hence provides a molecular barrier for cellular transformation. Montanaro *et al.* (2010) showed that increased p53 activity in breast cancer is dependent on dyskerin-mediated increase in IRES-mediated translation, but independent of effects on telomerase.

Expression induction of the aforementioned proteins provides a key factor for cancer cells to survive and proliferate under stress conditions, demonstrating the importance of IRES-mediated translation in the process of tumourigenesis and how the IRES structures may be considered important targets in the treatment of cancer.

1.5.2. IRES-related therapies

Combined gene therapy has emerged a few years ago as a promising strategy to improve treatments of many conditions, including cancer, cardiovascular and degenerative diseases. A significant feature of IRES elements is their ability to start protein synthesis via internal initiation, which allows multicistronic vectors expressing several genes from a single mRNA to be designed. IRES-mediated translation can occur under stress conditions, making IRES useful for therapeutical approaches, namely the IRES-based multicistronic vector concept (figure 1.12) (Renaud-Gabardos, 2015). The IRES-based expression cassette contains several genes, separated by IRES, controlled by the same promoter. This transcription unit gives rise to a single mRNA that codes the different genes. Translation initiation occurs at the 5' end via the cap-dependent mechanism, resulting in translation of the first ORF. Internal initiations of translation occur at each IRES, making the other ORFs to be translated. Thus, the multicistronic mRNA generates several proteins from a single transcription unit, allowing more stable, long-term expression and stable transgene ratio.

The first biomedical use of IRES in an expression vector was that of interleukin 12 subunit co-expressed with a gene of resistance to neomycin (Zitvogel *et al.*, 1994). Over the last decade, several studies have validated this concept using a cocktail of two vectors to simultaneously transfer two genes — a well-documented approach in the field of cardiovascular diseases and cancer, with therapeutic benefits obtained in various animal models, using different combinations of angiogenic and anti-angiogenic factors (Scappaticci *et al.*, 2001; Ohlfest *et al.*, 2005; Kupatt *et al.*, 2010). Moreover, a bicistronic IRES-based vector, co-expressing FGF2 and VEGF-A, has been assessed in a clinical assay of gene therapy on patients with refractory coronary disease (Kukuła *et al.*, 2011).

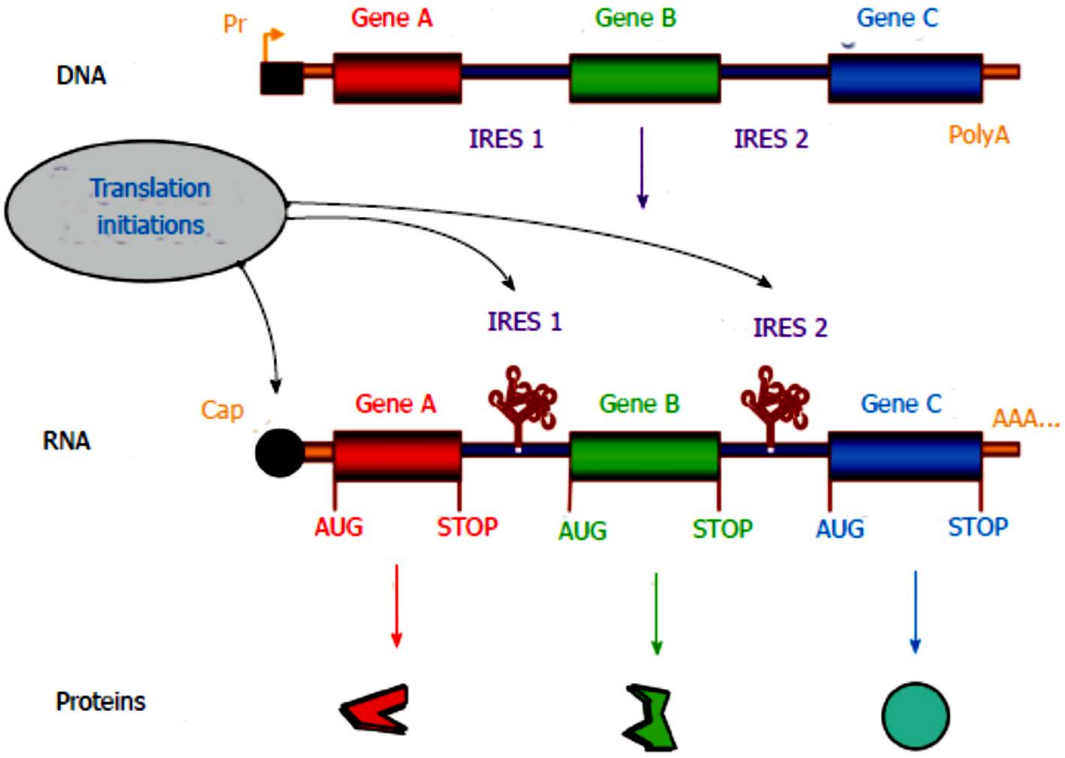


Figure 1.12 — IRES-based multicistronic vector concept. The IRES-based expression cassette contains several genes, separated by IRES, controlled by the same promoter. This transcription unit gives rise to a single mRNA that codes the different genes. Translation initiation occurs at the 5' end via the cap-dependent mechanism, resulting in translation of the first ORF (A). Internal initiations of translation occur at each IRES, making the other ORFs to be translated (B and C). Thus, the multicistronic mRNA generates several proteins from a single transcription unit, allowing more stable, long-term expression and stable transgene ratio. Adapted from Renaud-Gabardos (2015).

In 2012, Villaflores and co-workers developed an assay system using the bicistronic reporter constructs in the identification of compounds with activity against translation directed by amyloid-β precursor protein (APP) and tau IRES. This study aimed to determine the effects of curcumin and demethoxycurcumin on the IRES of APP and tau protein for screening of anti-Alzheimer's disease (AD) agents. They performed a bicistronic assay wherein the expression of the first cistron — a β-galactosidase gene under the control of a cytomegalovirus promoter — represents the canonical cap-dependent mechanism of translation initiation, while translation of the second is driven by the *APP* or the *tau* IRES elements in order to drive the expression of secreted alkaline phosphatase (SEAP) under a cap-independent mechanism. Bioactive natural products reported to have therapeutic potential for AD, such as curcumin and demethoxycurcumin, were screened in a murine neuroblastoma (N2A) cell model. Western blot analyses for the expression of APP C-terminal protein, human tau-1, and phosphorylated tau at serine 262 (pS262) and serine 396 (pS396) were performed after treatment of N2A cells with the test compounds. The results suggested that curcumin may play a role in AD pathology

alleviation through the APP and tau IRES-mediated translation mechanism, whereas demethoxycurcumin was observed to inhibit phosphorylation of both tau pS262 and pS396. These results reinforce the potential of the mentioned compounds as prophylactic and therapeutic anti-AD agents.

On the other hand, some diseases emerge because some proteins evade cell mechanisms of protein synthesis arrest and keep being produced via IRES-mediated translation initiation. This mechanism is considered a significant contributor to malignant phenotypes and chemoresistance. Therapeutic approaches that inhibit IRES-mediated translation initiation of proteins implicated in malignant phenotypes may interfere with this specialised mode of protein synthesis and therefore impair the growth and development of tumours. XIAP is an important member of the inhibitor of apoptosis protein (IAP) family that binds specifically to, and inhibits, the activated forms of caspases 3, 7 and 9 — the enzymes that induce the intrinsic (mitochondrial) apoptotic pathway, which is the major cell death mechanism that is triggered by radiotherapy and many chemotherapy drugs (Schimmer, 2004). Its expression is uniquely regulated by an IRES-dependent mechanism at translational level (Holcik *et al.*, 1999), which is activated when cells undergo stress, such as during chemotherapy (Lewis and Holcik, 2005). In a 2009 study, Gu *et al.* found that the MDM2 RING domain protein binds to the XIAP IRES, increasing IRES-mediated XIAP translation, which results in resistance to anticancer treatment. Recently, the same team found that binding of XIAP IRES to the MDM2 RING domain protein inhibited its ability for self-association and self-ubiquitination, which increased MDM2 protein stabilisation and cancer cell survival (Liu *et al.*, 2015). This study identified a new IRES RNA that interacts with MDM2 protein and regulates its stabilisation, suggesting that targeting of MDM2 through disruption of MDM2 protein-RNA interaction might be a useful strategy to develop novel anti-cancer therapeutics.

In an attempt to identify compounds capable of selectively inhibiting translation mediated through the insulin-like growth factor 1 receptor (IGF1R) IRES, Vaklavas *et al.* (2015) performed a cell-based empirical high-throughput screen. Results obtained using the bicistronic reporter system demonstrated selective inhibition of downstream cistron translation. Moreover, the identified compound and its structural analogues completely blocked *de novo* IGF1R protein synthesis in genetically unmodified cells, confirming activity against endogenous IRES. Their spectrum of activity extends beyond IGF1R to include the *c-myc* IRES. The small molecule IRES inhibitor differentially modulates synthesis of the oncogenic (p64) and growth-inhibitory (p67) isoforms of *Myc*, suggesting that the IRES controls not only translational efficiency, but also chooses the initiation codon. Sustained IRES inhibition has profound, detrimental effects on human tumour cells, inducing massive (>99%) cell death and complete loss of clonogenic survival in models of triple-negative breast cancer. The results begin to reveal new insights into the inherent complexity of gene-specific translational regulation, and the importance of IRES-mediated translation to tumour cell biology.

Also, IRES elements mediating translation of viruses causing lethal diseases in humans can be targeted. The internal ribosome entry site (IRES) of hepatitis C virus (HCV), which

governs the initiation of protein synthesis from viral RNA, represents an ideal target for antisense approaches. After establishing the sequence responsible for translational activity of *HCV* IRES, Alotte *et al.* (2008) designed five 6–10mer antisense molecules, i.e. short peptide nucleic acids (PNA), that strongly inhibited the highly conserved IIId or IV loop regions of the IRES in a rabbit reticulocyte lysate assay. This inhibition was highly specific, since corresponding PNAs with only one mismatch were inactive. A follow-up on this matter revealed that phosphorothioate antisense oligonucleotide 1 (S-ODN1) is completely efficient on HCV translation inhibition in hepatoma cells, but only partially efficient in peripheral blood mononuclear cells (PBMC) (Youssef *et al.*, 2014).

Another example of a drug inhibiting virus IRES-dependent translation is idarubicin[13] (IDR), a broad-spectrum enterovirus replication inhibitor that selectively targets enterovirus 71[14] (EV71) (Hou *et al.*, 2016). This study identified IDR as effectively blocking the synthesis of viral protein and RNA of EV species. Moreover, anthracyclines were shown to suppress *EV* IRES-mediated translation, but not that of *p53* IRES. In addition, IDR impaired binding between the *EV71* IRES RNA and hnRNP A1, a known host ITAF. All in all, this study identified an approved anticancer drug newly labelled as a selective *EV* IRES-binder and -inhibitor, providing leads for the development of novel antiviral therapies directed at the *EV* IRES RNA.

Not only antisense oligonucleotides but also small molecule compounds can be used to regulate IRES-mediated translation. According to Cammas *et al.* (2015), stabilisation of G-quadruplex at the VEGF IRES represses cap-independent translation, and the amount of human VEGF mRNA associated with polysomes decreases in the presence of a highly selective stabilising G-quadruplex ligand, resulting in reduced VEGF protein expression. These results uncover the existence of functionally silent G-quadruplex structures that are susceptible to conversion into efficient repressors of cap-independent mRNA translation. Together, these findings have implications for the *in vivo* uses of G-quadruplex-targeting compounds and for anti-angiogenic therapies.

1.6. Function of the proteins encoded by the transcripts studied in this work

As we will show, in the early stages of this work the main task was to identify *in silico* human proteins whose characteristics suggest their expression is regulated at translational level. According to the primary results obtained (c.f. Results, section I), UPF1 (up-frameshift 1 regulator of nonsense transcripts homolog), AGO1 (Argonaute RISC catalytic component 1) and MLH1 (mutL Homolog 1) were selected as putative candidates. In the following paragraphs, a characterisation of these proteins will be presented in order to better understand the results.

[13]Idarubicin is an anthracycline compound that is used therapeutically for certain types of tumour.

[14]This virus causes life-threatening diseases with neurological manifestations in young children, but whose treatment remained an unmet medical need (Hou *et al.*, 2016).

1.6.1. Human up-frameshift 1 (UPF1)

In 1997, Applequist *et al.* identified the first mammalian homologue of yeast UPF1. Human UPF1 (UPF1) is an evolutionarily conserved and ubiquitously expressed 130kDa phosphoprotein with RNA/DNA-dependent ATPase and RNA helicase activities, and contains 2 zinc finger motifs, 1 DEAD box and post-translational modifications in 6 amino acids (figure 1.13). It is encoded by the *UPF1* gene located in chromosome 19 (p13.2–p13.11, www.ncbi.nlm.nih.gov). The sequence identities among human, plant, fruit fly, nematode, and yeast UPF1 are between 40–62%, and reach over 90% among zebrafish, mouse and human, which makes UPF1 a highly conserved protein throughout eukaryotes (Culbertson and Leeds, 2003), suggesting it plays a key role in biological systems. The roles of UPF1 are quite diverse in mammalian cells and include RNA stability, DNA repair, cell cycle progression, DNA replication, and telomere metabolism, as further detailed below. UPF1 is indeed essential for embryonic viability in plant, fruit fly, zebrafish, and mice, and its loss-of-function inhibits cell growth and induces apoptosis in *Drosophila melanogaster* (Avery *et al.*, 2011). It shuttles between the nucleus and the cytoplasm via chromosomal maintenance 1 (CRM1)[15]; this characteristic conveys potential roles in both the nucleus and cytoplasm (Mendell *et al.*, 2002).

UPF1 has been initially characterised as an essential factor for NMD, a mechanism required for regulation of gene expression, and also a surveillance mechanism for rapid degradation of aberrant mRNA (cf. section 1.2.3.1) (Schoenberg and Maquat, 2012).

Apart from its function in NMD, UPF1 is also involved in Staufen1-mediated mRNA decay (SMD) through the direct binding with STAU1 (Kim *et al.*, 2005). Studies by Gong *et al.* (2009) revealed that SMD and NMD pathways fight over UPF1. STAU1-binding domain within UPF1 overlaps with UPF2, a core factor of NMD. Knock-down of STAU1, which inhibits SMD, increases the NMD activity, whereas knock-down of UPF2, which, in turn, inhibits NMD, increases SMD. Thus, the interaction between SMD and NMD pathways forms an important gene expression network, where UPF1 plays a central role. It is also involved in histone mRNA degradation through an interaction with stem-loop binding proteins at the end of S phase, or after the inhibition of DNA synthesis[16] (Kaygun and Marzluff, 2005). Its function in the triggering of this process is regulated by UPF1 phosphorylation (Kaygun and Marzluff, 2005; Müller *et al.*, 2007). Apart from its role in cell transcript degradation, several other functions have been ascribed to UPF1.

[15]Chromosomal maintenance 1, also known as Exportin 1, is the major mammalian export protein that facilitates the transport of large macromolecules including RNA and protein across the nuclear membrane to the cytoplasm (Nguyen *et al.*, 2012).

[16]Transcripts encoding histone proteins lack polyadenylated tails, although they are transcribed by RNA polymerase II (Kaygun and Marzluff, 2005). This conjures up an image of presence of special mechanisms for the regulation of histone mRNA stabilities (Imamachi, 2012). Actually, 3'UTRs of replication-dependent histone mRNA harbour the special stem-loop structure that is required for rapid regulatory degradation of histone mRNA (Kaygun and Marzluff, 2005; Marzluff *et al.*, 2008). The structure at the 3' end of histone mRNA interacts with the stem-loop binding proteins (SLBP) (Marzluff, 2005; Marzluff *et al.*, 2008).

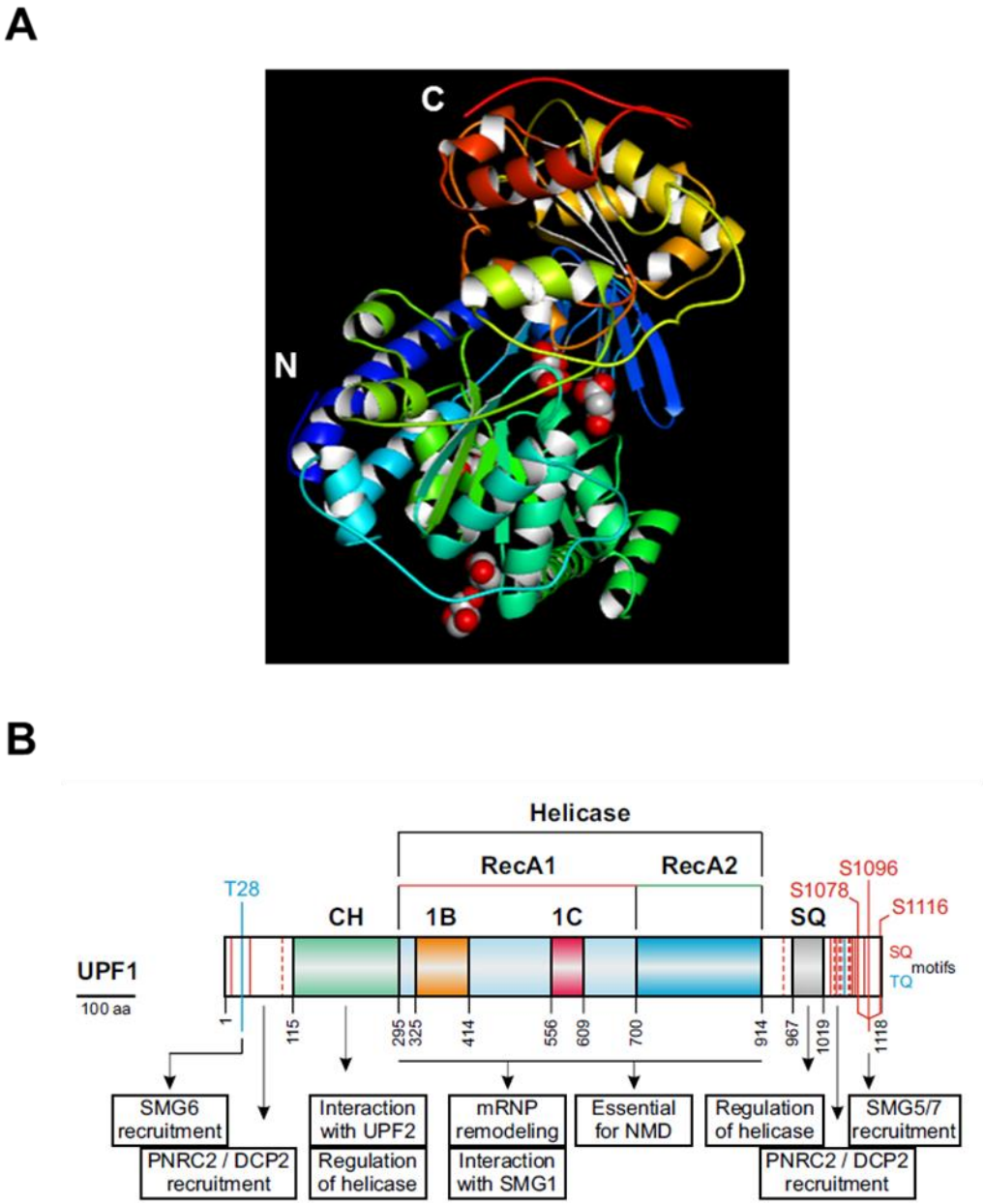


Figure 1.13 — Human up-frameshift 1 (UPF1) protein. (A) Schematic representation of UPF1 3-dimentional crystal structure (data from Protein Data Bank Japan, http://pdbj.org/). N is the N-terminal domain; C is the C-terminal domain (Chakrabarti *et al.*, 2011). (B) Schematic representation of the domain architecture cysteine-histidine-rich domain (CH, in green), the helicase core domains (RecA1 and 2, 1B and 1C, and UPF1 C-terminal unstructured region containing S/T-Q phosphorylation motifs (SQ) are indicated. Adapted from Fatscher *et al.* (2015).

Studies by (Azzalin and Lingner, 2006a; Azzalin and Lingner, 2006b) revealed UPF1 physically interacts with the p66 subunit and the p125 catalytic subunit of DNA polymerase δ[17], being crucial to S phase progression and DNA replication in an NMD-independent manner. They found that 4% of UPF1 proteins were bound to chromatin-associated protein fraction,

[17]DNA polymerase δ is involved in DNA replication and repair, and is the primarily used enzyme in both leading and lagging strand synthesis (Johnson *et al.*, 2015).

whereas UPF1 mostly exists in the soluble fraction. The amount of chromatin-associated UPF1 is low in M and early G1 phases, starts to increase in mid-G1, and is at its highest level during S phase (Imamachi, 2012). Depletion of UPF1 results in an early S phase arrest and stalls replication fork progression (Azzalin and Lingner, 2006a; Azzalin and Lingner, 2006b). Thus, UPF1 may be involved in DNA damage response during the S phase of the cell cycle, as the depletion of this protein also induces the accumulation of nuclear foci containing a sensitive marker for DNA damage, such as phosphorylated histone H2AX (γ-H2AX) (Azzalin and Lingner, 2006a; Azzalin and Lingner, 2006b).

Another role for UPF1 is related to telomere homeostasis. Telomeres, essential DNA-protein complex located at the end of linear eukaryotic chromosomes, are essential to chromosome stability. Although previously considered transcriptionally silent, mammalian telomeres are transcribed into telomeric repeat-containing RNA (TERRA) (Azzalin *et al.*, 2007). SMG proteins (effectors of NMD), are enriched at telomeres *in vivo*, negatively regulate TERRA association with chromatin, and protect chromosome ends from telomere loss. Thus, depletion of the NMD factors SMG1 and UPF1 results in a dramatic accumulation of telomere-bound TERRA, while total TERRA levels and turnover rate are unaffected (Azzalin *et al.*, 2007). Further, efficient replication of leading strand telomeres has been shown to require human UPF1 (Chawla *et al.*, 2011), as depletion of UPF1 results in fragile telomeres, a phenotype reflective of telomere replication-associated defects (Sfeir *et al.*, 2009), specifically involving leading strand telomeres. Based on these data, Azzalin and co-workers proposed a model in which UPF1 is required to the complete replication of telomeric DNA; they also suggested that in UPF1-depleted cells, replication fork progression through telomeric DNA is halted, generating DNA damage and single-stranded DNA, which eventually degenerates into DNA double strand breaks (DSBs), leading to the loss of entire telomeric tracts (Azzalin, 2012). Interestingly, yeast UPF1 is thought to localise exclusively to the cytoplasm (Atkin *et al.*, 1995), indicating that nuclear functions associated with UPF1 (cell cycle progression and regulation of telomere homeostasis) emerged late during evolution and might represent a unique feature of mammalian UPF1 proteins (Azzalin, 2012).

UPF1 has also been associated with human immunodeficiency virus type 1 (HIV-1) genomic RNA stability in an NMD-independent manner (Ajamian *et al.*, 2008). The *HIV-1* RNP consists of *HIV-1* genomic RNA, $pr55^{Gag}$ (the major structural protein), and STAU1 (the host protein) (Chatel-Chaix *et al.*, 2004). However, a more recent study revealed that UPF1 is one of the HIV-1 RNP components and is involved in *HIV-1* genomic RNA stability (Ajamian *et al.*, 2008). Based on the observation that the abundance of UPF1 was enhanced in the HIV-1 RNP, this study examined the function of UPF1 during *HIV-1* gene expression and showed that UPF1 knock-down resulted in a catastrophic decrease in *HIV-1* RNA and $pr55^{Gag}$ expression. The obtained results indicate that UPF1 enhances *HIV-1* mRNA translatability (Ajamian *et al.*, 2008). The function of UPF1 in HIV-1 expression is independent of the one in NMD, thus identifying novel functions for UPF1 in the maintenance of *HIV-1* RNA stability, and strongly supporting an essential role for this protein (Ajamian *et al.*, 2008). A follow-up on this matter revealed that HIV-

1 ensures nuclear export of the genomic RNA by recruiting UPF1, but excluding UPF2 (Ajamian *et al.*, 2015). In this report, the authors characterised the importance of the nuclear interaction between UPF1 and the genomic *HIV-1* RNA. They demonstrated that UPF1 shuttling promotes the nucleocytoplasmic export of genomic *HIV-1* RNA. By using *in situ* imaging analyses and *in silico* modelling of protein-protein interactions, they revealed that the association between UPF1 and UPF2 is of the utmost importance in the regulation of genomic *HIV-1* RNA nucleocytoplasmic export. Since UPF1 is a component of the Staufen1/HIV-1 RNP complex that excludes UPF2[18] (Milev *et al.*, 2012), it is possible that *HIV-1* mediates the association between UPF1 and Staufen1, blocking the ability of UPF2 to associate with UPF1 (Maquat and Gong, 2009). Hence, high expression levels of UPF2 would lead to the formation of UPF1-containing cytoplasmic complexes and limit the availability of UPF1 in the nucleus, resulting in a blockage of genomic *HIV-1* RNA export that is dependent on UPF1 nucleocytoplasmic shuttling (Ajamian *et al.*, 2015). Furthermore, it has been shown that UPF1 is crucial for the infectivity of HIV-1 progeny virions (Serquina *et al.*, 2013). The infectivity of HIV-1 virions produced in UPF1-depleted cells, or in cells expressing ATPase-defective UPF1 mutants, is markedly impaired due to a defect at the level of reverse transcription following entry into a new target cell, suggesting that UPF1 promotes an early post-entry step in HIV-1 replication (Serquina *et al.*, 2013). Thus, it is conceivable that UPF1 has a direct role in HIV-1 replication via the annealing of the tRNA primer to the viral genome, which is required to initiate reverse transcription (Serquina *et al.*, 2013). This is consistent with studies that have established that the efficiency of tRNA annealing and its ability to prime reverse transcription can both be promoted by a cellular RNA helicase (Xing *et al.*, 2011). Another possibility is that UPF1 would serve to remodel the viral RNP to facilitate reverse transcription, suggesting that UPF1 could act as an RNPase (Serquina *et al.*, 2013), a notion supported by the observation that its ATPase activity is required for the removal of proteins from partially degraded NMD substrates (Franks *et al.*, 2010).

1.6.2. Human argonaute RNA-induced silencing complex (RISC) catalytic component 1 (AGO1) protein

Argonaute proteins (AGOs) are essential effectors in RNA-mediated gene silencing pathways (Ender and Meister, 2010). They are ~100-kDa highly basic proteins that contain two common domains, PIWI/Argonaute/Zwille (PAZ) and P-element-induced whimpy testes (PIWI) (Cerutti et al., 2000). The first domain — consisting of 130 amino acids — has been identified in Argonaute proteins and Dicer (Bernstein *et al.*, 2001). Although it has no defined function, PAZ is thought to be a protein–protein interaction domain, potentially mediating both homo- and hetero-dimerization (Cerutti *et al.*, 2000). The C-terminal 300-amino acid PIWI domain also has no known function, but is highly conserved. There are eight AGO-like proteins in human cells grouped in two families, according to their sequence: the eIF2C/AGO subfamily and the PIWI

[18] UPF2 and Staufen1 compete for the same binding region in UPF1 (Maquat and Gong, 2009).

subfamily (Ender and Meister, 2010). AGO1, encoded by *eIF2C1* gene located on chromosome 1 (p35–p34, www.ncbi.nlm.nih.gov), displays the same domain architecture found in all Argonaute proteins, namely the four primary domains N, PAZ, Mid and PIWI with two linker regions L1 and L2 (figure 1.14) (Faehnle *et al.*, 2013). Argonaute family proteins have a role not only in RNAi, but also in developmental control, stem cell maintenance and tumourigenesis (Carmell *et al.*, 2002). Argonaute family proteins assemble with small RNA, including microRNA (miRs), small interfering RNA (siRNA), and PIWI-interacting RNA (piRNA), into the effector complex RISC, which mediates sequence-specific target gene silencing (Kobayashi and Tomari, 2016). RISC assembly is not a simple binding between a small RNA and AGO; rather, it follows an ordered multi-step pathway that requires specific accessory factors. Some steps of RISC assembly and RISC-mediated gene silencing are dependent on, or facilitated by, particular intracellular platforms, suggesting their spatial regulation (Kobayashi and Tomari, 2016).

Although mammalian AGO isoforms (AGO1–4) are considered to be functionally redundant as far as loading of miRs is concerned — immunoprecipitates with antibodies against individual isoforms contain nearly identical spectra of miRs —, there are some exceptions (Burroughs *et al.*, 2011; Dueck *et al.*, 2012). For instance, miR-451 is exclusively loaded on to, and processed by, AGO2 in a Dicer-independent manner (Yang *et al.*, 2010), whereas non-miR small RNA have been found to be specifically associated with AGO1 (Yamakawa *et al.*, 2014). Specific AGO-miR complexes may have different silencing effects on the same mRNA (Ghosh and Adhya, 2016). According to this study, depletion of either AGO1 or miR-1 resulted in early elevation of Ccnd1 mRNA, but there was no effect on the onset time of cyclin D1 (Ccnd1) translation; conversely, down-regulation of AGO2 affected the onset of Ccnd1 translation, but had no effect on mRNA levels. Thus, loss of the mir1-AGO1 complex had two distinct effects on Ccnd1 mRNA: it resulted in up-regulation at early times and led to inhibition of the normal rate of accumulation. One of the possible explanations for the effect on transcript accumulation is that the miR1-AGO1 complex up-regulates Ccnd1 transcription. Indeed, AGO1 has been reported to be associated with RNA polymerase II and to bind in close proximity to the transcription start site of a number of cell cycle genes, including Ccnd1 (Huang *et al.*, 2013).

It is apparent that Argonaute proteins are involved in the development of several tissues in different organisms. Actually, these proteins are also thought to have regulatory functions in stem cell self-renewal, including cancer stem cells. Alterations in Argonaute protein function have been shown to affect stem cells in a variety of tissues in a disparate group of organisms, indicating that this protein family may be part of the most basic mechanisms governing stem cell fate (Carmell *et al.*, 2002). Studies in *Drosophila* revealed that overexpression of AGO1 protein leads to germline stem cell (GSC) overproliferation, whereas loss of AGO1 results in the loss of GSCs (Yang *et al.*, 2007). Given that AGO1 serves as a key component of the miRNA pathway, these authors propose that an AGO1-dependent miRNA pathway probably plays an instructive role in repressing GSC/cystoblast differentiation. Adding to this, in *Arabidopsis*, AGO1, and its

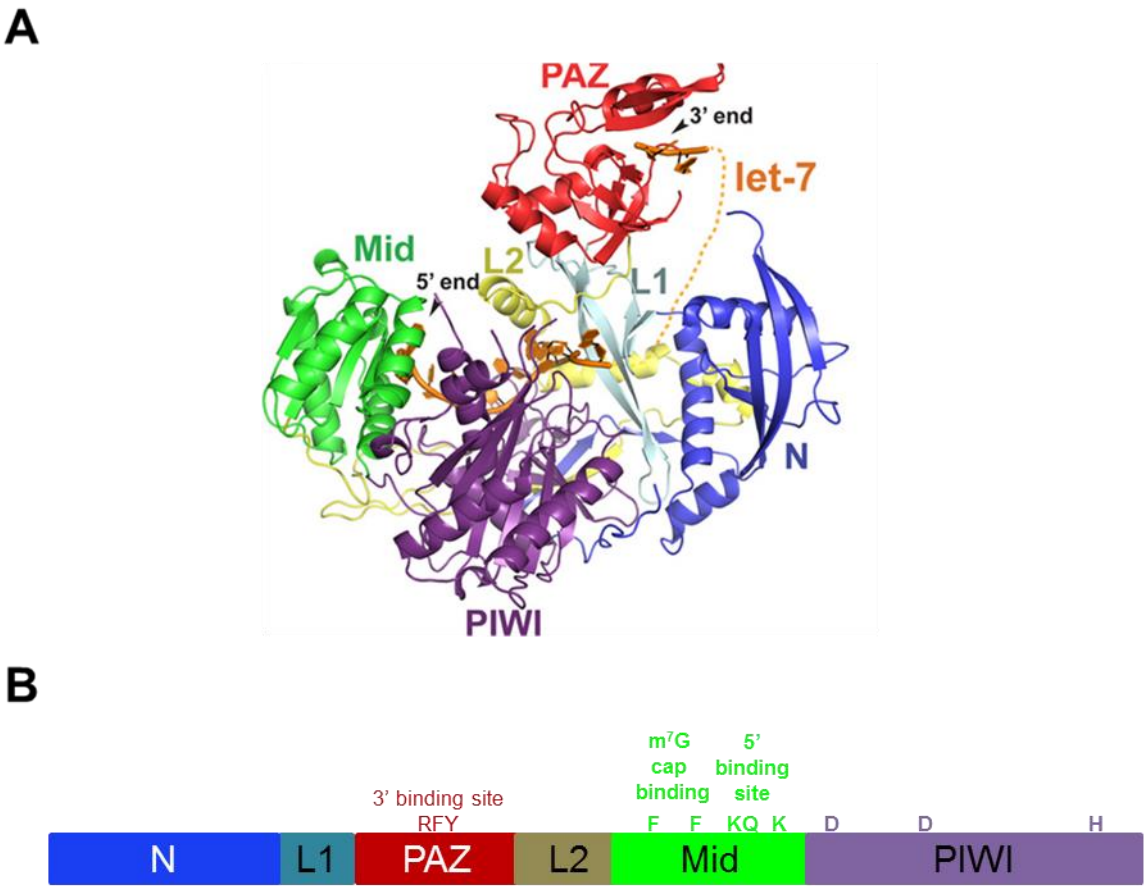


Figure 1.14 — Human Argonaute RNA-induced silencing complex (RISC) catalytic component 1 (AGO1) protein. (A) Schematic representation of AGO1 3-dimentional crystal structure in complex with lethal-7 (let-7) guide RNA. The individual domains of AGO1 are labelled and colour-coded. Let-7 miRNA is shown as an orange cartoon. Nucleotides (nts) 1–10 stretch from the middle (MID) domain and pass through L2 (linker region), the P-element-induced whimpy testes (PIWI) domain and L1 (linker region). A dashed line indicates the projected path of the disordered nts 11–20. Nts 21 and 22 are modelled in the PIWI/Argonaute/Zwille (PAZ) domain. Adapted from Faehnle *et al.* (2013). (B) Schematic representation of the human AGO1 protein. In the PAZ domain, residues important for the binding of small RNA 3' ends are indicated (R, arginine; F, phenylalanine; Y, tryptophan); in the Mid domain, the residues required for 5' end binding to small RNA and to the 7-methylguanine (m^7G) cap of target mRNA are shown (K, lysine; Q, glutamine); the PIWI domain catalytic residues are shown. Adapted from Höck and Meister (2008).

homologue AGO10, are required to maintain the correct temporal programme of floral stem cells (Ji *et al.*, 2011). In mammals, the *eIF2C1* gene encodes AGO1, a member of the former subfamily that is ubiquitously expressed at low-to-medium levels and highly conserved during evolution, reflecting its important physiological roles (Koesters *et al.*, 1999). AGO1 is 80% identical to AGO2 but lacks a key catalytic residue and cannot efficiently cleave RNA. It is associated with the loading of specific small RNA derived from the Epstein-Barr virus (Yamakawa *et al.*, 2014), and AGO1 and/or AGO3 is/are required for optimal resistance to influenza-A in mice (Van Stry *et al.*, 2012). Little is known about the function of AGO1 except that its overexpression slows neuroblastoma growth (Parisi *et al.*, 2011). Human AGO1

homologue, *eIF2C1*, was also identified in a screen of genes involved in Wilm's tumours (Dome and Coppes, 2002). Notably, it is expressed at low-to-medium levels in most tissues, but its expression is particularly high in embryotic kidney and lung, and also in tumours that lack the Wilm's tumour suppressor gene WT1 (Carmell *et al.*, 2002). Moreover, studies by Li *et al.* (2010a) concluded that eIF2C1 protein is overexpressed in colorectal cancer when compared to adjacent non-cancer tissue. Together, these findings could make human *eIF2C1* an interesting candidate gene to be involved in neoplastic development. It should be noted that positive reaction to each AGO in colon cancer tissue was significantly higher than that in adjacent non-cancerous tissues. The relationship of AGO subfamily with colon cancer has not been made fully clear yet. Perhaps through RNAi-related pathways or distinct mechanisms, AGO subfamily members have an important role in the progression of colon cancer (Li *et al.*, 2010a).

1.6.3. Human MutL homolog 1 (MLH1)

Human MLH1 (MLH1) is a 756-amino acid, 84 kDa protein that can be divided roughly into two halves: an N-terminal domain (NTD), where the ATPase activity resides, and a C-terminal domain (CTD), which is the site of dimerization with MLH1 paralogs (figure 1.15) (Guerrette *et al.*, 1999; Wu *et al.*, 2015). It is encoded by the *MLH1* gene located on chromosome 3 (p21.3; www.ncbi.nlm.nih.gov). It is a human homolog of the *E. coli* DNA mismatch repair gene, mutL, which mediates protein-protein interactions during mismatch recognition, strand discrimination, and strand removal (Li, 2008). It undergoes alternative splicing, which results in multiple transcript variants, encoding distinct isoforms (Genuardi *et al.*, 1998). Additional transcript variants have been described, but their full-length nature is yet to be determined (www.ncbi.nlm.nih.gov).

MLH1 protein is one of seven components of a DNA mismatch repair (MMR) system of proteins (MLH1, MLH3, MSH2, MSH3, MSH6, PMS1 and PMS2) that work co-ordinately in sequential steps to initiate repair of DNA mismatches in humans (Pal *et al.*, 2008). The main components of this repair system are proteins MLH1, MSH2, MSH6 and PMS2, which interact to recognise mismatches and excise them, therefore allowing resynthesis and religation of DNA strand by DNA polymerase δ and DNA ligase (Vilar and Gruber, 2010). Loss of function of one of the MMR system proteins is responsible for a deficient MMR system, leading to the accumulation of frameshift mutations (insertions/deletions) in microsatellites[19], which results in a genetic instability (Buecher *et al.*, 2013). Microsatellite instability (MSI) phenotype and/or loss of MMR protein expression, also known as deficient MMR (dMMR) phenotype, may have tumourigenic potential when occurring in coding regions of key genes involved in several cellular function and pathways (Turnpenny and Ellard, 2012).

This gene was recognised as a frequently mutated locus in hereditary nonpolyposis colon cancer (HNPCC). Yet, many cancers were identified as MLH1-deficient: stomach (Waki *et al.*,

[19]Microsatellites are short-tandem DNA repeat sequences of 1–6 bases, distributed throughout the genome (in coding and non-coding regions), which, due to their repeated structure, are especially prone to replication errors that are normally repaired by the MMR system (Turnpenny and Ellard, 2012).

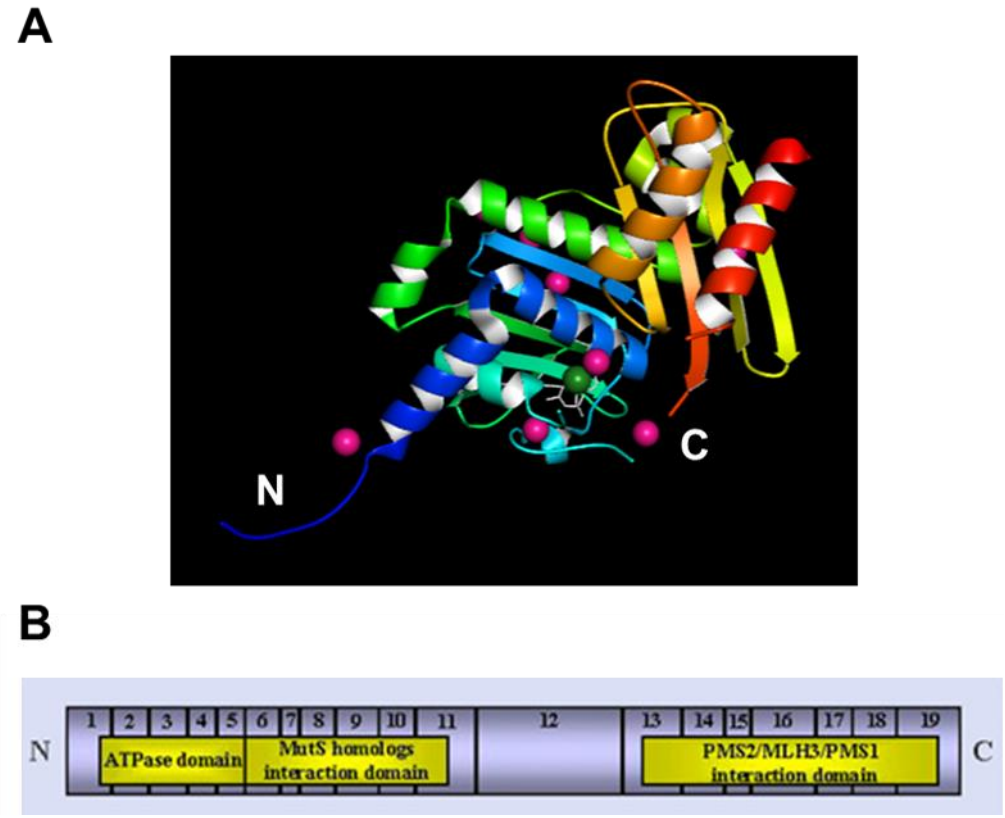


Figure 1.15 — Human MutL homolog 1 (MLH1) protein. (A) Schematic representation of MLH1 3-dimentional crystal structure. Data from Protein Data Bank Japan, http://pdbj.org/). N is the N-terminal domain; C is the C-terminal domain (Wu *et al.*, 2015). (B) Diagram of the MLH1 protein in scale. Each number inside a grey box indicates the exon from which each part of the protein is translated. The three yellow boxes represent the ATPase domain, the MutS homologs (MSH2, MSH3, MSH6) interaction domain and the PMS2/MLH3/PMS1 interaction domain; C is the carboxyl-terminal; N is the amino-terminal. Data from Atlas of Genetics and Cytogenetics in Oncology and Haematology, http://atlasonline.critt-informatique.fr/ accueil.aspx.

2002; Wani *et al.*, 2012; Kupčinskaitė-Noreikienė *et al.*, 2013), oesophageal (Chang *et al.*, 2015), head and neck squamous cell carcinoma (HNSCC) (Zuo *et al.*, 2009; Tawfik *et al.*, 2011), and non-small cell lung cancer (NSCLC) (Safar *et al.*, 2005). Although only a minority of sporadic cancers with a DNA repair deficiency have a mutation in a DNA repair gene, a majority of sporadic cancers with a DNA repair deficiency do have one or more epigenetic alterations that either reduce or silence DNA repair gene expression (Bernstein, 2015). Most of the deficiencies of MLH1 found in these cancers were due to methylation of the promoter region of MLH1 gene; nevertheless, another epigenetic mechanism reducing MLH1 expression is over-expression of miR-155, which targets MLH1 and MSH2 (Valeri *et al.*, 2010). These authors found an inverse correlation between the expression of miR-155 and the expression of MLH1 or MSH2 proteins in human colorectal cancer, and that a number of MSI tumours with unknown cause of MMR inactivation displayed miR-155 overexpression, providing support for miR-155 modulation of MMR as a mechanism of cancer pathogenesis. In an attempt to describe the frequency of MLH1 promoter methylation in colorectal cancer (CRC), Li *et al.* (2013) explored

the associations between MLH1 promoter methylation and clinicopathological and molecular factors. They found out that MLH1 promoter methylation may be significantly associated with gender, tumour location, tumour differentiation, MSI, MLH1 protein expression, and BRAF mutation. Thus, promoter hypermethylation plays a major role in cancers, such as CRC, through transcriptional silencing of critical genes, as was observed for MLH1 in 12% of cases (Haraldsdottir *et al.*, 2016). In fact, somatic hypermethylation of MLH1 is an accurate and cost-effective pre-screening method in the selection of patients that are candidates for MLH1 germline analysis when Lynch syndrome — responsible for MMR in 3% of CRC cases through germline mutations in MMR genes (Haraldsdottir *et al.*, 2016) — is presumed and MLH1 protein expression is absent (Gausachs *et al.*, 2012). A recent study to assess the differences in cancer-specific survival between Lynch syndrome-associated and MLH1-hypermethylation CRC, concluded they do not differ and suggested they carry a similar prognosis (Haraldsdottir *et al.*, 2016). However, Scarpa *et al.* (2016) evaluated the methylation status of some genes in the colonic mucosa without dysplasia or adenocarcinoma at different steps of sporadic and ulcerative colitis-related carcinogenesis and realised the methylation status of *MLH1*, among other tested genes, can be used as a marker of CRC. Summing up, transcriptional control of *MLH1* expression is tightly connected to the onset and development of several types of cancer, either due to germline mutations affecting its sequence and consequent protein product, or due to hypermethylation of the promoter region which inhibits transcription and subsequent translation.

1.7. Aims of the present work

Translation initiation is a major step in regulation of gene expression. As a consequence, a cell's ablility to control which proteins will be synthesized at a particular time and under a particular condition, or as a response to an external cue, is of utmost importance for its survival. In this regard, the use of non-canonical mechanisms of translation initiation play a pivotal role in allowing cells to adapt to environmental changes that deregulate the canonical cap- and scanning-dependent mechanism of protein synthesis. As already stated, this may lead to the onset and/or development of diseases, such as cancer, making it very important to understand which mechanisms cells use to control their protein synthesis in every situation.

Here, we aimed to identify proteins with potential relevance in colorectal cancer, which can be translated via non-canonical mechanisms of translation initiation. Bearing this in mind, we established several research objectives:

- *In silico* selection of putative proteins translated via non-canonical translation initiation mechanisms, based on their characteristics, the expression patterns in normal versus cancer tissues, and the corresponding mRNA levels;
- Experimental validation of the selected transcripts' 5'UTR's ability to mediate non-canonical translation initiation in colorectal cancer cells versus other cells;
- Identification of the 5'UTR core sequence that controls the non-canonical translation initiation mechanism, using deletional and mutational analyses;
- Study of the non-canonical mechanisms' ability to mediate translation initiation under conditions that impair canonical translation initiation;
- Identification of the alternative mechanism of translation initiation used by each selected protein.

Thus, we wish to understand how the mechanisms that govern translation of proteins with relevant functions in cell development and proliferation regulation work and, consequently, their role in controlling the onset and progression of diseases like colorectal cancer.

2.

Materials & Methods

2.1. *In silico* predictions

The Human Protein Atlas (http://www.proteinatlas.org/) and *Gene Expression Atlas* (http://www.ebi.ac.uk/rdf/services/atlas/) databases were used to gather information to select putative candidates translated via non-canonical mechanisms for further experimental validation. UPF1 (NM_001297549.1), AGO1 (NM_012199) and MLH1 (NM_000249) 5'UTR sequences are curated in NCBI (http://www.ncbi.nlm.nih.gov/) database as the most common variant. mFold software (http://mfold.rna.albany.edu/?q=mfold) was used to predict the secondary structure of human 5'UTR, applying the standard parameters defined by the software. Bioedit software (http://www.bioinformatics.org/sms2/color_align_cons.html) was used to align 5'UTR sequences among species. RNAalifold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi) was used to predict the degree of conservation of 5'UTR among species, according to the predicted secondary structure. GC content (%) of 5'UTR was calculated with Endmemo software (http://www.endmemo.com). Prediction of G-quadruplexes formation within 5'UTR sequences was performed using the software QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php) (Kikin *et al.*, 2006).

2.2. Plasmid constructs

The bicistronic plasmid was based on the commercially available vector psiCHECK^{TM}- 2 by Promega. It contains two reporter genes, *Renilla* Luciferase (RLuc; cap-dependent translated) and firefly Luciferase (FLuc; cap-independent translated). A stable hairpin (Candeias *et al.*, 2006) has been cloned downstream RLuc stop codon to prevent reinitiation, originating the empty vector *pR_F*, as previously described by (Marques-Ramos, 2013). The human β-globin 5'UTR (HBB, NM_000518), negative control for cap-independent translation, was PCR amplified, using primers #1 and #2. In parallel, a fragment from *pR_F* vector was amplified with primers #3 and #4. The respective fragments were subjected to SOEing (splicing by overlap extension) PCR-based method with primers #1 and #4. The resulting PCR products were digested with *XmaI/BsrGI* and cloned into *pR_F*, generating *pR_HBB_F* construct. The *c-Myc* IRES sequence, cellular positive control for cap-independent translation (minimal *c-Myc* IRES sequence described in Stoneley *et al.*[2000b]), was PCR amplified with primers #5 and #6 from *c-Myc* 5'UTR-containing pCDNA3 plasmid as template. At the same time, a fragment from *pR_F* vector was amplified using primers #7 and #8. After SOEing PCR with primers #5 and #8, both fragment and vector were digested with *EcoRI/AccI* restriction enzymes and cloned in *pR_F* vector. The resulting plasmid was, again, digested with *XmaI/BsrGI* and cloned into *pR_F*, generating *pR_MYC_F* construct. The *Encephalomyocarditis* virus (*EMCV*) IRES sequence, viral positive control for cap-independent translation (wild-type EMCV IRES sequence described in Bochkov and Palmenberg, [2006]), was PCR amplified from the EMCV sequence-containing pCDNA3 plasmid, using primers #9 and #10. In parallel, pR_F vector was amplified with primers #11 and #8. SOEing PCR was performed with the resulting PCR products using primers #9 and

#8. The generated fragment was digested with *EcoRI/AccI* and cloned into *pR_F*. To generate *pR_EMCV_F*, the previous plasmid was digested, again, with XmaI/BsrGI and cloned into *pR_F* vector. The *pR_AGO1_F* plasmid was also obtained by SOEing PCR: 5'UTR of human *AGO1* was PCR amplified using primers #12 and #13 and a fragment from *pR_F* using primers #14 and #8. After SOEing PCR with primers #12 and #8, both fragment and *pR_F* were digested with *EcoRI/AccI*. To generate *pR_AGO1_F*, the previous plasmid was digested, again, with *XhoI* and cloned into *pR_F* vector.

The same strategy was used for cloning MutL homolog 1 5'UTR, but with primers #15 to #17 and #8 and the enzymes XmaI/AccI; the resulting construct was called *pR_MLH1_F*.

Likewise for cloning Up-frameshift 1 5'UTR but with primers #24 to #26 and #8 and the enzymes *SmaI/BsrGI*; the resulting construct was called *pR_UPF1_F*.

UPF1 5'UTR 5' end deletional mutants were also obtained by SOEing PCR, as explained before: fragments from *UPF1* 5'UTR were amplified with primers #27–#30 and #31 from *pR_UPF1_F* template, digested with SmaI/BsrGI and cloned in *pR_F*, originating the constructs *pR_51-275_F*, *pR_101-275_F*, *pR_151-275_F* and *pR_201-275_F*, respectively

UPF1 5'UTR 3' end deletional mutants were also obtained by SOEing PCR as before: fragments from *UPF1* 5'UTR were amplified with primers #32–#35 and #36 from *pR_UPF1_F* template; fragments from *pR_F* were amplified with primers #37–#40 and primer #41. After SOEing PCR with primers #36 and #41, all fragments and *pR_F* were digested with *SmaI/BsrGI* and cloned in *pR_F* originating constructs *pR_1-50_F*, *pR1-100_F*, *pR_1-150* and *pR_1-200_F*.

Constructs with point mutations within *UPF1* 5'UTR were generated by site-directed mutagenesis (SMD) according to standard procedures using primers #42 and #43 (mutation at nts 39–40), primers #44 and #45 (mutation at nts 98–100), and primers #46 and #47 and #48 and #49 (mutations at nts 161–163 and 209–211, respectively). The resulting constructs were, respectively, *pR_AA_F*, *pR_ATA_F* and *pR_AAT-ATT_F*.

Construct with point mutations within *MLH1* 5'UTR were also generated by SDM, using primers #50 and #51 (mutation at nt -28), and primers #52 and #53 (mutation at nt -93). The resulting constructs were, respectively, *pR_MLH1-28_F* and *pR_MLH1-93_F*.

To generate the promoterless constructs, and remove the SV40 promoter and the chimeric intron, *pR_F* was digested with *NheI/BglII*, blunt-ended with Quick Blunting Kit (New England Biolabs) and re-ligated, originating the promoterless *p-R_F* plasmid. *pR_AGO1_F* plasmid was digested with EcoRV/BsrGI and the resulting fragments were cloned into p-R_F, originating the promoterless *p-R_AGO1_F*. *pR_MLH1_F* and *pR_UPF1_F* plasmids were digested with XmaI/BsrGI and the resulting fragments were cloned into *p-R_F*, originating the promoterless *p-R_MLH1_F* and *p-R_UPF1_F*, respectively.

Monocistronic reporter constructs used were obtained by removing RLuc ORF sequence from *pR_F* by SOEing PCR using primers #18 to #23. PCR product and *pR_F* were then digested with *NheI/BsrGI* and the resulting vector and insert were ligated; the resulting construct was called *p_F*. The 5'UTR of *HBB*, *MLH1*, *UPF1* and *AGO1*, as well as the IRES sequences of

c-Myc and *EMCV*, were cloned using the same restriction enzymes used for the bicistronic constructs explained above, generating the constructs *p_HBB*_F, *p_MLH1*_F, *p_UPF1*_F, *p_AGO1*_F, *p_MYC_F* and *p_EMCV*_F, respectively.

All restriction enzymes used in this work were from New England Biolabs, except *XmaI* (NZYTech) and *ECORI* (Amersham) and T4 DNA ligase was from Thermo Fisher Scientific. Digestions and ligations were performed according to manufacturer's instructions. Clones were generated in NZY5α competent cells (NZYTech) and plasmid DNAs were extracted with NZYMiniprep kit (NZYTech)

All primer sequences are provided in table 2.1.

Table 2.1 — Sequences of the primers used to generate the constructs needed for this study. All presented sequences are oriented from 5' to 3'.

Primer	Sequence
#1	TCCCCCCGGGGGGAACATTTGCTTCTGACACAAC
#2	CATCGGCCATGGTGTCTGTTTGAGGT
#3	ACAGACACCATGGCCGATGCTAAGAACA
#4	GTGAGAGAAGCGCACACAG
#5	GGAATTCCAATTCCAGCGAGAGGCAGAG
#6	TAGCATCGGCCATCGTCTAAGCAGCTGCAAGGAGA
#7	GCAGCTGCTTAGACGATGGCCGATGCTAAGAACA
#8	GCAAATCAGGTAGCCCAGG
#9	GGAATTCCTTCCGCCCCTCTCCCTCCCC
#10	TAGCATCGGCCATACAATGGGGTACCTTCTGG
#11	ACCCCATTGTATGGCCGATGCTAAGAACA
#12	GGAATTCCACTGGCAGCTGGCCGGGCGCTCGCAGTGGGAG
#13	TAGCATCGGCCATCCCATATACCCGTGCGGAGGTCA
#14	ACGGGTATATGGGATGGCCGATGCTAAGAACA
#15	TCCCCCGGGGGAGAAGAGACCCAGCAACCCAC
#16	TAGCATCGGCCATTTTGGCGCCAGAAGAGC
#17	GGCGCCAAAATGGCCGATGCTAAGAACA
#18	GTCTCGAACTTAAGCTGCAG

#19	CGATCGCCTAGAAGGTGGCTAGCCTATAGTGAGT
#20	GGCTAGCCACCTTCTAGGCGATCGCTCGAGCT
#21	GTGAGAGAAGCGCACACAG
#22	GTCTCGAACTTAAGCTGCAG
#23	GTGAGAGAAGCGCACACAG
#24	GGAATTCCCACGGCGACGGCGGCGGTGGCGGCAGTTCCT
#25	CATCGGCCATGGTGCCTCCGGGTAGGGCCCTCGGGCCGGT
#26	CCTACCCGGAGGCACCATGGCCGATGCTAAGAACA
#27	TCCCCCGGGGGATGGCGGCTTCGAGGGGAGCT
#28	TCCCCCGGGGGAGCGGCTCGGCACTGTTACCT
#29	TCCCCCGGGGGATTGGTCCTTTCCGGGCGCG
#30	TCCCCCGGGGGAGCGGCCTAGGCCTCAGCGCG
#31	GTGAGAGAAGCGCACACAG
#32	TAGCATCGGCCATGCCGCTCGCAGCCTAGAGCA
#33	TAGCATCGGCCATCGCCGCT GCCGCCGAGCCCCTCC
#34	TAGCATCGGCCATACCGCCCGCCCCGGCGCCAG
#35	TAGCATCGGCCATAGGCCTCGGGTCGCTGCCGC
#36	GGCTTGTCTGGCCTTTCACTA
#37	GCTGCGAGCGGCATGGCCGATGCTAAGAACAT
#38	TCGGCGGCAGCGGCGATGGCCGATGCTAAGAACAT
#39	GCCGGGGCGGGCGGTATGGCCGATGCTAAGAACAT
#40	GACCCGAGGCCTATGGCCGATGCTAAGAACAT
#41	GTGAGAGAAGCGCACACAG
#42	TGGCGGCAGTTCCTGCTCTAGAATGCGAGCGGCTGGCGGCTTCGA
#43	TCGAAGCCGCCAGCCGCTCGCATTCTAGAGCAGGAACTGCCGCCA
#44	AGGGGCTCGGCGGCAGCGATAGCGGCTCGGCACTGTTA
#45	TAACAGTGCCGAGCCGCTATCGC TGCCGCCGAGCCCCT
#46	TGGCGCCGGGGCGGGCGGTTTGGTCCTTTAATGGCGCGCGGGGGCGACAGCG CAGCGA

#47	TCGCTGCCGCTGTCGCCCCCGCGCGCCATTAAAGGACCAAACCGCCCGCCCCG GCGCCA
#48	AGCGACCCGAGGCCTGCGGCCTAATTCTCAGCGCGGCGGCGGGCTCGA
#49	TCGAGCCCGCCGCCGCGCTGAGAATTAGGCCGCAGGCCTCGGGTCGCT
#50	CACTTCCGTTGAGCATCTAGACGTTTCCTTGGCTCTTCTGG
#51	CCAGAAGAGCCAAGGAAACGTCTAGATGCTCAACGGAAGTG
#52	GGATGGCGTAAGCTACAGCTAAAGGAAGAACGTGAGCACGA
#53	TCGTGCTCACGTTCTTCCTTTAGCTGTAGCTTACGCCATCC

2.3. *In vitro* transcription

Each bicistronic plasmid (described above) was linearized using *ClaI* restriction enzyme (New England Biolabs). Linearized fragments were purified with DNA Clean & Concentrator™-5 (Zymo Research) and 1 µg of this purified product was *in vitro* transcribed and capped with mMessage mMachine T7 kit (Ambion), according to manufacturer's instructions. RNA samples were treated with Turbo DNase (Ambion) and, then, poly-adenylated with poly(A) tailing kit (Ambion), according to manufacturer's instructions. The resulting transcripts were purified by phenol:chloroform (pH=4.7, Ambion) extraction, precipitated with absolute ethanol and eluted in RNAe-free water, according to standard procedures. The quality of all obtained transcripts was analysed by denaturing formaldehyde-agarose gel electrophoresis.

Monocistronic reporter plasmids (p_F, p_*HBB*_F, p_*MYC*_F, p_*EMCV*_F, p_*AGO1*_F, p_*UPF1*_F and p_*MLH1*_F) were linearized with ClaI (New England Biolabs). Linearized fragments were purified with DNA Clean & Concentrator™-5 (Zymo Research) and 1µg of each purified product was *in vitro* transcribed with HiScribe™T7 Quick High Yield RNA Synthesis Kit (New England Biolabs). During *in vitro* transcription, each transcript was capped with either m^7G(5')ppp(5')G RNA Cap Structure Analog or G(5')ppp(5')A RNA Cap Structure Analog (New England Biolabs), according to manufacturer's instructions. The resulting transcripts were then poly-adenylated with *E. coli* Poly(A) Polymerase (New Englans Biolabs), according to manufacturer's instructions. The resulting transcripts were extracted with phenol:chloroform (pH=4.7, Ambion), precipitated with absolute ethanol and eluted in RNAe-free water, according to standard procedures. The quality of all obtained transcripts was analysed by denaturing formaldehyde-agarose gel electrophoresis.

2.4. Cell culture

HeLa[20] and HCT116[21] cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) foetal bovine serum (FBS), whereas NCM460[22] cells were cultured in Roswell Park Memorial Institute–1640 (RPMI) medium supplemented with 10% (v/v) FBS. Cells were maintained at 37 ºC in a humidified atmosphere with 5% CO_2.

Cells were seeded in 35-mm plates 24h prior to transfection in a manner such as cell confluency would be ~30–40% or ~70–80% at the time of transfection with either siRNA or plasmid DNA/*in vitro* transcribed mRNA, respectively.

2.5. Transfections with plasmid DNA or *in vitro* transcribed mRNA

Cells were transfected with either 1.5 μg of plasmid DNA or 4 μg of *in vitro* transcribed mRNA using Lipofectamine 2000 Transfection Reagent (Invitrogen), according to manufacturer's instructions. Briefly, per 35 mm well, DNA or mRNA was diluted in 250 μl of Optimem medium, and 4 μl of Lipofectamine were diluted in 250 μl of Optimem and rest for 5 min. The latter solution was added to the former and rest for 20 min. Meanwhile, old culture medium was removed and fresh medium was added to the cell culture dishes. After 20 min, cells were transfected dropwise and incubated at 37 ºC for either 20–24h (plasmid DNA) or 4–8 h (*in vitro* transcribed mRNA). When mentioned, cells were co-transfected with 500 ng β-galactosidase-encoding plasmid (pSV-β-Galactosidase control vector by Promega).

2.6. siRNA transfection

The siRNA oligonucleotides used for knocking down eIF4E (5'-AAGCAAACCUGCGG CUGAUCU-3'), GFP (5'-GGCUACGUCCAGGAGCGCAC-3'), RLuc (5'-GCUGCAAGC AAAUGAACGU-3'), and FLuc (5'-GGACGAGGACGAGCACUUC-3') were designed with 3'-dTdT overhangs and purchased as annealed, ready-to-use duplexes (Thermo Fisher Scientific). Cells were transfected with 200 pmol of each siRNA, according to manufacturer's instructions (cf. section 2.5). Twenty four hours post siRNA transfection, cells were transfected with the plasmids of interest, and harvested 48 h post siRNA transfection. For experiments requiring mRNA transfection, cells were transfected with *in vitro* transcribed mRNA approximately 40 h post siRNA transfection and harvested 4–8 h later.

2.7. Drug treatments

Four to six hours post DNA transfection, cells were treated with: 200 μM of cobalt chloride ($CoCl_2$) (Sigma-Aldrich) for 20 h; 1 μM of thapsigargin (Sigma-Aldrich) for 20 h; or 200 μM of 4EGI-1 (Calbiochem) for 20 h. Fourteen hours post DNA transfection cells were treated with 80

[20]Human cervical cancer-derived cell line.
[21]Human pre-metastatic colorectal carcinoma-derived cell line.
[22]Human normal intestinal mucosa-derived cell line.

nM of rapamycin (Sigma-Aldrich) for 6 h. In parallel, cells were treated with the corresponding control vehicle: H_2O for $CoCl_2$, or DMSO for the other drugs. All cells were harvested 20–24h after transfection.

Two hours post RNA transfection, cells were treated with either 200 μM of cobalt chloride ($CoCl_2$) (Sigma-Aldrich), 1 μM of thapsigargin (Sigma-Aldrich), or with 80 nM of rapamycin (Sigma-Aldrich), for 6 h. In parallel, cells were treated with the corresponding control vehicle: H_2O for $CoCl_2$, or DMSO for the other stimuli. All cells were harvested 8–10 h posttransfection.

2.8. Luminometry assays

Twenty four hours (DNA transfection) or 8 h (RNA transfection) posttransfection, cells were rinsed with pre-chilled 1x (v/v) phosphate-buffered saline (PBS) and lysed with 100 μl of 1x (v/v) passive lysis buffer (Promega). Cleared cell lysate (10 μL) were used to perform the luminometry assays. The Dual Glo Assay System (Promega) was used to assess both RLuc and FLuc relative luciferase activity, and the Beta Glo Assay System was used to assess β-galactosidase activity, both according to manufacturer's instructions. Briefly, 40 μl of Luciferase Assay Reagent (LAR, contains the substrate for firefly luciferase) was added to the sample and luminescence was read in a GloMax® 96 Microplate Luminometer (Promega); then, 40 μl of Stop & Glo Reagent (stops reaction between LAR and firefly luciferase, and contains the substate for *Renilla* luciferase) was added to the sample and luminescence was read in the same Luminometer. The resuts were obtained in arbitrary light units.

2.9. SDS-PAGE and Western blot

SDS sample buffer (5x volume/volume, v/v) was added to 20 μl of clear whole cell lysate. Samples were then denatured at 95 ºC for 10 min, resolved by polyacrylamide gel electrophoresis (SDS-PAGE), for 1 h in a 10% acrylamide gel and transferred, for 1 h, to PVDF membranes (Bio-Rad), previously activated with methanol. Membranes were then blocked with either 5% (weight/volume, w/v) non-fat dry milk or bovine serum albumin (BSA, Sigma), 1x (v/v) tris-buffered saline (TBS), 0.05% (v/v) Triton x-100 (Sigma-Aldrich) or 0.1% (v/v) Tween 20 (Sigma-Aldrich), as specified below, and analysed by immunoblotting, according to standard procedures. Blots were probed with the following antibodies: rabbit anti-HIF1-α (Sigma-Aldrich) diluted 1:1000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), at room temperature (rt), for 1 h, to control the effect of $CoCl_2$; rabbit anti-PARP (Cell Signaling) diluted 1:1000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.1% (v/v) Tween 20 (Sigma-Aldrich), overnight (o/n) at 4 ºC, to control the effect of 4EGI-1; rabbit anti-eIF2α-Phosphorylated (Invitrogen) diluted 1:750 in 5% (w/v) BSA, 1x (v/v) TBS, 0.1% (v/v) Tween 20 (Sigma-Aldrich), o/n at 4 ºC, to control the effect of thapsigargin; rabbit anti-p70-S6K-Phosphorylated (Cell Signaling) diluted 1:1000 in 5% (w/v) BSA, 1x (v/v) TBS, 0.1% (v/v) Tween 20 (Sigma-Aldrich), o/n at 4 ºC, to control the effect of rapamycin; rabbit anti-eIF4E subunit

(Ambion), diluted 1:1000 in 5% (w/v) non-fat dry milk, 1x v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), o/n at 4 ºC, to control knock-down of eIF4E subunit; and mouse anti-α-tubulin (Sigma-Aldrich) diluted 1:50 000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), to control sample loading. Detection was performed using secondary peroxidase-conjugated anti-mouse IgG (Bio-Rad) or anti-rabbit IgG (Bio-Rad) antibodies diluted 1:4000 or 1:3000, respectively, in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), at rt, for 1 h, with gentle shaking, followed by Enhanced Chemiluminescence.

In cases we wished to observe proteins with similar molecular weights, we stripped off the membrane from previously used antibodies and probed blots with different antibodies, according to standard procedures. Briefly, dried membranes were re-activated with methanol, blocked in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich) or 0.1% (v/v) Tween 20 (sigma-Aldrich), and blots were probed with: rabbit anti-eIF2α (Cell Signaling), diluted 1:750 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.1% (v/v) Tween 20 (sigma-Aldrich), at rt, for 1 h; or rabbit anti-p70-S6K (Cell Signaling) diluted 1:1000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), at rt for 1 h. Detection was performed using secondary peroxidase-conjugated anti-rabbit IgG (Bio-Rad) antibodies diluted 1:3000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), at rt, for 1 h, with gentle shaking, followed by Enhanced Chemiluminescence.

2.10. Co-Immunoprecipitation

Cells were rinsed with pre-chilled dialysed culture medium, and lysed with 150 µl of pull-down buffer (NP40 buffer with protease inhibitor diluted 1:100). Lysates were centrifuged for 10 min, at 4 ºC, 5000 rpm in a refrigerated tabletop microcentrifuge (Eppendorf 5415R). Clear supernatant (20 µl) was transferred to a tube containing 5 µl 5x (v/v) SDS sample buffer, and denatured for 10 min at 95 ºC (pre-IP lysate). The remaining supernatant was transferred to a clean tube, 1 µl of rabbit anti-eIF4E antibody was added to the tube and it was incubated o/n, at 4 ºC, on a spinning rotator. Protein G agarose beads (30 µl; Roche) were added to each tube and they were incubated o/n, at 4 ºC on a spinning rotator. Tubes were centrifuged for 30 s at 4 ºC, as before. Supernatant (20 µl) was transferred to a clean tube containing 5 µl of 5x (v/v) SDS sample buffer, and denatured for 10 min at 95 ºC (post-IP lysate). The remaining supernatant was discarded and, after washed with pull-down buffer, beads were denatured for 10 min at 95 ºC with 30 µl of 2x (v/v) SDS sample buffer, resulting in a bead-free lysate (IP lysate). All obtained lysates (pre-IP, IP and post-IP) were then analysed by Western blot analysis. Blots were probed with the following antibodies: rabbit anti-eIF4E subunit (Ambion), diluted 1:1000 in 5% (w/v) non-fat dry milk, 1x (v/v)TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), o/n at 4 ºC; rabbit anti-eIF4G subnit (Cell Signaling), diluted 1:1000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), o/n at 4 ºC; and mouse anti-α-

tubulin (Sigma-Aldrich) diluted 1:50 000 in 5% (w/v) non-fat dry milk, 1x (v/v) TBS, 0.05% (v/v) Triton x-100 (Sigma-Aldrich), to control sample loading. Detection was performed as before.

2.11. Quantification of total protein amount using Bradford's reagent

Standard calibration curve was obtained using Bovine Serum Albumin (BSA) as a reference protein. Sequential dilutions with a known reference concentration (0.2 mg/ml, 0.4 mg/ml, 0.6 mg/ml, 0.8 mg/ml, 1.0 mg/ml, 1.2 mg/ml, 1.4 mg/ml) were used as standard solutions. 10 µl of each solution were thoroughly mixed with 200 µl of NZYBradford reagent (NZYTech). After 2 min, absorbance of each standard solution was measured at 595 nm in a spectrophotometer (NanoVue Plus spectrophotometer, GE Healthcare Life Sciences). Samples were diluted 1:5 in ddH_2O and 10 µl of diluted sample were thoroughly mixed with 200 µl NZYBradford reagent (NZYTech). After 2 min, absorbance of each sample was measured at 595 nm in the same spectrophotometer and compared to the calibration curve previously created, thus obtaining the actual protein amount in each sample.

2.12. RNA isolation

Total RNA from transfected cells was isolated using Nucleospin II RNA extraction kit (Macherey-Nagel), following the manufacturer's instructions. All samples were treated with RNAe-free DNase I (Ambion). RNA was then extracted with equal amount of phenol:chloroform, pH=4.7 (Ambion), precipitated with absolute ethanol and eluted in RNAe-free water, according to standard procedures.

2.13. RT-PCR analysis

cDNA was synthesized with NZY Reverse Transcriptase (NZYTech) according to manufacturer's instructions, using random hexameres (Invitrogen) and 1 µg of total RNA. PCR was performed according to standard procedures, using the resulting cDNA as template. Two sets of primers spanning the whole transcript were used to check the integrity of the latter. Set I: 5'-GTCTCGAACTTAAGCTGCAG-3' (fwd); and 5'-GCAAATCAGGTAGCCCAGG-3' (rev). Set II: 5'-ATGGCTTCCAAGGTGTACGA-3' (fwd); and 5'-ATCGATTTTACCACATTTGTAGAGG-3' (rev).

2.14. Data analysis and statistics

Regarding bicistronic reporter constructs, RLuc is the internal control for transfection efficiency and therefore FLuc activity was normalised to RLuc activity from the same construct. Then, FLuc/RLuc relative luciferase activity of each construct was normalised to that from the empty counterpart to calculate variations in FLuc expression. Under stress conditions, all relative luciferase activities obtained were normalised to those from the empty construct under control conditions. In situations in which FLuc and RLuc activities had to be addressed

separately, either FLuc or RLuc activity was normalised to β-galactosidase activity (derived from the co-transfected β-galactosidase-containing plasmid, used as a control of transfection efficiency). FLuc/β-galactosidase or RLuc/β-galactosidase relative luciferase activities were then normalised to those from the empty counterpart to determine variations in FLuc or RLuc expression. As far as monocistronic reporter constructs are concerned, FLuc activity was normalised to that of β-galactosidase. Then, FLuc/β-galactosidase relative luciferase activity of each construct was normalised to its modified counterpart (e.g.: hairpin-containing construct vs. construct without hairpin, or A-capped transcript vs. G-capped transcript), and, eventually, the relative activity obtained was normalised to that of the empty constructs.

All results are expressed as mean ± standard deviation. Two-tailed Student's *t*-test was used for estimation of statistical significance. Significance for statistical analysis was defined as $p<0.05$, considering different variances among samples. All presented data result at least from three independent experiments.

3.

Results

I. *In silico* selection of candidates

The main goal of this work was to identify proteins whose synthesis could occur via an alternative mechanism of translation initiation. In order to find such proteins, we have searched the available bibliographic resources for evidence that might lead us to them. These would have to include proteins whose expression would be maintained under conditions compromising the canonical mechanism of translation initiation in human cells, either by impairing cap recognition and binding of eIF4F complex — eliciting cap-independent mechanisms of translation initiation, such as IRES-mediated translation initiation — or by impairing the scanning of 5'UTR — triggering mechanisms that allow the ribosome to bypass the secondary structures, such as the ribosome shunting mechanism. Also, proteins with altered expression in some conditions, such as cancer, would be of relevance for choosing candidates. In this regard, we sought proteins whose expression would be altered in several cancer types. A tight scrutiny of the literature showed many proteins overexpressed in these conditions, such as: transient receptor potential vanilloid 3 (TRPV3) (Li *et al.*, 2016); Kirsten rat sarcoma viral oncogene homolog (KRAS) (Kempf *et al.*, 2016); high mobility group A protein 2 (HMGA2) (Palumbo *et al.*, 2016); moesin and cytokeratin 17 (KRT17) (Luo *et al.*, 2004; Shin, 2011); casein Kinase 2 subunit α (CK2α) (Zou *et al.*, 2011); the Argonaute protein family members (Li *et al.*, 2010a); ribosomal protein genes (Pogue-Geile *et al.*, 1991); tyrosine kinases (Leroy *et al.*, 2009); small cell adhesion glycoprotein (SMAGP) (Tarbé *et al.*, 2005); or centromere protein-A (Tomonaga *et al.*, 2003), among many others. Furthermore, the expression of several translation-related proteins, such as eukaryotic initiation factors, has been proven to be altered in various types of cancer — e.g. eIF2B (Gallagher *et al.*, 2008); eIF3a (Shen *et al.*, 2014); eIF3f (Cheng *et al.*, 2014; Li *et al.*, 2014); eIF3i (Qi *et al.*, 2014); eIF3m (Goh *et al.*, 2011); eIF4E (Wang *et al.*, 1999; Shuda *et al.*, 2000; Wang, 2012; Yin *et al.*, 2014); eIF4G (Connolly *et al.*, 2011); or eIF5A2 (Guan *et al.*, 2001; Liu *et al.*, 2014), etc. Similarly, some translation-related proteins like up-frameshift 1 (UPF1) or mammalian target of rapamycin (mTOR) can also maintain their expression levels during cell cycle S phase progression and G2/M, respectively (Azzalin and Lingner, 2006b; Liu *et al.*, 2007). Thus, our preliminary selection featured the aforementioned proteins — UPF1, KRAS, Argonaute family proteins, CK2α, and KRT17. We search them in *The Human Protein Atlas* (http://www.proteinatlas.org) to verify their expression in several tissues, normal and cancerous ones, at protein and RNA levels. We have also checked the information curated in the Gene Expression Atlas (http://www.ebi.ac.uk/gxa/home), in order to understand the differential expression of the corresponding mRNA in different conditions, such as disease, infection, or external stimuli. After such a thorough analysis, we have curtailed our assortment of candidates to UPF1, KRAS and argonaute family proteins, and, finally, managed to confine our research to UPF1 and AGO1 as they seem suitable candidates, as further detailed below.

I.1. *In silico* analysis of UPF1 expression

We have gathered information on UPF1 protein and mRNA expression patterns, from *The Human Protein Atlas*. Figure 3.1 shows an overview of RNA and protein expression. Expression of UPF1 is higher in brain, lung, male and female tissues, whereas adipose and soft tissues present the lowest expression levels. Interestingly, the RNA expression levels are seldom concomitant with the counterpart protein's, as the former are, on average, expressed in lower levels.

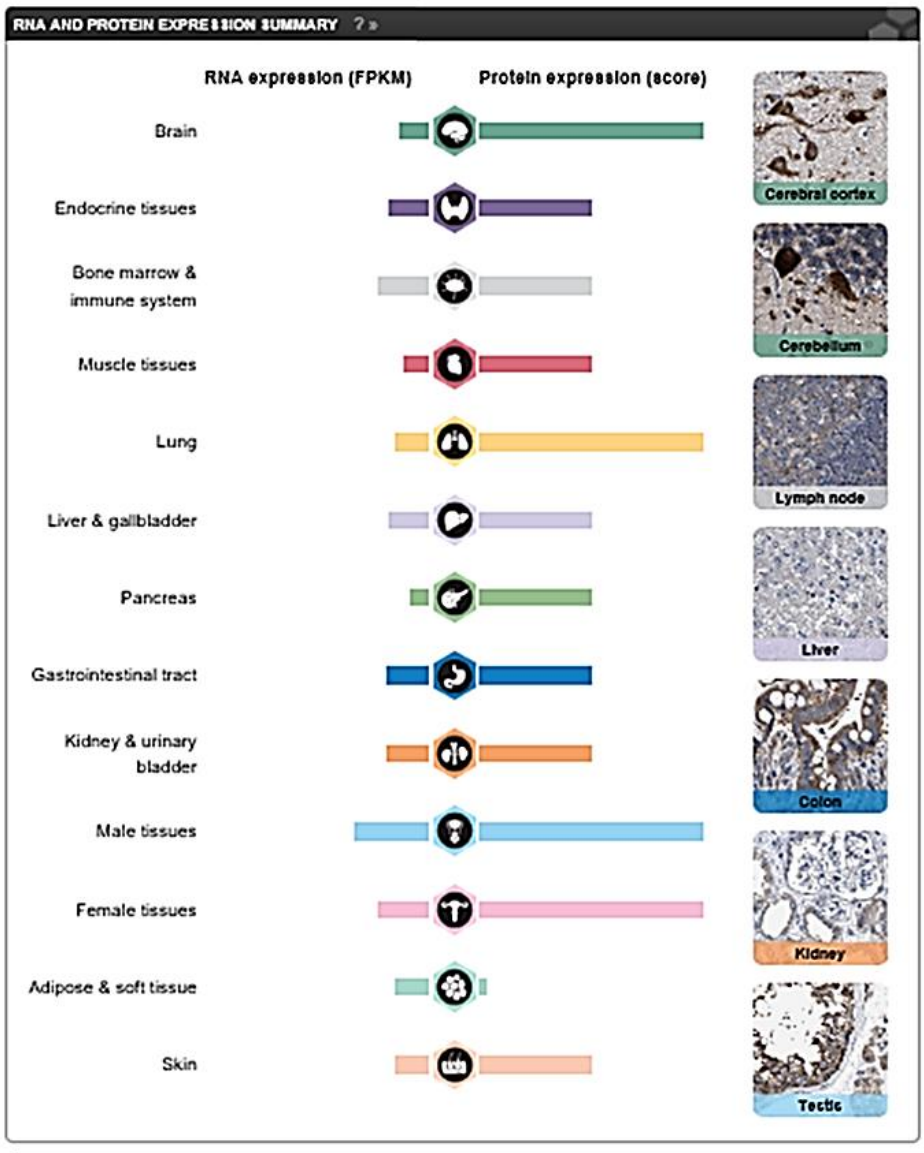


Figure 3.1 — Overview of UPF1 protein and RNA expression data. Analysed tissues are divided into 13 colour-coded groups, according to common functional features. Images of selected tissues give a visual summary of the protein expression profile (panels on the right). RNA expression results were obtained from RNA-seq analysis and are reported as number of fragments per kilobase gene model and million reads (FPKM). "Protein expression (score)" represents the highest expression score found in a particular group of tissues, corresponding to the staining profile of the used antibody. Data from http://www.proteinatlas.org.

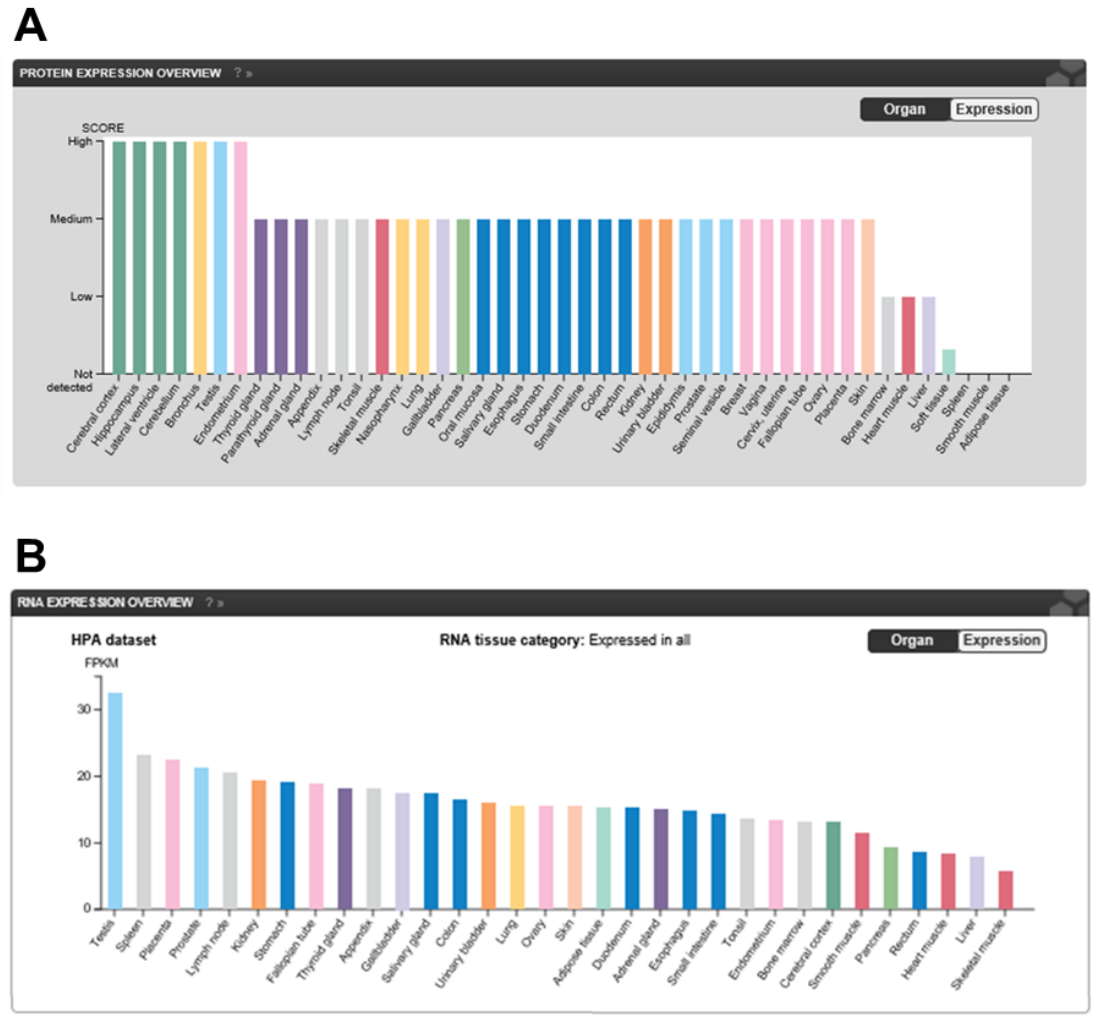


Figure 3.2 — Expression pattern of UPF1 in different tissues. (A) UPF1 protein expression levels in different tissues, according to antibody immunostaining intensity (not detected, low, medium or high). The generated tissue microarrays include samples from 144 individuals, corresponding to 44 different normal tissue types. (B) *UPF1* mRNA expression levels were obtained as RNA-seq data, from 32 tissues. These are referred to as mean FPKM (fragments per kilobase gene model and million reads), corresponding to mean values of the different individual samples from each tissue type. FPKM thresholds are set for categorisation of transcript expression levels into “not detected” (0–0.5); “low” (0.5–10); “medium” (10–50); or “high” (>50). Data from http://www.proteinatlas.org.

Figure 3.2 shows protein and RNA expression in further detail. According to the data, UPF1 protein is highly expressed in annotated tissues from brain, bronchus, testis and endometrium, whereas its expression in bone marrow, heart muscle, liver and soft tissues is low (figure 3.2.A). It has not been detected in spleen, smooth muscle and adipose tissue; in all other annotated tissues, UPF1 expression levels are medium (figure 3.2.A). As far as the RNA expression levels are concerned, no tissue presents a high expression level, i.e. >50 FPKM (fragments per kilobase gene model and million reads) (figure 3.2.B). Indeed, some of the analysed tissues display no increased levels of RNA expression; nevertheless, the corresponding protein levels are higher, as is the case of rectum, skeletal muscle, or cerebral cortex tissues, for instance. Such information

leads us to the conclusion that some post-transcriptional events may take place in order to translate UPF1 protein, despite the low levels of corresponding mRNA.

Regarding cancer tissues, most malignancies display weak-to-moderate cytoplasmic staining, whereas thyroid, several colorectal and endometrial cancers are strongly stained (figure 3.3).

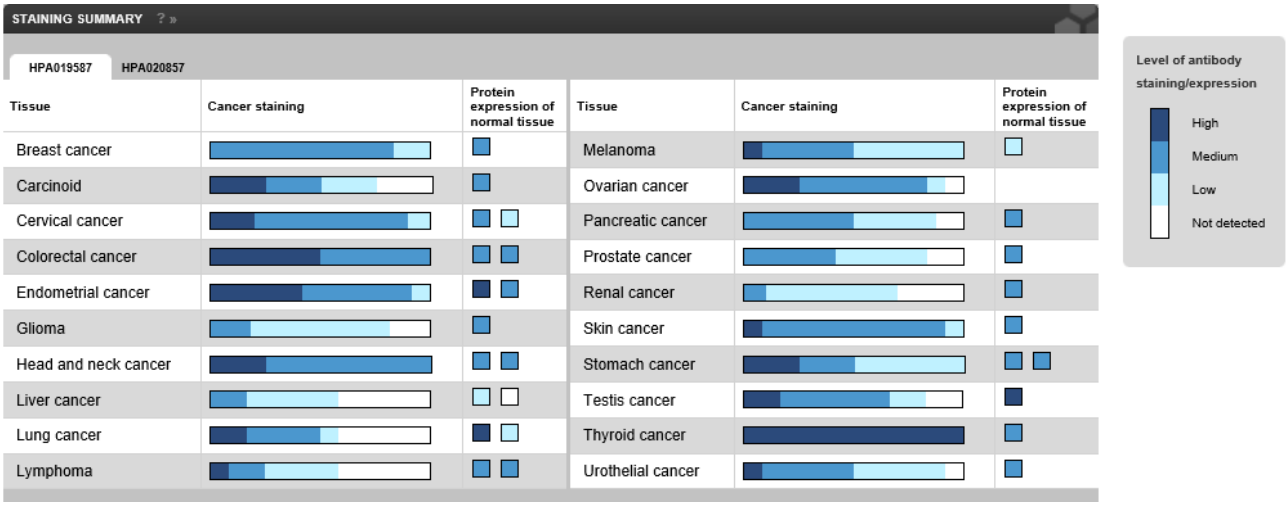


Figure 3.3 — Expression of UPF1 in different cancer tissues. For each cancer, the fraction of samples with antibody staining/protein expression levels — high, medium, low, or not-detected — are provided by the blue scale colour code, as described in the box on the right. The bar length represents the number of patient samples analysed (max=12 patients). Next to the cancer staining data, the protein expression data of normal tissues corresponding to each cancer are shown and protein expression levels are indicated by the blue scale colour code. Data from http://www.proteinatlas.org.

However, protein expression of normal colorectal and thyroid tissues is only medium, suggesting this protein is up-regulated in these cancer tissues (figure 3.2). Likewise, cervical cancer tissue staining is medium-to-high, whereas normal tissue staining is only low-to-medium, suggesting, again, an up-regulation of UPF1 protein in this cancer type. On the other hand, renal and liver cancers show the lowest expression levels of UPF1 protein and the counterpart normal tissues behave accordingly. In all other annotated cancer tissues, UPF1 expression is weak-to-moderate and the counterpart normal tissues follow the same tendency (figure 3.3 *versus* figure 3.2). In order to better understand how *UPF1* gene expression varies and whether such variation correlates with protein expression alterations, we sought information regarding mRNA expression levels in the *Gene Expression Atlas*, as depicted in figure 3.4. According to such data, *UPF1* mRNA is down-regulated in several conditions, especially upon infection by some *Staphylococcus* strains, but also, to a lesser extent, upon herpes virus infection, in osteosarcoma, or in cells treated with tumour necrosis factor (TNF) or interleukin 4 (IL4). In a different way, expression of this gene is up-regulated in some conditions, such as breast carcinoma or ovarian cancer, as well as upon treatment with trovafloxacin (4^{th}-generation antibiotics) or epoxomicin (a proteosome inhibitor with anti-inflammatory activity).

Gene	Organism	Comparison	Log_2-fold change
UPF1	Homo sapiens	'Staphylococcus aureus strain 11490' at '6 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 10254' at '6 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain Col' at '6 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 9897' at '6 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 252' at '6 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'100 microgram per milliliter trovafloxacin' vs 'none'	
UPF1	Homo sapiens	'p53KO; BREFELDIN A' vs 'p53KO; control'	
UPF1	Homo sapiens	'breast carcinoma' vs 'normal'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 9897' at '3 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'10 unit per milliliter; JFHA' vs 'control'	
UPF1	Homo sapiens	'casiopeina II-gly' vs 'control'	
UPF1	Homo sapiens	'Staphylococcus aureus strain Col' at '3 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'herpes simplex virus G207' vs 'none'	
UPF1	Homo sapiens	'HOP-62; adenocarcinoma of lung; P04' vs 'HOP-62; adenocarcinoma of lung; P0'	
UPF1	Homo sapiens	'wild type; BREFELDIN A' vs 'wild type; control'	
UPF1	Homo sapiens	'HOP-62; adenocarcinoma of lung; P10' vs 'HOP-62; adenocarcinoma of lung; P0'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 10254' at '3 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 9897' at '2 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 11490' at '3 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'NCI-H460; large cell lung carcinoma; P01' vs 'NCI-H460; large cell lung carcinoma; P0'	
UPF1	Homo sapiens	'ovarian cancer' vs 'normal' in 'stromal cell'	
UPF1	Homo sapiens	'p53KO; tunicamycin' vs 'p53KO; control'	
UPF1	Homo sapiens	'siERK2; estradiol' at '4 hour' vs 'siERK2; ethanol' at '4 hour'	
UPF1	Homo sapiens	'AsPC-1; pancreatic adenocarcinoma; P04' vs 'AsPC-1; pancreatic adenocarcinoma; P0'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 10254' at '2 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 11490' at '2 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'blood alcohol content 0.02%declining' vs 'control' in 'time point 5'	
UPF1	Homo sapiens	'SR; poorly differentiated B cell origin leukemia; P04' vs 'SR; poorly differentiated B cell origin leukemia; P0'	
UPF1	Homo sapiens	'Staphylococcus aureus strain 252' at '3 hour' vs 'none' at '0 hour'	
UPF1	Homo sapiens	'NCI-H460; large cell lung carcinoma; P10' vs 'NCI-H460; large cell lung carcinoma; P0'	
UPF1	Homo sapiens	'pEGZ-transfected (empty vector)' vs 'wild type' in 'influenza A virus infected'	
UPF1	Homo sapiens	'freshly isolated' vs 'monolayer'	
UPF1	Homo sapiens	'SR; poorly differentiated B cell origin leukemia; P10' vs 'SR; poorly differentiated B cell origin leukemia; P0'	
UPF1	Homo sapiens	'TNF-treated' vs 'untreated'	
UPF1	Homo sapiens	'IL4-treated' vs 'untreated'	
UPF1	Homo sapiens	'HOP-92; large cell undifferentiated lung; P10' vs 'HOP-92; large cell undifferentiated lung; P0'	
UPF1	Homo sapiens	'osteosarcoma' vs 'normal'	
UPF1	Homo sapiens	'K-562; chronic myelogenous (blast crisis) leukemia; P01' vs 'K-562; chronic myelogenous (blast crisis) leukemia; P0'	
UPF1	Homo sapiens	'epoxomicin; 1 micromolar' at '6 hour' vs 'none' at '0 hour'	

Figure 3.4 — Expression of *UPF1* mRNA under different conditions or experimental factors. Stronger colour saturation means higher absolute log_2-fold change value. Blue indicates the gene is down-regulated and red means it is up-regulated. Log_2-fold changes are not directly comparable across experiments. Data from http://www.ebi.ac.uk.

These data, together with data from *The Human Protein Atlas*, informed us that the levels of UPF1 protein are not a direct consequence of alterations at transcriptional level, suggesting some post-transcriptional mechanism could be involved in the regulation of its expression and, hence, the existence of a non-canonical mechanism governing its translation is likely to occur.

I.2. *In silico* analysis of AGO1 expression

Regarding AGO1 protein expression levels, *The Human Protein Atlas* provides no information. However, studies in the literature state that this protein is overexpressed in colorectal cancer and is, for that reason, a potential biomarker (Li *et al.*, 2010b). As for the mRNA levels, both *The Human Protein Atlas* and *Gene Expression Atlas* provide information on this gene. Figure 3.5 provides an overview of RNA expression in different tissue groups — *AGO1* mRNA expression levels are similar in most organs but pancreas, where its expression is very low. A deeper look into *AGO1* mRNA expression throughout different tissues shows it is expressed in low levels in all annotated tissues, as indicated by the FPKM values below 10 (figure 3.6). This agrees with data publish in literature that state *eIF2C1* — the gene coding for AGO1 protein — is expressed at low-to-medium levels (Koesters *et al.*, 1999). These data are also strengthened by the information curated in the *Gene Expression Atlas* (figure 3.7). *AGO1* expression is down-regulated in most tested conditions, including viral or bacterial infection, or upon treatment with alcohol or lipopolysaccharides. On the other hand, in renal adenocarcinoma or in squamous cell carcinoma, *AGO1* mRNA is down-regulated compared to the counterpart normal cells.

Since this gene is constitutively expressed at low-to-medium levels, but its expression at protein level is not always concomitant, as is the case of its expression in colorectal cancer cells, in which it has been identified as a potential biomarker (Li *et al.*, 2010b), we predict that its expression may be regulated by a non-canonical mechanism of translation initiation that allow protein expression from only small amounts of the corresponding mRNA.

Taking all the above information into account, we pursued the experimental validation of these two candidates (UPF1 and AGO1) and sought for evidence of their being translated via an alternative mechanism of translation initiation. Based on the results obtained *in silico* on these proteins' expression being altered in colorectal and cervical cancers (CRC and CC, respectively), we selected CRC and CC cell lines as experimental models for such validation.

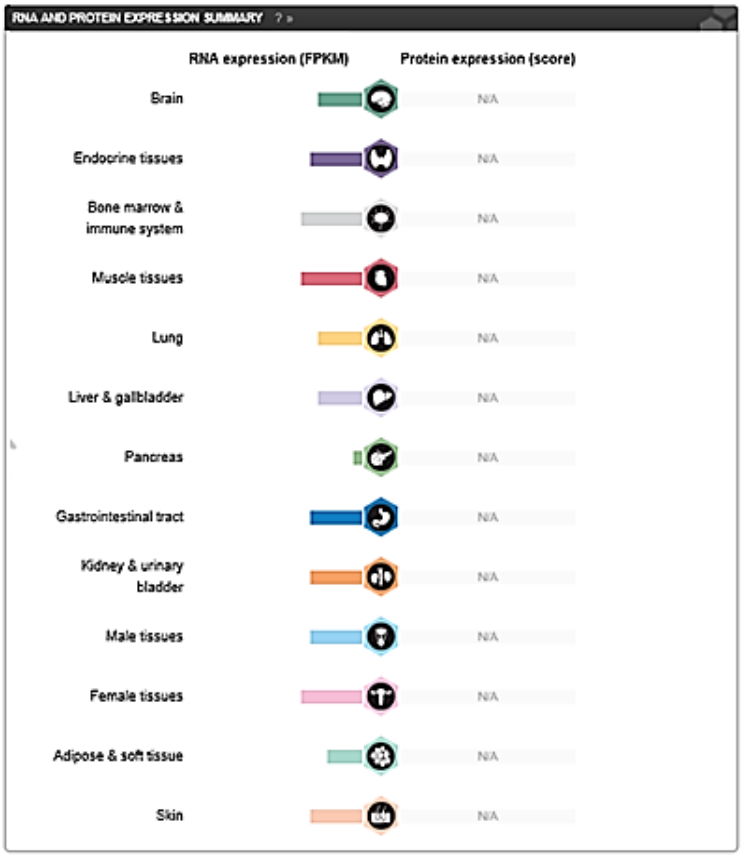


Figure 3.5 — Overview of *AGO1* RNA expression in different tissues. Analysed tissues are divided into 13 colour-coded groups according to common functional features. RNA expression results were obtained from RNA-seq analysis and are reported as number of fragments per kilobase gene model and million reads (FPKM). N/A indicates non-available information regarding AGO1 protein expression. Data from http://www.proteinatlas.org.

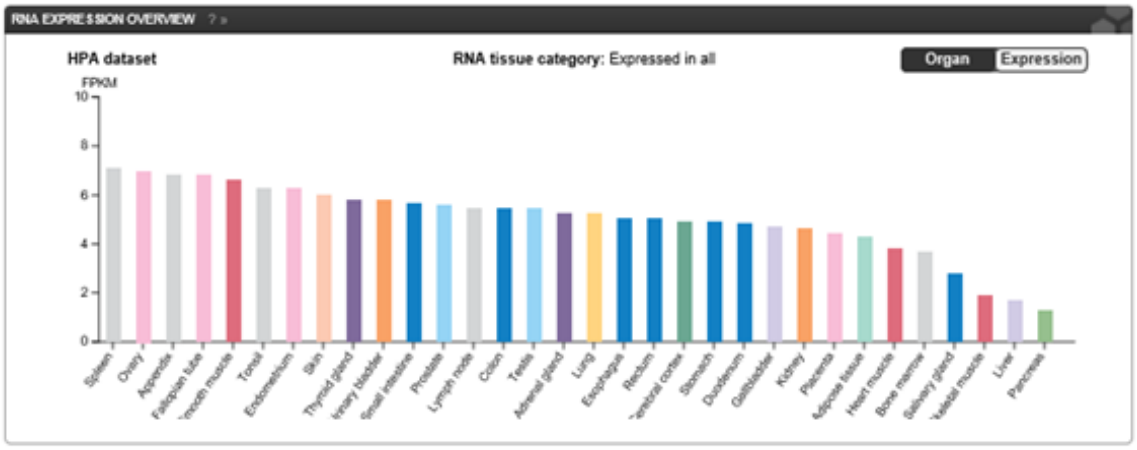
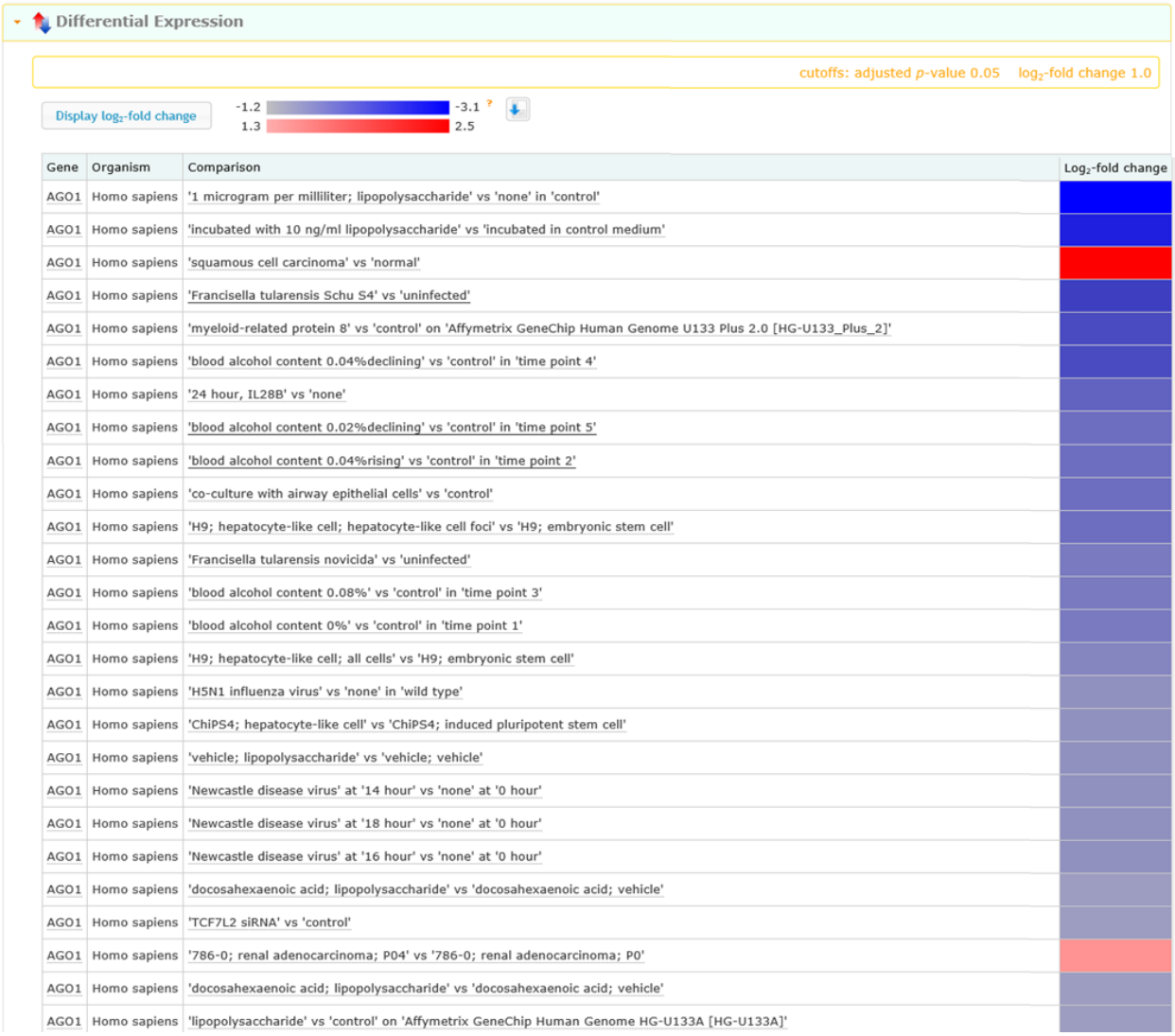


Figure 3.6 — Expression pattern of *AGO1* mRNA in different tissues. *AGO1* mRNA expression levels were obtained as RNA-seq data from 32 tissues. These are reported as mean FPKM (fragments per kilobase gene model and million reads), corresponding to mean values of the different individual samples from each tissue type. FPKM thresholds are set for categorisation of transcript expression levels into: “not detected” (0–0.5); “low” (0.5–10); “medium” (10–50); or “high” (>50). Data from http://www.proteinatlas.org.



Gene	Organism	Comparison	Log_2-fold change
AGO1	Homo sapiens	'1 microgram per milliliter; lipopolysaccharide' vs 'none' in 'control'	
AGO1	Homo sapiens	'incubated with 10 ng/ml lipopolysaccharide' vs 'incubated in control medium'	
AGO1	Homo sapiens	'squamous cell carcinoma' vs 'normal'	
AGO1	Homo sapiens	'Francisella tularensis Schu S4' vs 'uninfected'	
AGO1	Homo sapiens	'myeloid-related protein 8' vs 'control' on 'Affymetrix GeneChip Human Genome U133 Plus 2.0 [HG-U133_Plus_2]'	
AGO1	Homo sapiens	'blood alcohol content 0.04%declining' vs 'control' in 'time point 4'	
AGO1	Homo sapiens	'24 hour, IL28B' vs 'none'	
AGO1	Homo sapiens	'blood alcohol content 0.02%declining' vs 'control' in 'time point 5'	
AGO1	Homo sapiens	'blood alcohol content 0.04%rising' vs 'control' in 'time point 2'	
AGO1	Homo sapiens	'co-culture with airway epithelial cells' vs 'control'	
AGO1	Homo sapiens	'H9; hepatocyte-like cell; hepatocyte-like cell foci' vs 'H9; embryonic stem cell'	
AGO1	Homo sapiens	'Francisella tularensis novicida' vs 'uninfected'	
AGO1	Homo sapiens	'blood alcohol content 0.08%' vs 'control' in 'time point 3'	
AGO1	Homo sapiens	'blood alcohol content 0%' vs 'control' in 'time point 1'	
AGO1	Homo sapiens	'H9; hepatocyte-like cell; all cells' vs 'H9; embryonic stem cell'	
AGO1	Homo sapiens	'H5N1 influenza virus' vs 'none' in 'wild type'	
AGO1	Homo sapiens	'ChiPS4; hepatocyte-like cell' vs 'ChiPS4; induced pluripotent stem cell'	
AGO1	Homo sapiens	'vehicle; lipopolysaccharide' vs 'vehicle; vehicle'	
AGO1	Homo sapiens	'Newcastle disease virus' at '14 hour' vs 'none' at '0 hour'	
AGO1	Homo sapiens	'Newcastle disease virus' at '18 hour' vs 'none' at '0 hour'	
AGO1	Homo sapiens	'Newcastle disease virus' at '16 hour' vs 'none' at '0 hour'	
AGO1	Homo sapiens	'docosahexaenoic acid; lipopolysaccharide' vs 'docosahexaenoic acid; vehicle'	
AGO1	Homo sapiens	'TCF7L2 siRNA' vs 'control'	
AGO1	Homo sapiens	'786-0; renal adenocarcinoma; P04' vs '786-0; renal adenocarcinoma; P0'	
AGO1	Homo sapiens	'docosahexaenoic acid; lipopolysaccharide' vs 'docosahexaenoic acid; vehicle'	
AGO1	Homo sapiens	'lipopolysaccharide' vs 'control' on 'Affymetrix GeneChip Human Genome HG-U133A [HG-U133A]'	

Figure 3.7 — Expression of *AGO1* mRNA under different conditions or experimental factors. Greater colour saturation means higher absolute log_2-fold change value. Blue indicates the gene is down-regulated and red means up-regulated. Log_2-fold changes are not directly comparable across experiments. Data from http://www.ebi.ac.uk.

I.3. *In silico* analysis of MLH1 expression

We chose CRC and CC as the experimental models to be used in this work. Considering the panel of genes involved in CRC (Cragun *et al.*, 2014) and that of those involved in CC (Giarnieri *et al.*, 2000), we observed that MLH1 and MSH2 are common to both cancer types. After a comprehensive analysis of both genes and their expression patterns, we added MLH1 to our pool of putative candidates based on the evidence provided by *The Human Protein Atlas* and *Gene Expression Atlas* as shown below. An overview of RNA and protein expression levels of MLH1 in different groups of tissues (figure 3.8) indicates that MLH1 protein is expressed in high levels in most tissues but muscle, adipose and soft tissues, whereas RNA expression levels are mostly low in every tissue, especially in pancreas. A deeper look into the available information revealed a

high expression level of MLH1 protein in most tissues (figure 3.9.A), particularly those included in the gastrointestinal tract. This protein is also highly expressed in the immune system, lung and some female tissues. Conversely, it is not detected in skeletal or smooth muscle, liver, prostate, adipose nor parathyroid gland tissues. Regarding the counterpart RNA levels, reported FPKM indicate *MLH1* RNA is expressed at medium levels in most tissues, but in small intestine, skin, liver, salivary gland and pancreas, where its expression levels are low (i.e., below 10 FPKM).

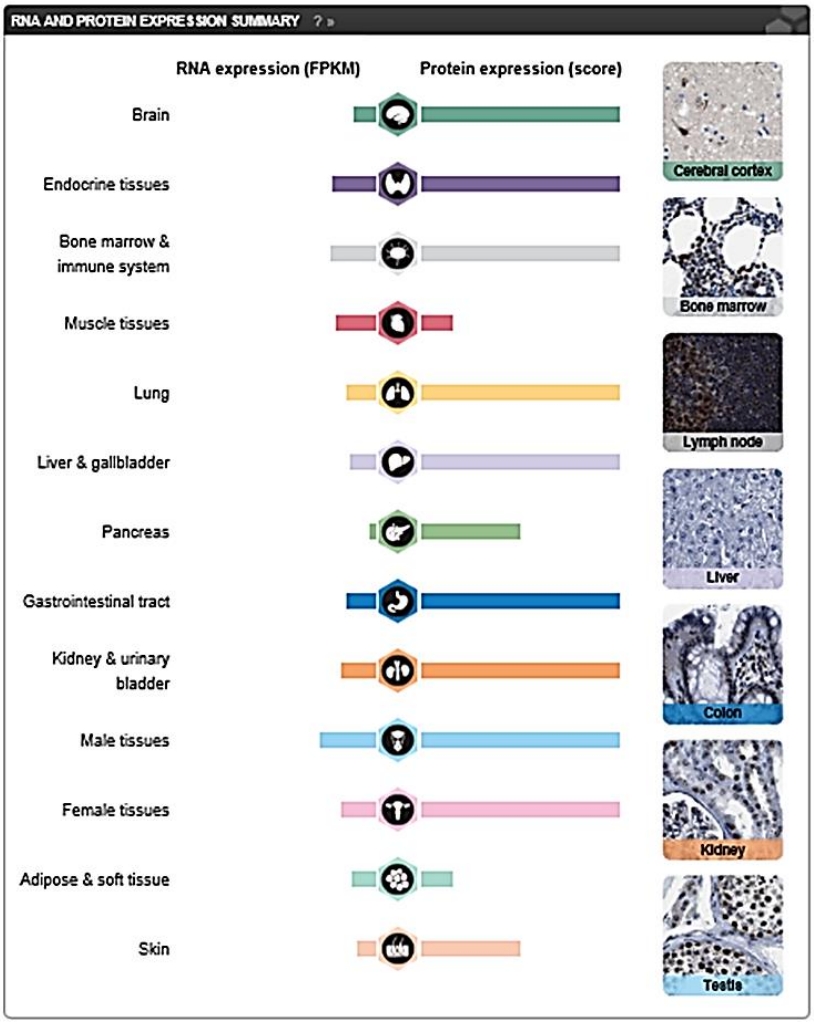


Figure 3.8 — Overview of MLH1 protein and mRNA expression data. Analysed tissues are divided into 13 colour-coded groups according to common functional features. Images of selected tissues give a visual summary of the protein expression profile (panels on the right). RNA expression results were obtained from RNA-seq analysis and are reported as number of fragments per kilobase gene model and million reads (FPKM).Protein expression scores represent the highest expression score found in a particular group of tissues, corresponding to the staining profile of the used antibody. Data from http://www.proteinatlas.org.

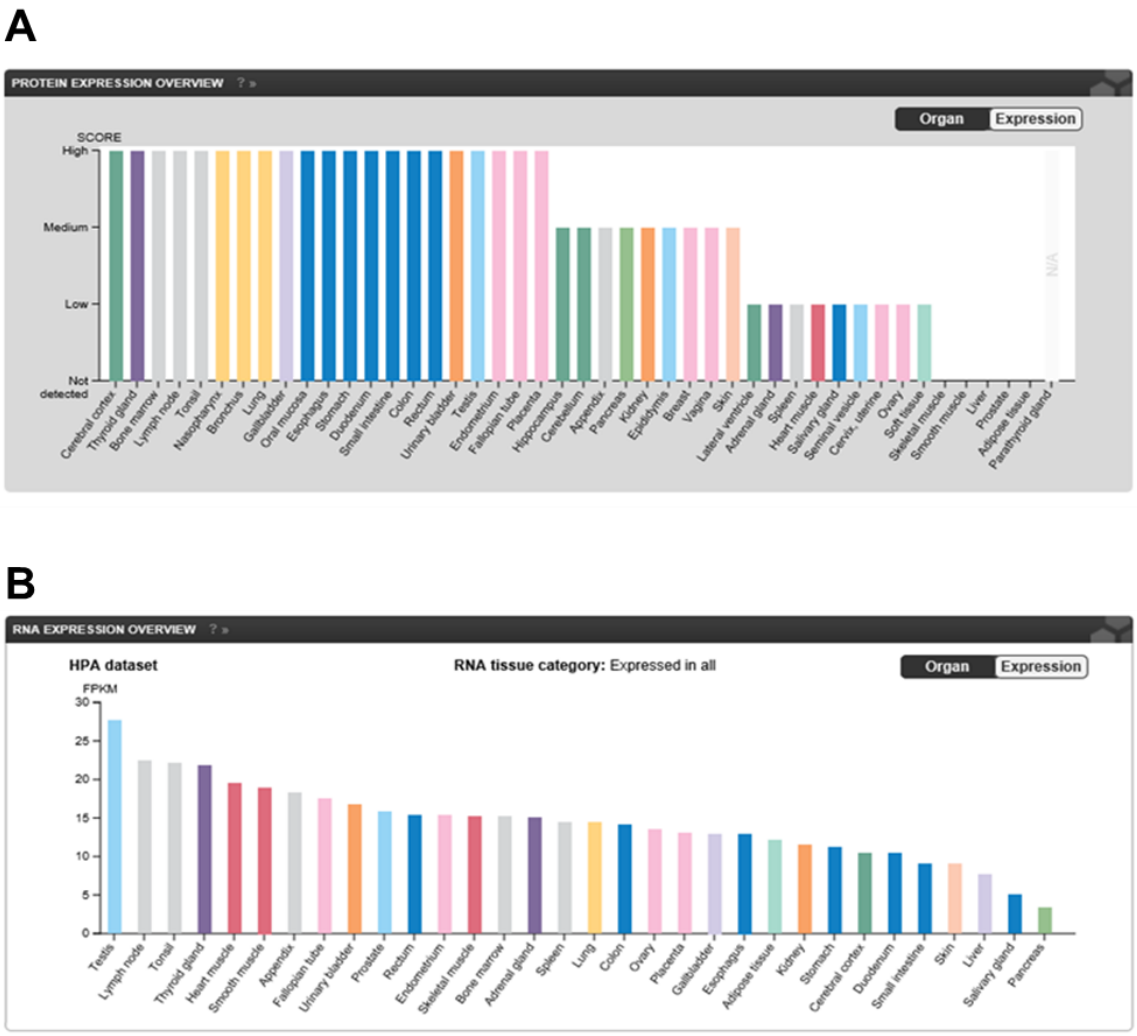


Figure 3.9 — Expression pattern of MLH1 in different tissues. (A) MLH1 protein expression levels in different tissues according to antibody immunostaining intensity (not detected, low, medium or high). The generated tissue microarrays include samples from 144 individuals corresponding to 44 different normal tissue types. (B) *MLH1* mRNA expression levels were obtained as RNA-seq data from 32 tissues. These are reported as mean FPKM (fragments per kilobase gene model and million reads), corresponding to mean values of the different individual samples from each tissue type. FPKM thresholds are set for categorisation of transcript expression levels into: “not detected” (0–0.5); “low” (0.5–10); “medium” (10–50); or “high” (>50). Data from http://www.proteinatlas.org.

As far as cancer tissues are concerned, data curated in *The Human Protein Atlas* indicates that glioma, melanoma and lymphoma, as well as skin, testicular and breast cancers, exhibited moderate-to-strong positivity, and that the remaining tumour cells mainly display weak-to-moderate staining (figure 3.10). Interestingly, nevertheless, is that cervical cancer presents a medium-to-strong MLH1 expression, whereas in the counterpart normal tissue its expression is low. A similar state of affairs comes about in prostate and liver cancers, in which MLH1 is expressed (from low to high levels) while in normal tissues it is not detected.

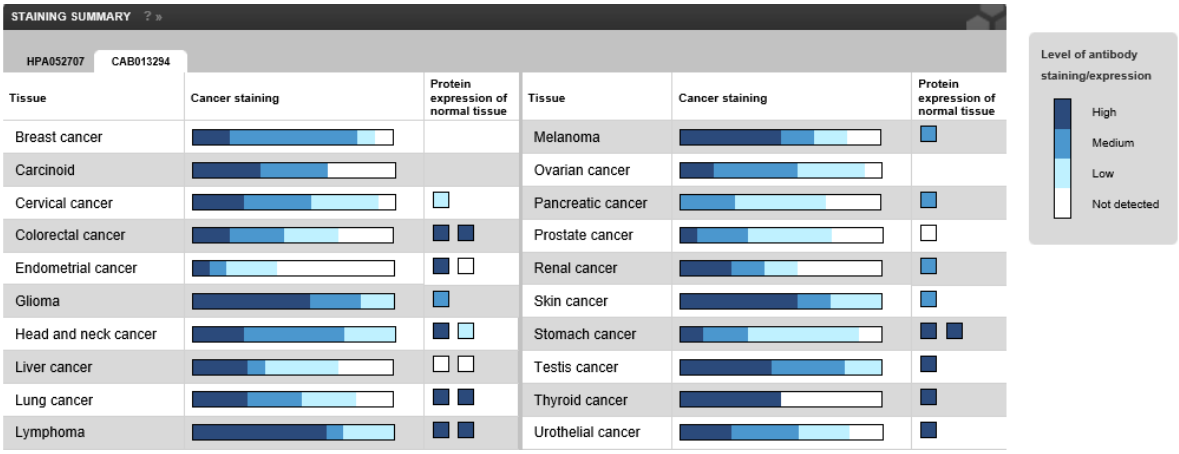
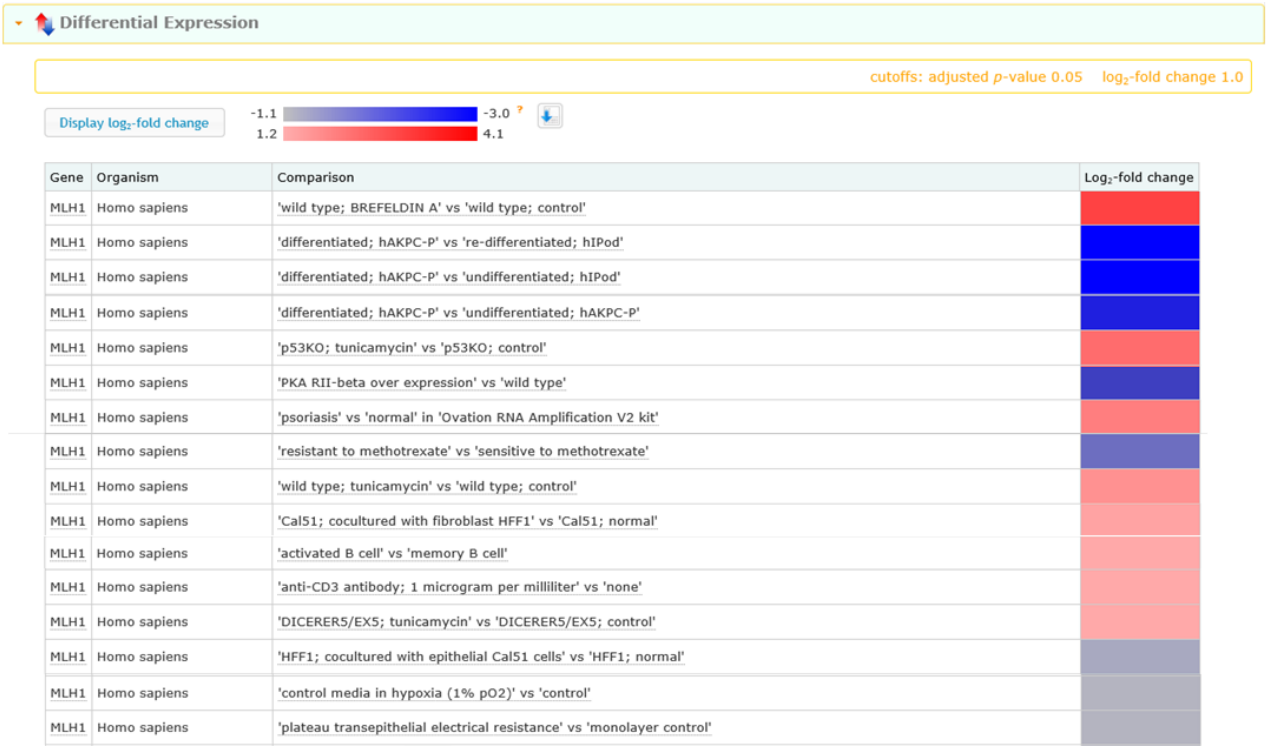


Figure 3.10 — Expression of MLH1 in different cancer tissues. For each cancer, the fraction of samples with antibody staining/protein expression levels — high, medium, low, or nor detected — are provided by the blue scale colour code, as described in the box to the right. The bar length represents the number of patient samples analysed (max=12 patients). Next to the cancer staining data, the protein expression data of normal tissues corresponding to each cancer are shown and protein expression levels are indicated by the blue scale colour code. Data from http://www.proteinatlas.org.

Such evidence may indicate that alternative mechanisms of translation initiation may account for the possibility of translating RNA poorly expressed.

In order to further understand to what extent *MLH1* RNA expression varies and whether such variation is the cause of protein expression alterations, we sought information about it in the *Gene Expression Atlas* (figure 3.11). *MLH1* is down-regulated in cells resistant to methotrexate *versus* cells sensitive to such drug, as well as in cells with overexpression of protein kinase cAMP-dependent type II regulatory subunit beta (PKA RII-β) compared to wild-type ones, or in hypoxia *versus* control conditions. On the other hand, treatments with Brefeldin A (a lactone antibiotic that inhibits intracellular protein transport), or tunicamycin, lead to an up-regulation of *MLH1* RNA levels. Also, *MLH1* RNA is up-regulated in activated B cells compared to memory T cells.

Altogether, these data indicate that MLH1 expression pattern is not straightforward as far as RNA and protein expression relate to each other and therefore the possibility of a non-canonical mechanism of translation initiation being governing MLH1 protein synthesis cannot be ruled out. In this regard, we compelled ourselves to experimentally validate this potential candidate alongside the other two previously selected and test their 5'UTR for the presence of possible *cis*-acting elements capable of driving translation initiation either via cap-dependent or -independent mechanisms.



Differential Expression

cutoffs: adjusted *p*-value 0.05 log_2-fold change 1.0

Display log_2-fold change -1.1 -3.0 1.2 4.1

Gene	Organism	Comparison	Log_2-fold change
MLH1	Homo sapiens	'wild type; BREFELDIN A' vs 'wild type; control'	
MLH1	Homo sapiens	'differentiated; hAKPC-P' vs 're-differentiated; hIPod'	
MLH1	Homo sapiens	'differentiated; hAKPC-P' vs 'undifferentiated; hIPod'	
MLH1	Homo sapiens	'differentiated; hAKPC-P' vs 'undifferentiated; hAKPC-P'	
MLH1	Homo sapiens	'p53KO; tunicamycin' vs 'p53KO; control'	
MLH1	Homo sapiens	'PKA RII-beta over expression' vs 'wild type'	
MLH1	Homo sapiens	'psoriasis' vs 'normal' in 'Ovation RNA Amplification V2 kit'	
MLH1	Homo sapiens	'resistant to methotrexate' vs 'sensitive to methotrexate'	
MLH1	Homo sapiens	'wild type; tunicamycin' vs 'wild type; control'	
MLH1	Homo sapiens	'Cal51; cocultured with fibroblast HFF1' vs 'Cal51; normal'	
MLH1	Homo sapiens	'activated B cell' vs 'memory B cell'	
MLH1	Homo sapiens	'anti-CD3 antibody; 1 microgram per milliliter' vs 'none'	
MLH1	Homo sapiens	'DICERER5/EX5; tunicamycin' vs 'DICERER5/EX5; control'	
MLH1	Homo sapiens	'HFF1; cocultured with epithelial Cal51 cells' vs 'HFF1; normal'	
MLH1	Homo sapiens	'control media in hypoxia (1% pO2)' vs 'control'	
MLH1	Homo sapiens	'plateau transepithelial electrical resistance' vs 'monolayer control'	

Figure 3.11 — Expression of *MLH1* RNA under different conditions or experimental factors. Greater colour saturation means higher absolute log_2-fold change value. Blue indicates the gene is down-regulated and red means up-regulated. Log_2-fold changes are not directly comparable across experiments. Data from http://www.ebi.ac.uk.

II. Expression of human UPF1 is regulated by a cap-independent translation initiation mechanism

After selecting UPF1 as a putative candidate whose expression might be regulated via an alternative mechanism of translation initiation, experimental validation is required to either confirm or rule out the existence of *cis*-elements within its 5'UTR capable of driving cap-independent translation initiation or an alternative cap-dependent mechanism that function in non-canonical conditions.

II.1. *In silico* analysis of *UPF1* 5'UTR characteristics reveals features common to many IRES-containing 5'UTR

We first decided to submit *UPF1* 5'UTR to an *in silico* analysis prior to experimental validation, in order to understand whether its characteristics would support the hypothesis of elements capable of mediating cap-independent translation initiation existing therein. This sequence is similar to the reference in NCBI (NM_001297549.1), but instead of a C in position 145 there is a G, a polymorphism common to 88% of world human population, according to NCBI information (figure 3.12.A). We decided to use the sequence containing the latter for considering its better representing of the most frequent occurrence. The 5'UTR is composed of 275 nucleotides (figure 3.12.A), with overall 78% GC content, distributed over regions of very high GC content (up to 85% at regions adjacent to 3' terminus) and others with lower content (minimum 75% at regions adjacent to 5' terminus, figure 3.12.B), and tendency to fold into structures of predicted stability — ΔG = -141.35kcal/mol (prediction with mFold, figure 3.12.C). According to this prediction, the secondary structure formed within *UPF1* 5'UTR includes three stem loops (SL) — I, II and III. SLIII corresponds to the region containing the highest GC content, suggesting a great stability for the predicted stem loop. The formation of such structures may allow ribosome recruitment to the vicinity of the main AUG and thus may help in facilitating cap-independent translation initiation via internal entry of the ribosome (cap-independent mechanism). On the other hand, these structures may impair the regular scanning of the 5'UTR and, hence, promote mechanisms of ribosome shunting that force the ribosome to bypass them and reach the AUG in a non-canonical way. However, we looked for sequence similarities between UPF1 5'UTR and ribosomal RNA and found none (data not shown), which considerably reduces the possibility of ribosomes being shunted across UPF1 5'UTR. Furthermore, there are no upstream AUG within this untranslated region. However, there are 7 CUG codons, which cannot allow us to completely rule out possible uORFs regulating UPF1 protein expression.

Sequence conservation throughout evolution may also provide some clues on the importance of the sequence under analysis and how it may be involved in the process of translation initiation. For that purpose, we decided to compare the human *UPF1* 5'UTR to those of other mammalian species to evaluate how conserved among species this sequence is and also to what extent the formation of the predicted stem loops is maintained among different species. Using Bioedit software, we compared the sequence of *UPF1* 5'UTR from different mammals (human, chimpanzee, orangutan, rat, mouse, Guinea pig, dog, cow and cat, figure 3.13.A). Out of these, human, chimpanzee, and orangutan *UPF1* 5'UTR are almost perfectly aligned, as only a few

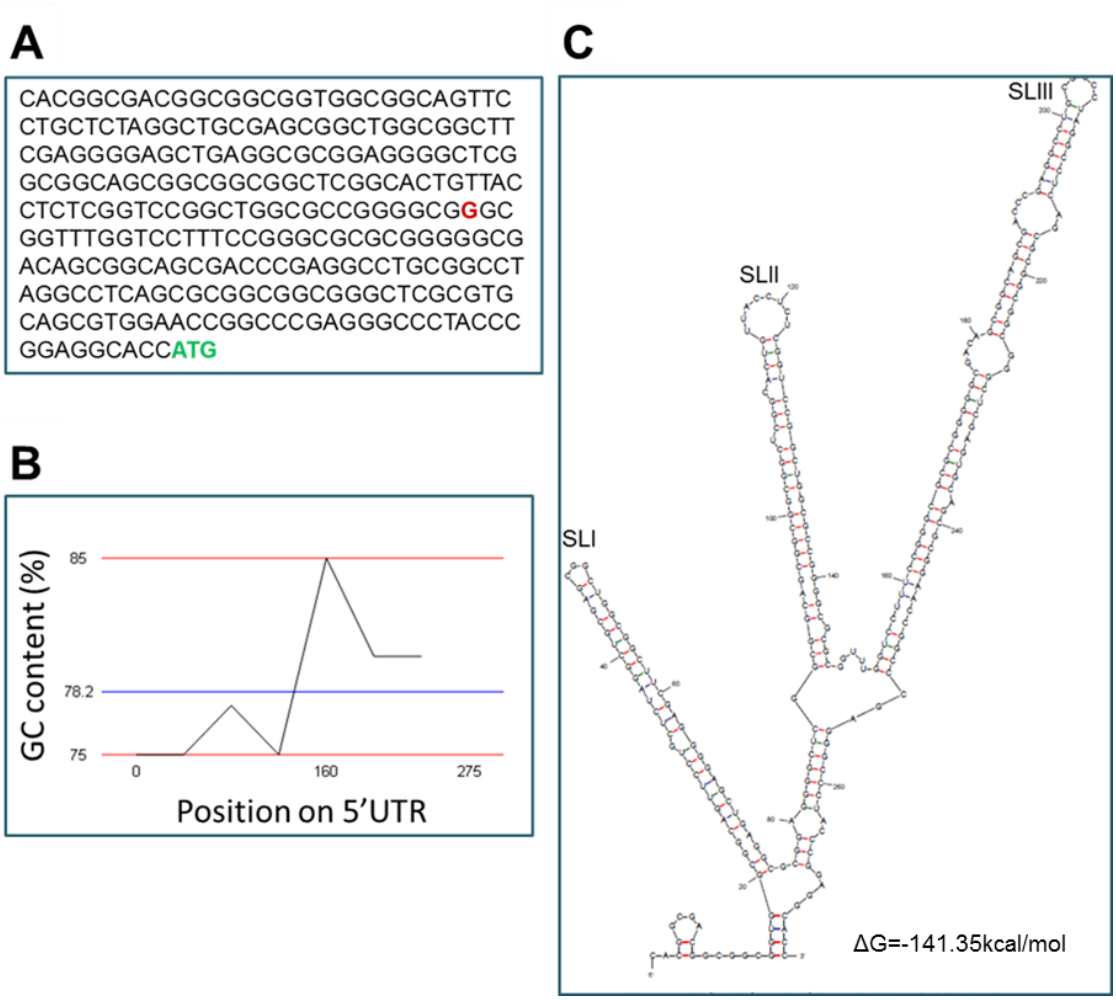


Figure 3.12 — Multiple features of human *UPF1* 5'UTR predicted *in silico*. (A) Nucleotide sequence of human *UPF1* 5'UTR used in this work. **G** represents a polymorphism at position 145 common in 88% of world population, instead of the reference **C**. **ATG** is the UPF1 translation initiation codon. (B) Calculation of GC contents (%) of different *UPF1* 5'UTR regions (http://www.endmemo.com). The content (average 78%) ranges from 75% to 85% and the highest percentages tend to localise towards the 3' end of the 5'UTR. **C)** *UPF1* 5'UTR RNA secondary structure predicted by mFold software (http://mfold.rna.albany.edu/?q=mfold). Based on this prediction, three stem loops (SL) — I, II and III — are formed within UPF1 5'UTR and the structure is very stable (ΔG = -141.35kcal/mol)**.**

nucleotides do not match among them. Mouse *UPF1* 5'UTR includes a 26-nucleotide portion spanning from nucleotide 102 to 128 that is absent from all other considered species. As for the others, there is a great degree of conservation among all the sequences, especially at the 3' portion of the untranslated region — from nucleotide 163 of human *UPF1* 5'UTR to the end of the sequence. Furthermore, the predicted consensus secondary structure of the set of aligned sequences obtained using RNAalifold webserver (figure 3.13.B) is composed of 3 stem loops exactly as in the mFold prediction (figure 3.12.C). According to the results obtained, formation of SLIII seems to be conserved throughout evolution, whereas formation of SLI is the less conserved. All in all, these results indicate the sequence comprising nucleotides 163 to 275 is highly conserved and predictably able to be folded in an utterly stable stem loop, suggesting it

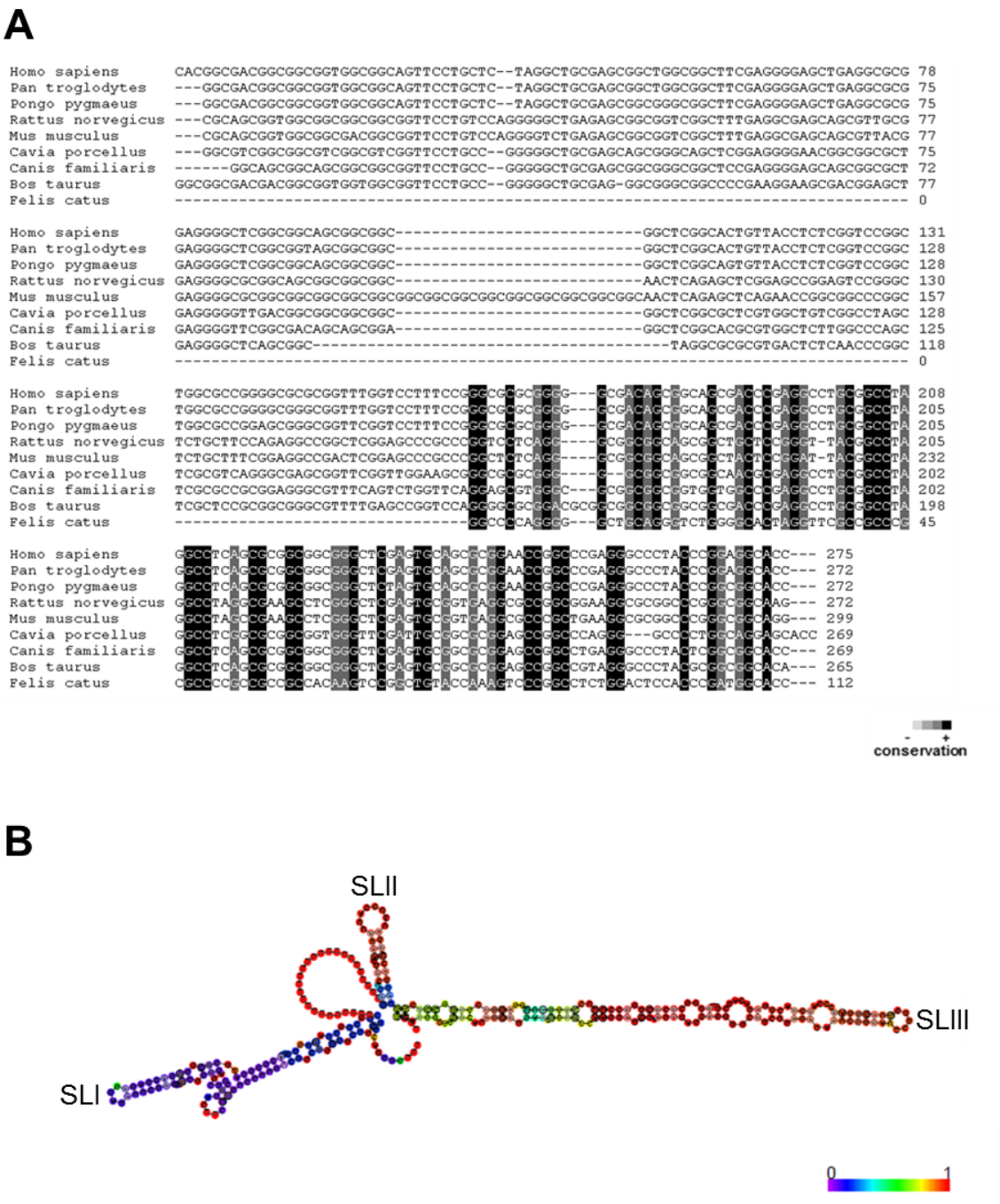


Figure 3.13 — Conservation of *UPF1* 5'UTR sequence among mammalian species. *(*A) Sequence alignment of UPF1 5'UTR among human, chimpanzee, orangutan, rat, mouse, Guinea pig, dog, cow and cat obtained using Biodit software (http://www.bioinformatics.org/sms2/color_align_cons.html). Grey scale indicates the degree of conservation among species for each nucleotide. White (-) indicates least conserved; black (+) indicates most conserved. (B) Predicted consensus secondary structure of the set of aligned sequences using RNAalifold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi). Coloured scale indicates the degree of conservation of the predicted secondary structure. Purple (0) indicates no secondary structure conservation; Red (1) indicates full secondary structure conservation.

may be involved in a conserved mechanism of gene expression regulation. Another interesting feature observed in *UPF1* 5'UTR is the prediction of G-quadruplex formation. These structures can dramatically influence translation effectiveness, because they block cap-dependent translation initiation, acting as a means of regulating protein levels in the cell (Beaudoin and Perreault, 2010). On the other hand, cap-independent translation can be enhanced by the formation of G-quadruplexes, which can add an extra layer of gene expression regulation. According to QGRS Mapper software, which predicts Quadruplex-forming G-rich sequences (QGRS), human *UPF1* 5'UTR is predicted to form several G-quadruplex structures as illustrated in table 3.1. G-quadruplexes might contribute to control translation initiation driven by *UPF1* 5'UTR.

Although many of the analysed features so far suggest the possibility of non-canonical mechanisms of translation initiation to occur, experimental validation is required and the only way to confirm or rule out such hypothesis. In this regard, we decided to start our approach by evaluating the possibility of this sequence being able to mediate a cap-independent mechanism of translation initiation.

Table 3.1 — Quadruplex-forming G-rich sequences (QGRS) found in human *UPF1* 5'UTR*

Position	Length	QGRS	G-Score
4	20	GGCGACGGCGGCGGTGGCGG	21
55	29	GGCTTCGAGGGGAGCTGAGGCGCGGAGGG	21
91	19	GGCAGCGGCGGCGGCTCGG	20
139	16	GGGGCGGGCGGTTTGG	20
163	22	GGGCGCGCGGGGGGCGACAGCGG	15
203	25	GGCCTAGGCCTCAGCGCGGCGGCGG	16
243	26	GGAACCGGCCCGAGGGCCCTACCCGG	17

*The underlined GG represent those putatively involved in the formation of G-quadruplex structures. Position designates the first nucleotide of the QGRS sequences. The putative G-quadruplexes are identified using the motif $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$, where x is the number of guanine tetrads in the G-quadruplex, and y1, y2, y3 are the length of gaps, i.e., the length of the loops connecting the guanine tetrads. The motif consists of four equal length groups of guanines, separated by arbitrary nucleotide sequences with at least two tetrads (x≥2) and maximum length of 30 bases. The maximum length of 30 bases restricts G-groups to a maximum size of 6. G-score is a classification attributed by the software that evaluates a QGRS for its likelihood to form a stable G-quadruplex. Higher scoring sequences will make better candidates for G-quadruplex. The scoring method considers the following principles: shorter loops are more common than longer loops; G-quadruplexes tend to have loops roughly equal in size; the greater the number of guanine tetrads, the more stable the quadruplex. The highest possible G-score, using the default maximum QGRS length of 30, is 105. (http://bioinformatics.ramapo.edu/QGRS/index.php)

For that purpose, we utilised a bicistronic system — the *pR_F*, based on the psicheck-2 vector by Promega. It has been proven to be useful for identifying sequences capable of mediating cap-independent translation initiation, because it includes two cistrons encoding for two reporter proteins within the same plasmid — one, further upstream, whose translation is cap-dependent, and another one, further downstream, whose translation will occur if ever a sequence upstream its AUG can drive its translation initiation —that will be transcribed as a single mRNA. In the chosen system, *pR_F*, the 5' cistron is RLuc, which encodes for the ORF of *Renilla reniformis* luciferase protein, whereas the 3' cistron is FLuc, which encodes for the ORF of firefly (*Photynus pyralis)* luciferase protein. Both proteins catalyse chemical reactions in which either luciferin is transformed in oxyluciferin (in the case of FLuc) or coelenterazine is transformed in coelenteramide (in the case of RLuc). In both cases, one of the reaction products is light that can be quantified in a luminometer and is directly proportional to the amount of enzyme being expressed. Thus, in our system, RLuc luminescence acts as an internal control and FLuc luminescence will indicate the amount of protein synthesized under the control of the sequence cloned upstream FLuc AUG. Due to the extreme sensitivity, this system allows quantification of even small changes in protein synthesis.

II.2. *UPF1* 5'UTR drives FLuc expression in a bicistronic context

Taking the above information into account, we, then, cloned *UPF1* 5'UTR, or the counterpart control sequences (*HBB* 5'UTR[23], *c-Myc* IRES[24] and *EMCV*[25] IRES) upstream FLuc ORF in the bicistronic plasmid previously described, in order to obtain the constructs depicted in figure 3.14. In these constructs, RLuc translation is cap-dependent and represents an internal control, whereas FLuc will be translated if the upstream cloned sequence can drive its translation; transcription of the bicistronic plasmid is under the control of SV40 promoter and enhancer; and the stable hairpin cloned downstream RLuc ORF prevents translation reinitiation events. The empty vector, *pR_F*, is the negative control for any non-canonical activity, as it does not contain any insert between RLuc and FLuc ORF sequences. *HBB* 5'UTR-containing vector, *pR_HBB_F*, is the negative control for non-canonical translation initiation, because it cannot mediate alternative mechanisms of translation initiation. The positive cellular control for non-canonical activity is the IRES sequence included in the *c-Myc* 5'UTR. This sequence was first identified by Stoneley *et al.* (1998, 2000) and, although there is some controversy regarding whether it is a true IRES (Bert *et al.*, 2006), it is for sure capable of driving cap-independent translation and is widely used as a positive control for cap-independent translation initiation activity (Ozretić *et al.*, 2015). *Encephalomiocarditis* IRES sequence (Bochkov and Palmenberg, 2006) was our choice as a viral positive control. This is a very strong IRES element that works very well in human cells, thus providing us a trustworthy control that the system is working properly. The relative FLuc activity measured from each of the aforementioned constructs provides us a reliable control for cap-independent activity. Whenever such activity is significantly greater than 1, as in the case of

[23]Human β-globin (*HBB*) 5' untranslated region.
[24]v-myc avian myelocytomatosis viral oncogene homolog (*c-Myc*) IRES sequence.
[25]*Encephalomyocarditis* virus (*EMCV*) IRES sequence

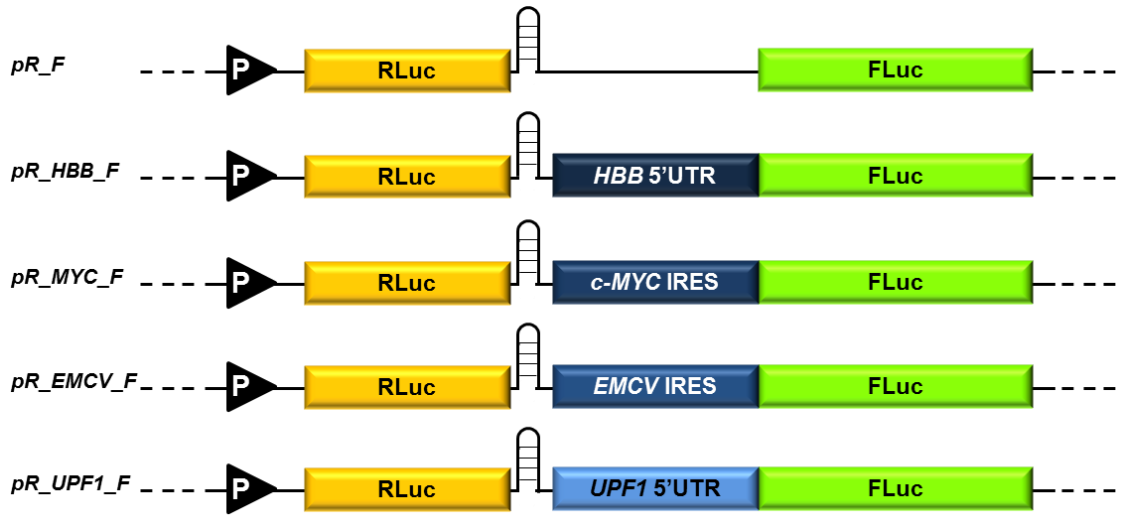


Figure 3.14 — Schematic representation of the constructs used to check whether *UPF1* 5'UTR is able to drive cap-independent translation initiation. RLuc is the *Renilla* luciferase cap-dependent translated cistron (yellow box) and FLuc the firefly cap-independent-translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Boxes in different shades of blue represent the different sequences cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_HBB_F,* the human *β-globin* (*HBB*) 5'UTR-containing vector, is the negative control for cap-independent activity; *pR_MYC_F,* the *c-MYC* IRES-containing vector, is the cellular positive control for cap-independent activity; *pR_EMCV_F,* the *EMCV* IRES-containing vector, is the viral positive control for IRES activity; *pR_UPF1_F,* the *UPF1* 5'UTR-containing vector, is the sequence under study. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

the plasmids containing *c-MYC* or *EMCV* IRES sequences, that means the cloned sequence is able to mediate a cap-independent mechanism of translation initiation; conversely, if the relative FLuc activity is ≤1, i.e., similar to that of the empty plasmid, as is the case of the *HBB* 5'UTR-containing plasmid, that means the cloned sequence cannot drive cap-independent translation initiation. Thus, this system allows us to validate the experimental conditions for cap-independent activity and understand whether a sequence is able to drive translation initiation in non-canonical conditions. All the used controls, apart from granting the reliability of the bicistronic system, allow us to evaluate the activity of any detected cap-independent activity and to understand how strong such activity might be. As a final point, all the constructs used in this study contain a stable hairpin (Candeias *et al.*, 2006) downstream RLuc ORF that helps blocking ribosomes and preventing false positive results as a consequence of reinitiation events.

Having established all positive and negative controls required to assess the ability of *UPF1* 5'UTR to drive cap-independent translation initiation, we transfected the previously selected cell lines — HeLa, NCM460 and HCT116 — with the plasmids described above — *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F* and *pR_UPF1_F* — and, by luminometry assays, compared the relative FLuc expression from each of the transfected constructs with that from the empty counterpart. Thus, in all tested cell lines, there was a significant increase in relative FLuc expression from *pR_UPF1_F* compared to *pR_F* (figure 3.15). The increase in relative FLuc

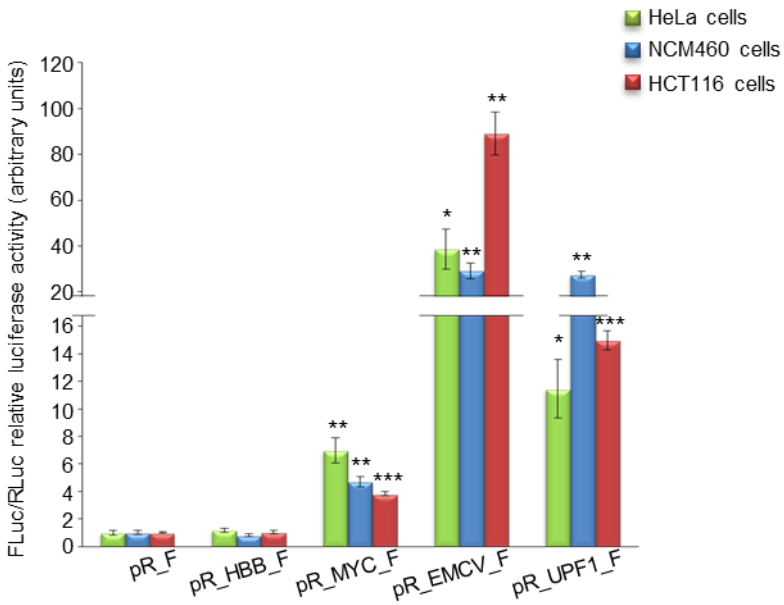


Figure 3.15 — Expression of FLuc reporter protein is mediated by *UPF1* 5'UTR in a bicistronic context. HeLa (green bars), NCM460 (blue bars), and HCT116 (red bars) cells were transfected either with *UPF1* 5'UTR-containing plasmid (*pR_UPF1_F*) or with one of the controls used in the experiment: *pR_F* (empty plasmid), *pR_HBB_F* (*HBB* 5'UTR-containing plasmid), *pR_MYC_F* (*c-Myc* IRES-containing plasmid), or *pR_EMCV_F* (*EMCV* IRES-containing plasmid). Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. * $P<0.05$, ** $P<0.01$, *** $P<0.001$

expression was 11.5-fold in HeLa cells, 27.4-fold in NCM460 cells and 15.0-fold in HCT116 cells, indicating *UPF1* 5'UTR mediates FLuc expression in all tested cell lines. This suggests *UPF1* 5'UTR is able to drive a non-canonical mechanism responsible for the increased expression of FLuc. Also, this mechanism seems to be stronger in normal cells (NCM460) than in cancer ones (HeLa and HCT116), as the increase in relative FLuc expression was much greater in the former than in the latter. Regarding the constructs containing the chosen positive and negative controls for cap-independent activity, we observed a significant increase in relative FLuc expression from plasmids containing *c-MYC* and *EMCV* IRES sequences but not from the one containing *HBB* 5'UTR sequence, which reflects the expected outcome of the experiment. Accordingly, the relative FLuc expression from *pR_HBB_F* was similar to that from *pR_F* in all cell lines (1.2-fold in HeLa cells, 0.8-fold in NCM460 cells and 1.0-fold in HCT116 cells), confirming *HBB* 5'UTR sequence is not able to mediate translation of a downstream ORF in a non-canonical manner.

As for relative FLuc expression levels from *pR_MYC_F*, we observed 7.0-, 4.7- and 3.8-fold increase compared to that from *pR_F* in HeLa, NCM460 and HCT116 cells, respectively, indicating this sequence can drive FLuc translation in a bicistronic context in all tested cell lines. Furthermore, in HeLa cells such activity is greater than in colon-derived cells (NCM460 and HCT116), suggesting that c-Myc cap-independent activity may vary depending on the tissue. As for as the relative FLuc expression levels from *pR_EMCV_F,* we observed a 29.7-, 38.6- and

89.2-fold increase compared to the empty counterpart in HeLa, NCM460 and HCT116 cells, respectively. These results indicate the viral sequence is able to mediate cap-independent translation initiation in all tested cell lines, being mainly active in colorectal carcinoma cells. In all three analysed cell lines, relative FLuc expression from *pR_UPF1_F* is greater than that from *pR_MYC_F*, but lower than that from *pR_EMCV_F*. All in all, the outcome of this experiment indicates *UPF1* 5'UTR can drive FLuc expression in a non-canonical manner, similar to what *c-Myc* and *EMCV* IRES elements are able to do.

II.3. *UPF1* 5'UTR contains a cryptic promoter

One caveat of the bicistronic system is the possibility of generating false-positive results due to the existence of cryptic promoters or alternative splicing events. The existence of a cryptic promoter within the sequence under study may originate a monocistronic transcript encoding only FLuc that will be translated via the canonical cap-dependent translation initiation mechanism. This would dramatically increase the levels of FLuc protein measured by luminometry assays, as these would be the result of cap-dependent translation initiation from the monocistronic transcript plus the possible FLuc expression resulting from a cap-independent mechanism of translation initiation.

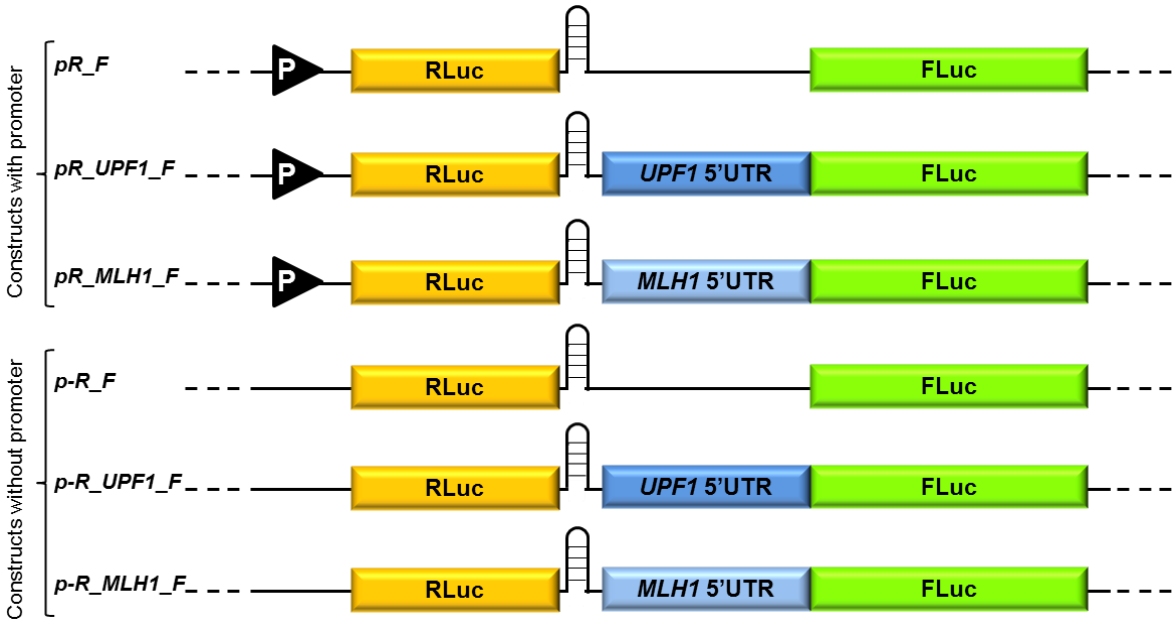


Figure 3.16 — Schematic representation of the constructs used to check whether *UPF1* 5'UTR contains a cryptic promoter. RLuc is the *Renilla* luciferase cap-dependent-translated cistron (yellow box) and FLuc the firefly cap-independent-translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Boxes in different shades of blue represent the different sequences cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_UPF1_F,* the *UPF1* 5'UTR-containing vector, is the sequence under study; *pR_MLH1_F*, the *MLH1* 5'UTR-containing vector, is the positive control for the presence of cryptic promoters. *p-R_F*, *p-R_UPF1_F* and *p-R_MLH1_F* are the counterpart promoterless plasmids. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

Thus, we created a promoterless construct containing *UPF1* 5'UTR to rule out the presence of false positives and, as a positive control for this matter, we used a human *MLH1* 5'UTR-containing plasmid, as such sequence has been described to include a cryptic promoter (Ito et al., 1999; Arita et al., 2003) (figure 3.16). We co-transfected HeLa, NCM460 and HCT116 cells with each of the bicistronic constructs used to evaluate the presence of cryptic promoters — promoter-containing *pR_F*, *pR_UPF1_F* and *pR_MLH1_F*, and the corresponding promoterless counterparts: *p-R_F*, *p-R_UPF1_F* and p-*R_MLH1_F* — and the β-galactosidase-encoding plasmid (pSV-β-Galactosidase Control Vector by Promega, a control vector for monitoring transfection efficiencies of mammalian cells). We measured relative RLuc and FLuc expression levels in every tested cell lines and observed FLuc was expressed from promoterless plasmids containing *UPF1* 5'UTR, which indicates this sequence is able to drive transcription and the concomitant production of a monocistronic transcript translated via the canonical cap-dependent mechanism of translation initiation (figure 3.17). Regarding relative RLuc expression, its levels were expected to be much greater in constructs containing the SV40 promoter than in those it has been removed. Accordingly, RLuc expression levels from promoterless constructs are actually virtually inexistent. This may be explained by some alternative splicing event occurring between RLuc and the 5'UTR sequence, which is reducing RLuc expression levels. In NCM460, all promoter-containing plasmids produce similar levels of relative RLuc expression, suggesting the cloned sequence did not affect transfection efficiency in this cell line (figure 3.17.B). As far as relative FLuc expression is concerned, we observed a significant increase in its levels from *pR_UPF1_F* compared to those from *pR_F* in all tested cell lines, corroborating the previously obtained results. Relative FLuc expression levels from *pR_MLH1_F* were also significantly greater than those from the empty vector, as expected, due to the presence of the cryptic promoter. Accordingly, the relative FLuc expression levels from promoterless *MLH1* 5'UTR-containing plasmid were also significantly greater than that from *p-R_F*, the empty promoterless plasmid. Interestingly, promoterless *p-R_UPF1_F* also expresses FLuc reporter protein in levels significantly greater than those from *p-R_F*, in all tested cell lines. This result suggests that UPF1 5'UTR sequence contains a cryptic promoter region that originates a monocistronic transcript whose translation occurs in a cap-dependent manner. The presence of such cryptic promoter originates false-positive results from the bicistronic system as the relative FLuc expression levels measured from the monocistronic transcript encoding FLuc ORF only mask the relative FLuc expression levels derived from a putative cap-independent mechanism of translation initiation mediated by *UPF1* 5'UTR and the latter cannot be detected. Although this unexpected result stands in the way of our initial purpose, it is something we cannot ignore, since it is the first time a promoter region is described within the sequence assigned to the 5'UTR of the *UPF1* transcript. To map which portion of this region is required for promoter activity, we did a deletional analysis using the deletional mutants depicted in figure 3.18. These mutants are the result of sequential deletions performed either from 5' to 3' end or the other way around by removing 50 nucleotides at a time. Thus, the 5' deletional mutants — *pR_51-275_F*, *pR_101-275_F*, *pR_151_275_F* and *pR-201-275_F* — and the counterpart promoterless plasmids — *p-R_51-275_F*, *p-R_101-275_F*,

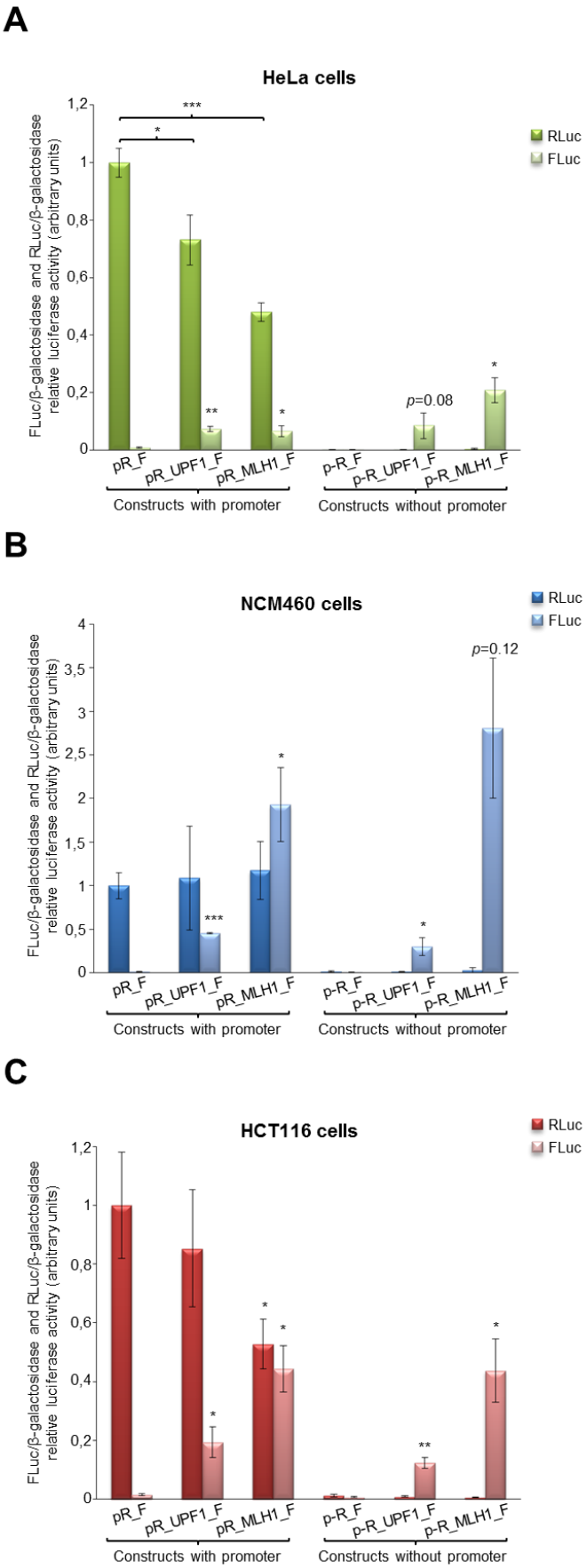


Figure 3.17 — *UPF1* 5'UTR contains a cryptic promoter active in all tested cell lines. HeLa (A), NCM460 (B) and HCT116 (C) cells were transfected with promoter-containing constructs (*pR_F*, *pR_UPF1_F* and *pR_MLH1_F*) or promoterless constructs (*p-R_F*, *p-R_UPF1_F* and *p-R_MLH1_F*), and co-transfected with *β-galactosidase*-encoding plasmid (*pSV-β-Galactosidase* Control Vector), an internal control for transfection efficiency in mammalian cells. Relative RLuc (dark green, dark blue and red bars) and FLuc (light green, light blue and pink bars) expression levels were obtained by normalising each of them to those from β-galactosidase-expressing plasmid, all measured by luminometry assays. Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. **P*<0.05, ***P*<0.01, ****P*<0.001

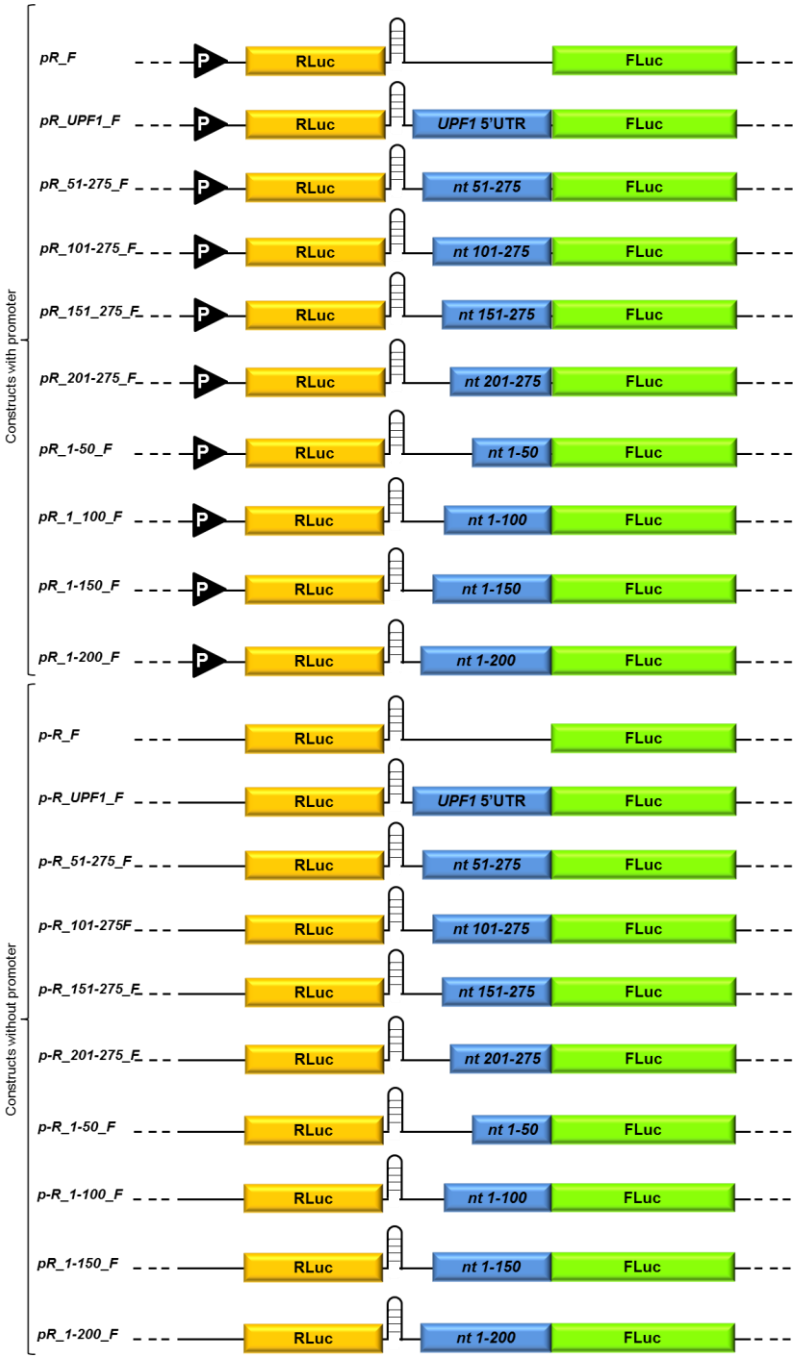


Figure 3.18 — Schematic representation of the deletional mutant constructs used to localise the cryptic promoter sequence. RLuc is the *Renilla* luciferase cap-dependent-translated cistron (yellow box) and FLuc the firefly cap-independent-translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Different size blue boxes represent the different deletional mutants cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_UPF1_F*, the full-length *UPF1* 5'UTR-containing plasmid; *pR_51-275_F*, the *UPF1* 5'UTR nucleotides (nt) 51–275-containing plasmid; *pR_101-275_F*, the *UPF1* 5'UTR nt 101–275-containing plasmid; *pR_151-275_F*, the *UPF1* 5'UTR nt 151–275-containing plasmid; *pR_201-275_F*, the *UPF1* 5'UTR nt 201–275-containing plasmid; *pR_1-50_F*, the *UPF1* 5'UTR nt 1–50-containing plasmid; *pR-1-100_F*, the *UPF1* 5'UTR nt 1–100-containing plasmid; *pR-1-150_F*, the *UPF1* 5'UTR nt 1–150-containing plasmid; *pR-1-200_F*, the *UPF1* 5'UTR nt 1–200-containing plasmid. *p-R_F*, *p-R_UPF1_F*, *p-R_51-275_F*, *p-R_101-275_F*, *p-R_151_275_F*, *p-R_201-275_F*, *p-R_1-50_F*, *p-R_1-100_F*, *p-R_1-150_F* and *p-R_200_F* are the counterpart promoterless constructs. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

p-R_151_275_F and *p-R-201-275_F* — were obtained by removing 50, 100, 150 or 200 nucleotides, respectively, from the 5' end of *UPF1* 5'UTR full sequence. As for the 3' deletional mutants — *pR_1-50_F*, *pR_1-100-F*, *pR_1-150_F* and *pR_1-200_F* — and the counterpart promoterless plasmids — *p-R_1-50_F*, *p-R_1-100-F*, *p-R_1-150_F* and *p-R_1-200_F* —, they were obtained by removing 225, 175, 125 or 75 nucleotides, respectively, from the 3' end of *UPF* 5'UTR full sequence. HeLa cells were transfected with each of these deletional mutants or the counterpart negative and positive controls for this experiment — empty vector and full-length *UPF1* 5'UTR-containing vector, respectively — and co-transfected with β-galactosidase-encoding plasmid (pSV-β-Galactosidase control vector) as an internal control for transfection efficiency. Using luminometry assays, we assessed relative RLuc and FLuc expression levels from every transfected construct and observed relative FLuc expression levels from *pR_UPF1_F* and *p-R_UPF1_F*, i.e, *UPF1* 5'UTR-containing plasmids, either with or without promoter, respectively, were significantly greater than those from the counterpart empty vector, whereas those from the plasmids containing deletional mutants, either with or without promoter, were not (figure 3.19). In addition, relative FLuc expression levels from promoter-containing plasmids including at least the first 50 nucleotides of *UPF1* 5'UTR 5' end (*pR_1-50_F*, *pR_1-100_F*, *pR_1-150_F* and *pR_1-200_F*) were circa 2-fold those from the empty vector; however, these values were not statistically significant as they failed Student's *t*-test. Overall, these results suggest the entire sequence is required for cryptic promoter activity, nevertheless nucleotides 1–50 seem to be crucial for such activity. Deletions of segments of *UPF1* 5'UTR sequence are either disrupting the promoter itself or some enhancers required for the binding of transcriptions factors, thus abrogating the ability of this sequence to act as a transcription start site. On the other hand, RLuc expression levels measured from every construct with promoter that contains either full-length or mutant *UPF1* 5'UTR sequences are similar but approximately half of that measured from *pR_F*. This may be explained by some alternative splicing event that is disrupting full RLuc ORF and, hence, reducing its expression.

II.4. *UPF1* 5'UTR mediates cap-independent translation initiation

As shown before, *UPF1* 5'UTR sequence has cryptic promoter activity that maybe masking potential cap-independent translation activity. To avoid this situation, henceforth, all experiments were performed by transfecting cells with *in vitro* transcribed, capped and polyadenylated bicistronic mRNA. By using *in vitro* transcribed mRNA, we circumvent not only the occurrence of false-positive results derived from the cryptic promoter but also those arisen from alternative splicing events. Thus, the relative FLuc expression levels measured in cells transfected with each of the transcripts reflect the ability of the sequence cloned upstream FLuc AUG to mediate a cap-independent translation initiation mechanism. For that, we produced mRNA from the previously described plasmids — *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F* and *pR_UPF1_F* — and obtained the corresponding transcripts — *R_F*, *R_HBB_F*, *R_MYC_F*,

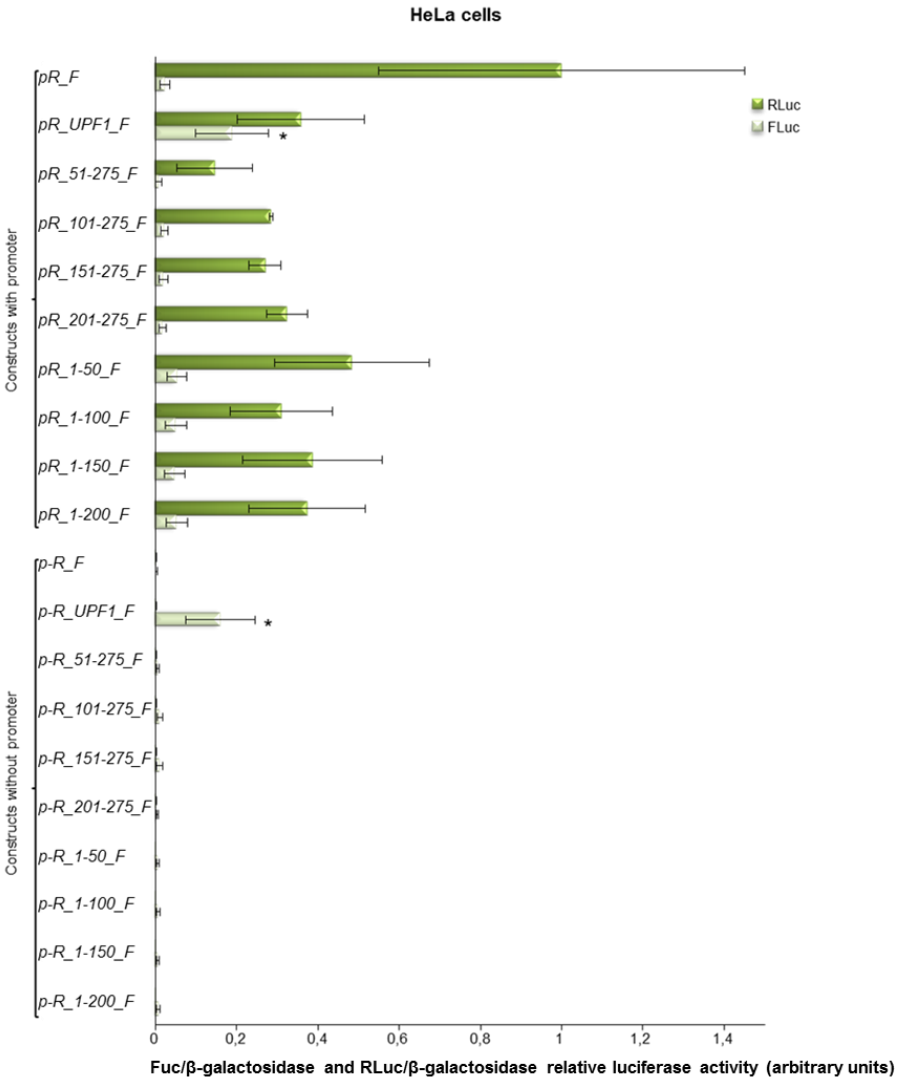


Figure 3.19 — Deletion of *UPF1* 5'UTR sequence disrupts cryptic promoter activity. HeLa cells were transfected with either promoter-containing or promoterless constructs (*pR_F*, *pR_UPF1_F*, *pR_51-275_F*, *pR_101-275_F*, *pR_151-275_F*, *pR_201-275_F*, *pR_1-50_F*, *pR_1-100_F*, *pR_1-150_F* and *pR_1-200_F*, or *p-R_F*, *p-R_UPF1_F*, *p-R_51-275_F*, *p-R_101-275_F*, *p-R_151-275_F*, *p-R_201-275_F*, *p-R_1-50_F*, *p-R_1-100_F*, *p-R_1-150_F* and *p-R_1-200_F*, respectively) depicted in figure 3.18. Cells were co-transfected with β-galactosidase-encoding plasmid (pSV-β-Galactosidase control vector) as an internal control. Dark green bars represent relative RLuc expression levels and light green bars indicate relative FLuc expression levels. Asterisks (*) indicate statistical significance in relation to the promoter-containing empty vector. *P<0.05

*R_EMCV*_F and *R_UPF1_F* (figure 3.20). HeLa, NCM460 and HCT116 cells were transfected with such transcripts for 4 h instead of the customary 24 h when cells are transfected with DNA, in order to preserve the integrity of the *in vitro* produced transcript. The relative FLuc expression from each transcript was assessed and compared to that from the empty construct, arbitrarily set to 1 (figure 3.21). In all tested cell lines, we observed a significant increase in relative FLuc

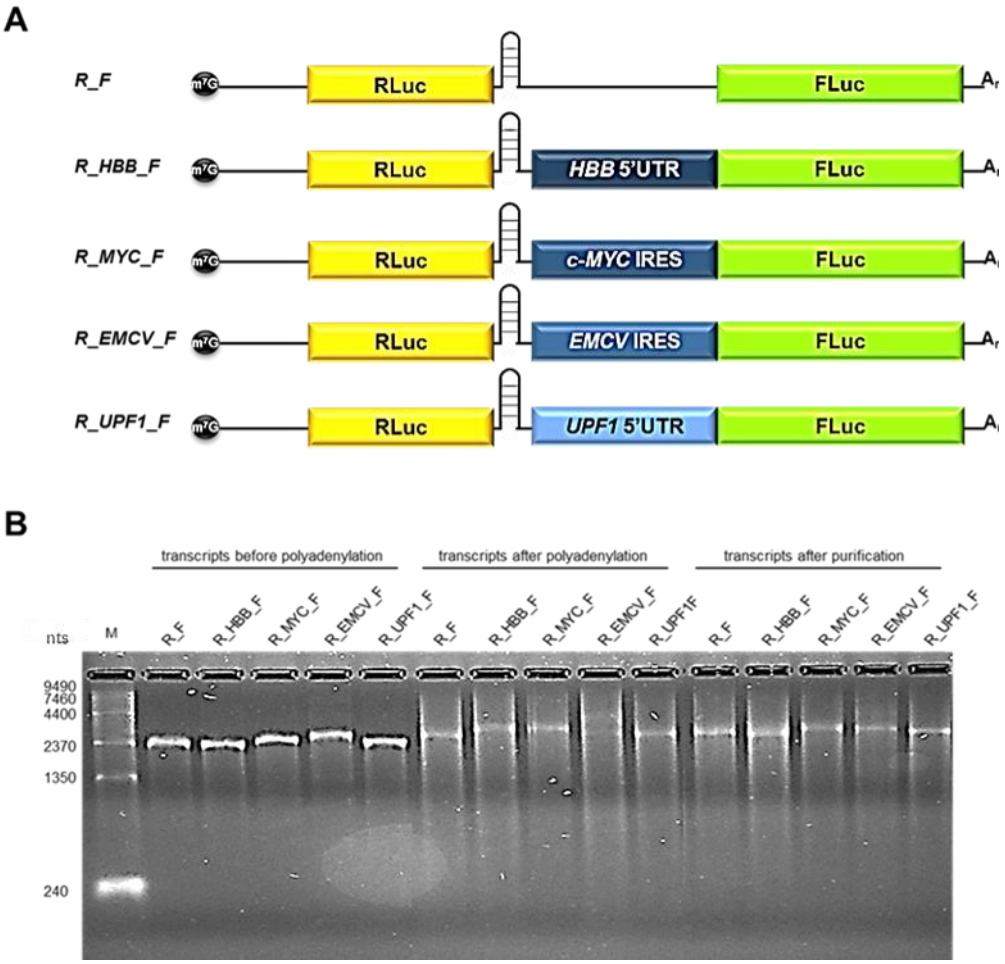


Figure 3.20 — *In vitro* transcribed mRNA used to detect *UPF1* 5'UTR-mediated cap-independent translation activity. (A) Schematic representation of the capped (m^7G) and polyadenylated (A_n) *in vitro* transcribed mRNA produced from the corresponding plasmids (cf. figure 3.14). RLuc is the *Renilla* luciferase cap-dependent-translated cistron (yellow box) and FLuc the firefly luciferase cap-independent-translated cistron (green box). Boxes in different shades of blue represent the different sequences cloned upstream FLuc AUG. *R_F* is the empty transcript and *R_HBB_F,* the human *β-globin* (*HBB*) 5'UTR-containing transcript — both are negative controls for cap-independent activity; *R_MYC_F,* the *c-MYC* IRES-containing transcript, is the cellular positive control for cap-independent activity; *R_EMCV_F,* the *EMCV* IRES-containing transcript, is the viral positive control for IRES activity; *R_UPF1_F,* the *UPF1* 5'UTR-containing transcript, contains the sequence under study. All transcripts contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation. (B) Denaturing agarose–formaldehyde gel showing the integrity of the produced mRNA. Expected transcript sizes before polyadenylation: 3004 nts (*R_F*); 3054 nts (*R_HBB_F*); 3344 nts (*R_MYC_F*); 3585 nts (R_*EMCV*_F); 3279 nts (*R_UPF1_F*). M: 0.24–9.5 kb RNA Ladder (Invitrogen); nts: weight, in bases, of RNA ladder fragments. Transcripts before polyadenylation: *in vitro* capped transcripts without poly(A) tail. Transcripts after polyadenylation: *in vitro* capped transcripts after addition of poly(A) tail. Transcripts after purification: *in vitro* capped and polyadenylated transcripts after RNA purification by phenol–chloroform extraction.

expression levels from the *UPF1* 5'UTR-containing transcript compared to those from *R_F* — 2.1-fold in HeLa cells, 2.4-fold in NCM460 cells and 2.5-fold in HCT116 cells. Also, the fold-

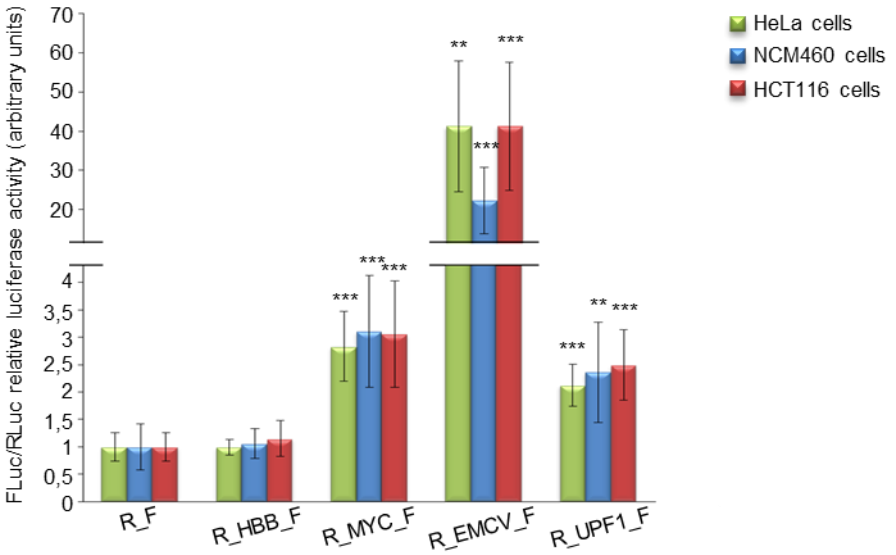


Figure 3.21 — *UPF1* 5'UTR mediates cap-independent translation initiation in HeLa, NCM460 and HCT116 cell lines. HeLa (green bars), NCM460 (blue bars), and HCT116 (red bars) cells were transfected with *in vitro* transcribed mRNA containing either *UPF1* 5'UTR (*R_UPF1_F*) or one of the controls used in the experiment: *R_F* (empty transcript), *R_HBB_F* (*HBB* 5'UTR-containing transcript), *R_MYC_F* (*c-Myc* IRES-containing transcript), or *R_EMCV_F* (*EMCV* IRES-containing transcript). Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. ***P*<0.01, ****P*<0.001

increase in relative FLuc expression levels from *R_UPF1_F* is similar to those from *R_MYC_F* in all tested cell lines — 2.8-fold in HeLa cells and 3.1-fold in both NCM460 and HCT116 cells. In addition, the relative FLuc expression levels from *R_EMCV_F* were 41.3-, 22.2- and 41.2-fold those from *R_F* in HeLa, NCM460 and HCT116 cells, respectively, whereas those from *R_HBB_F* were similar to those from the empty transcript. These results indicate, on the one hand that the system is reliable for the detection of cap-independent translation initiation activity as the used positive and negative controls behaved exactly according to what expected: *c-Myc* and *EMCV* IRES sequences are able to mediate translation of the downstream ORF in a cap-independent manner, whereas *HBB* 5'UTR is not. We also observed that the levels of relative FLuc expression mediated by *UPF1* 5'UTR are similar among cell lines, suggesting cap-independent translation initiation is neither tissue-specific nor differentially activated in normal versus cancer cells. The fact that relative FLuc expression levels from *R_UPF1_F* were similar to those from *R_MYC_F* strengthen our conclusion of *UPF1* 5'UTR being able to mediate cap-independent translation initiation in a manner similar to that of *c-Myc* IRES, a cellular sequence capable of driving cap-independent translation initiation. The much greater levels of relative FLuc expression from *R_EMCV_F* than from the transcripts containing cellular sequences corroborate the fact that viral IRES are much stronger and more active than the cellular ones.

II.5. Mutational analysis of *UPF1* 5'UTR reveals stem loop I and stem loop III are required for cap-independent translation initiation

After understanding that *UPF1* 5'UTR drives cap-independent translation initiation, we checked what the minimal required sequence for such activity might be. For that purpose, we performed a deletional analysis of the sequence by removing 50 nucleotides at a time either in a 5' to 3' direction or in a 3' to 5' direction, similar to the deletions performed previously to identify the sequence required for cryptic promoter activity. Using mFold software, we first performed an *in silico* analysis of the predicted secondary structures formed in the absence of each deleted sequence and compared it to the secondary structure predicted to the full-length sequence. We evaluated how the original structure would be affected by the deletions and, specifically whether the stem loops predicted to be formed by the full-length sequence were maintained or disrupted. Figures 3.22 and 3.23 give an overview of the secondary structures predicted to each case.

Deletion of nts 1–50 (figure 3.22.B) disrupts SLI but maintains SLII and III, although SLII does not maintain its original spatial conformation (figure 3.22.A). Deletion of the first 100 nts (figure 3.22.C) completely abolishes SLI and part of SLII, whereas SLIII is maintained as in the full-length sequence. Deletion of the first 150 nts is predicted to entirely abolish SLI and II but maintain SLIII intact as in the full-length prediction (figure 3.22.D). As for the deletion of the first 200 nts of *UPF1* 5'UTR, it abolishes SLI and III and disrupts SLIII (figure 3.22.E). These results suggest the formation of SLIII is well-maintained with a spatial conformation similar to that in the full-length sequence, whereas SLII formation as in the full-length sequence is wobbly. Regarding deletions performed in a 3' to 5' direction, the obtained prediction is as follows: by deleting nts 51–275 (figure 3.23.B), the whole structure is disrupted and only a portion of SLI is formed; deletion of nts 101–275 (figure 3.23.C) allows the formation of SLI with a spatial conformation similar to that predicted for the full-length sequence (figure 3.23.A); by deleting the last 125 nts of *UPF* 5'UTR (figure 3.23.D), SLI is maintained and SLII is formed; deletion of the last 75 nts leads to the complete formation of SLI and II, but SLIII is disrupted. These results suggest that SLI is well-preserved and maintains a spatial conformation similar to that of the full-length sequence, whereas, again, SLII is uneven and does not always reflects the prediction for the full-length structure. Overall, these results indicate that the formation of SLI and SLIII is well-kept, suggesting these structures may be of consequence for cap-independent translation activity by playing a role in the direct recruitment of the ribosome to the vicinity of the initiator AUG. To complement this deletional analysis, and because deleting part of the sequence may lead to the formation of abnormal structures, we disrupted the predicted stem loops of the wild-type sequence by point mutating groups of 2 or 3 nts within the loops of interest. Figure 3.24 (B, C and D) shows the predicted secondary structure for mutated *UPF1* 5'UTR and compared it to the prediction obtained for the wild-type sequence (figure 3.24.A). Mutations at nts 39–40 (GC→AA) lead to disruption of SLI, whereas formation of SLII and III is maintained, nevertheless SLII spatial conformation is altered compared to that of the wild-type (figure

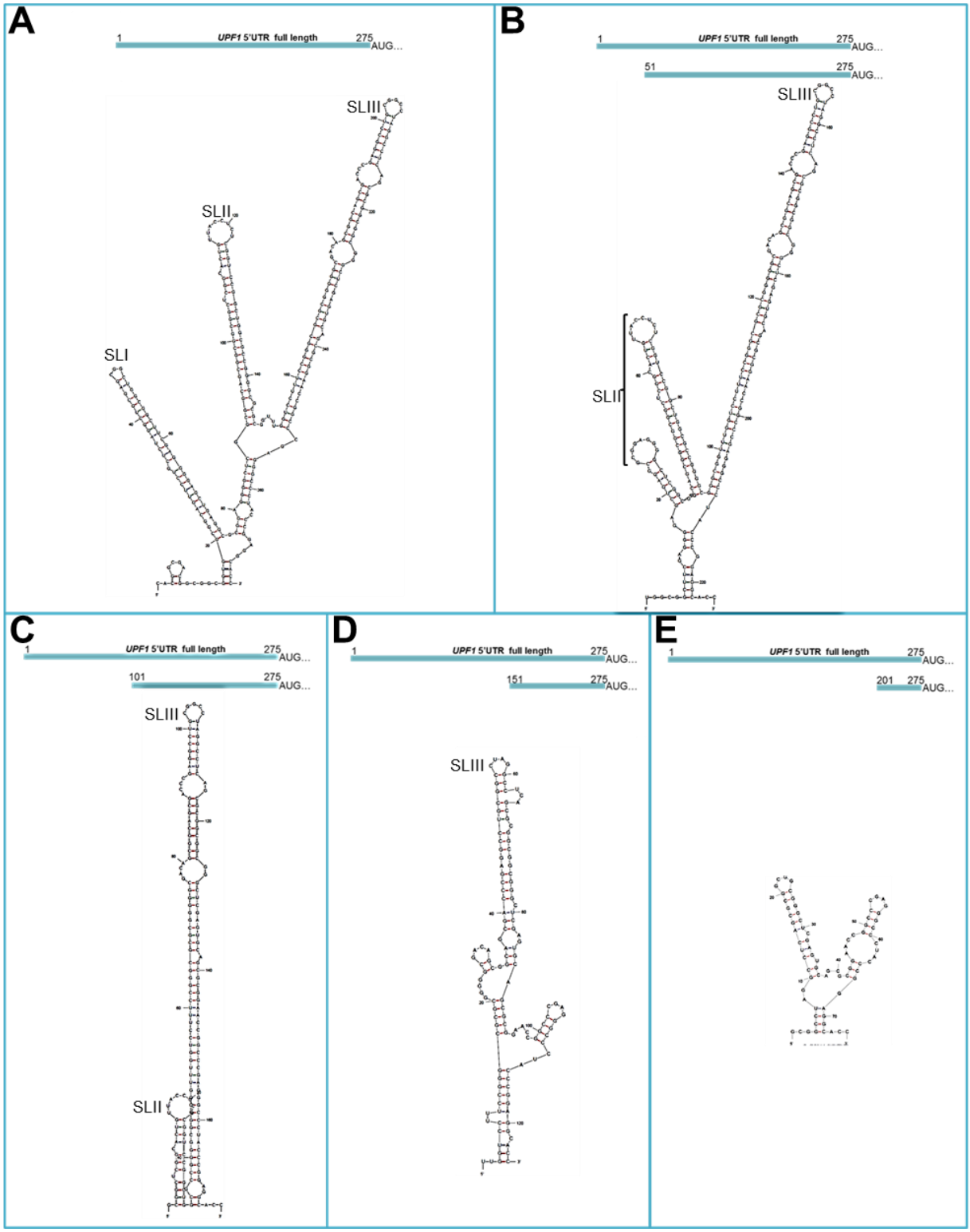


Figure 3.22 — *In silico* predicted secondary structures of *UPF1* 5'UTR with 5' end sequential deletions. (A) Full-length *UPF1* 5'UTR; (B) nucleotides (nts) 51–275 of *UPF1* 5'UTR; (C) nts 101–275 of *UPF1* 5'UTR; (D) nts 151–275 of *UPF1* 5'UTR; (E) nts 200–275 of *UPF1* 5'UTR. Stem loop (SL) I, II and III in each structure represent the predicted formation of stem loops, according to those identified in the full-length sequence. Predictions were obtained with mFold software (http://mfold.rna.albany.edu/?q=mfold) using default parameters. Blue lines indicate the length of the sequence compared to full-length.

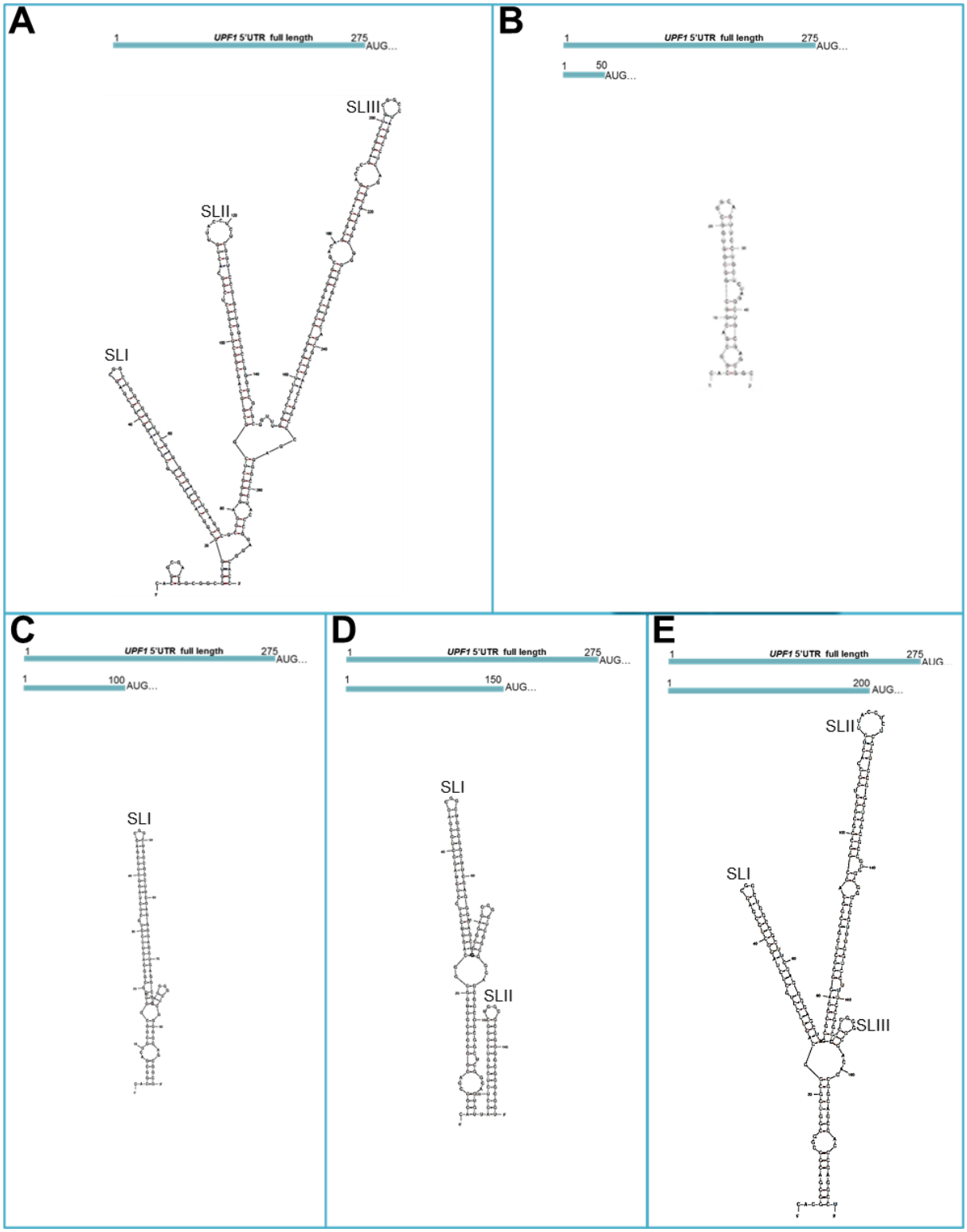


Figure 3.23 — *In silico* predicted secondary structures of *UPF1* 5'UTR with 3' end sequential deletions. (A) Full-length *UPF1* 5'UTR; (B) nucleotides (nts) 1–50 of *UPF1* 5'UTR; (C) nts 1–100 of *UPF1* 5'UTR; (D) nts 1–150 of *UPF1* 5'UTR; (E) nts 1–200 of *UPF1* 5'UTR. Stem loop (SL) I, II and III in each structure represent the predicted formation of stem loops, according to those identified in the full-length sequence. Predictions were obtained with mFold software (http://mfold.rna.albany.edu/?q=mfold) using default parameters. Blue lines indicate the length of the sequence compared to full-length.

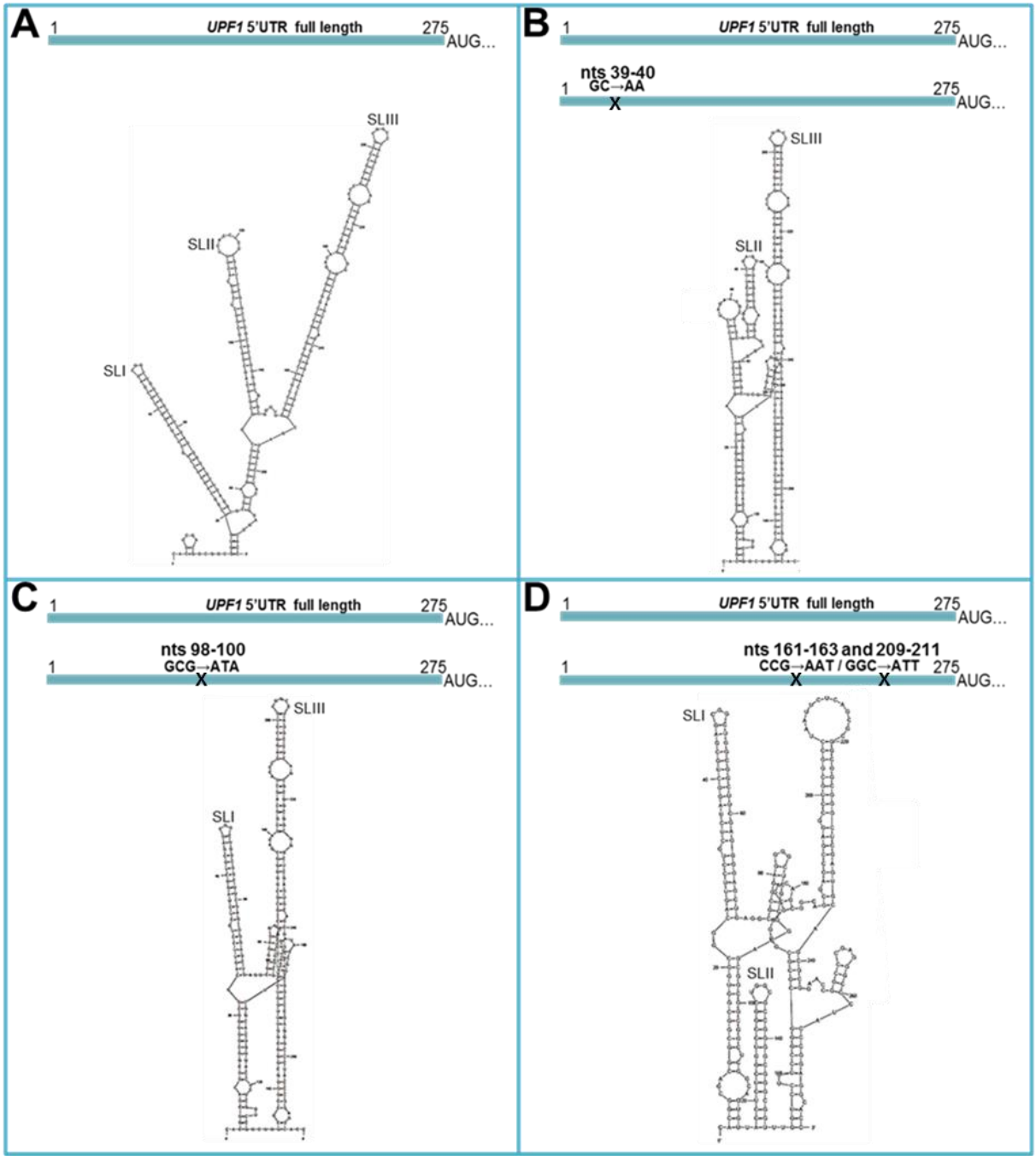


Figure 3.24 — *In silico* predicted secondary structures of *UPF1* 5'UTR with point mutations. (A) Full-length *UPF1* 5'UTR; (B) *UPF1* 5'UTR mutated at nucleotides (nts) 39–40 (GC→AA); (C) *UPF1* 5'UTR mutated at nts 98–100 (GCG→ATA); (D) *UPF1* 5'UTR mutated at nts 161–163 (CCG→AAT) and 209–211 (GGC→ATT). Stem loop (SL) I, II and III in each structure represent the predicted formation of stem loops, according to those identified in the full-length sequence. Predictions were obtained with mFold software (http://mfold.rna.albany.edu/?q=mfold) using default parameters. Black "X" indicates the relative positions of the mutations compared to the wild-type sequence.

3.14.B). By mutating nts 98–100 (GCC→ATA), we observed SLII is disrupted, but SLI and III are maintained in a spatial conformation similar to that in the wild-type sequence (figure 3.24.C). As for mutations at nts 161–163 (CCG→AAT) and nts 209–211 (GGC→ATT), they have an impact on the formation of SLIII and completely disrupt the formation of SLIII as it is predicted for the wild-type sequence, and formation of both SLI and II is maintained in a spatial conformation similar to the one predicted for the wild-type sequence (figure 3.24.D). These results indicate that the formation of SLI and III is preserved even if the remaining structure is damaged, indicating these two structures may have a relevant role in regulating cap-independent translation activity mediated by *UPF1* 5'UTR. Contrariwise, SLII seems not to be effortlessly maintained as its spatial conformation alters when the rest of the structure (cf. figure 3.13) is damaged, suggesting it may not be essential for cap-independent translation activity. These results corroborate the outcome of the previous deletional analysis. We evaluated the effect of all these alterations (deletions and point mutations) on *UPF1* 5'UTR cap-independent translation activity. For that, we produced *in vitro* capped and polyadenylated transcripts containing each of the desired alteration (figure 3.25) and confirmed the integrity of such transcripts in a denaturing agarose**–**formaldehyde gel (figure 3.26). We transfected HeLa cells with each of the transcripts described above (figure 3.25) for 4 h and used luminometry assays to assess relative FLuc expression levels from each of the transcripts. We compared the cap-independent translation of FLuc from the transcripts containing either deletional or point mutations to that from the empty or the *UPF1* 5'UTR-containing transcripts. When transfecting cells with transcripts containing deletional mutations (figure 3.27.A), we observed that both R_*151-275*_F and R_*1-100*_F reached almost full activity of *R_UPF1_F* (91% and 99%, respectively), whereas the relative FLuc expression levels from the other analysed deletional mutant transcripts were similar to those from *R_F*. This means that the first 100 and the last 125 nts are required and each of them alone is suffcient for the cap-independent translation initiation mediated by *UPF1* 5'UTR. However, when cells were transfected with transcripts containing sequences spanning from nt 101 to nt 150 of *UPF1* 5'UTR but not the sequences comprising nts 1–100 and 151–275 together, i.e., R_*51-275*_F, R_*101-275*_F, R_*1-150*_F, and R_*1-200*_F, the relative FLuc expression levels from these transcripts were drastically reduced to levels similar to those from *R_F*. R_*1-50*_F and R_*201-275*_F — transcripts containing only half the sequences required for cap-independent translation activity also retrieved relative FLuc expression levels similar to those from *R_F*. By transfecting cells with transcripts containing point-mutated *UPF1* 5'UTR sequences (figure 3.27.B), we saw mutations disrupting either SLI or III completely abolish cap-independent translation activity, whereas mutations disrupting SLII maintained cap-independent activity (84% of that from *R_UPF1_F*). Altogether, these results show the first 100 nts (corresponding to the predicted SLI) and the last 125 nts (corresponding to SLIII) are essential for cap-independent activity to occur. However, the central portion of the sequence (corresponding to SLII) seems to inhibit that activity when only nts 1–100 or nts 151–275 are present.

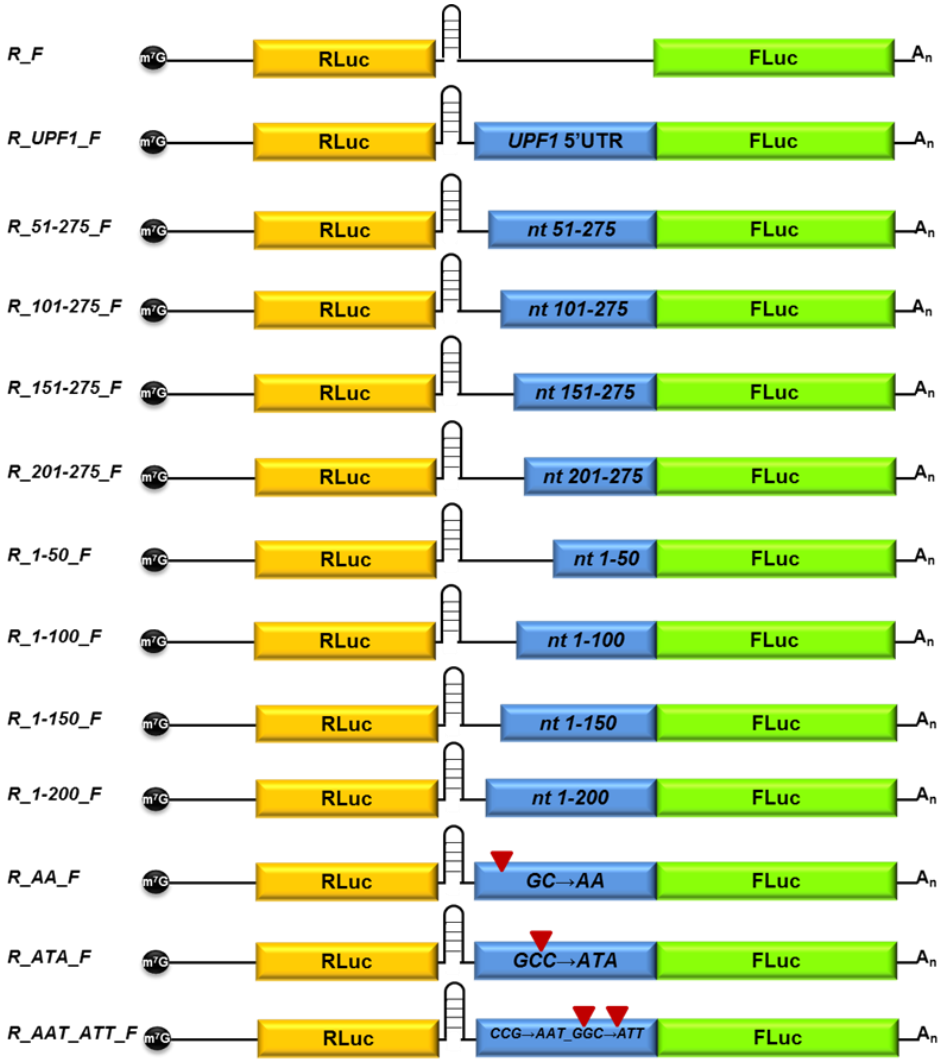


Figure 3.25 — Schematic representation of the *in vitro* transcribed, capped (m^7G) and polyadenylated (A_n) mRNA used to identify the minimal sequence of *UPF1* 5'UTR required for cap-independent translation activity. RLuc is the *Renilla* luciferase cap-dependent translated cistron (yellow box) and FLuc the firefly cap-independent translated cistron (green box). Blue boxes of different sizes represent the different lengths of the deletional sequences cloned upstream FLuc AUG. Red triangles indicate the relative position of the mutations. *R_F* is the empty transcript; *R_UPF1_F*, the full-length *UPF1* 5'UTR-containing transcript; R_51-275_F, the *UPF1* 5'UTR nucleotides (nts) 51–275-containing transcript; *R_101-275_F*, the *UPF1* 5'UTR nts 101–275-containing transcript; *R_151-275_F*, the *UPF1* 5'UTR nts 151–275-containing transcript; *R_201-275_F*, the *UPF1* 5'UTR nts 201–275-containing transcript; *R_1-50_F*, the *UPF1* 5'UTR nts 1–50-containing transcript; *R-1-100_F*, the *UPF1* 5'UTR nts 1–100-containing transcript; *R-1-150_F*, the *UPF1* 5'UTR nts 1–150-containing transcript; *R-1-200_F*, the *UPF1* 5'UTR nts 1–200-containing transcript; *R_AA_F*, the *UPF1* 5'UTR mutated at nts 39–40-containing transcript; *R_ATA_F*, the *UPF1* 5'UTR mutated at nts 98–100-containing transcript; *R_AAT_ATT_F*, the *UPF1* 5'UTR mutated at nts 161–163 and 209–211-containing transcript. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

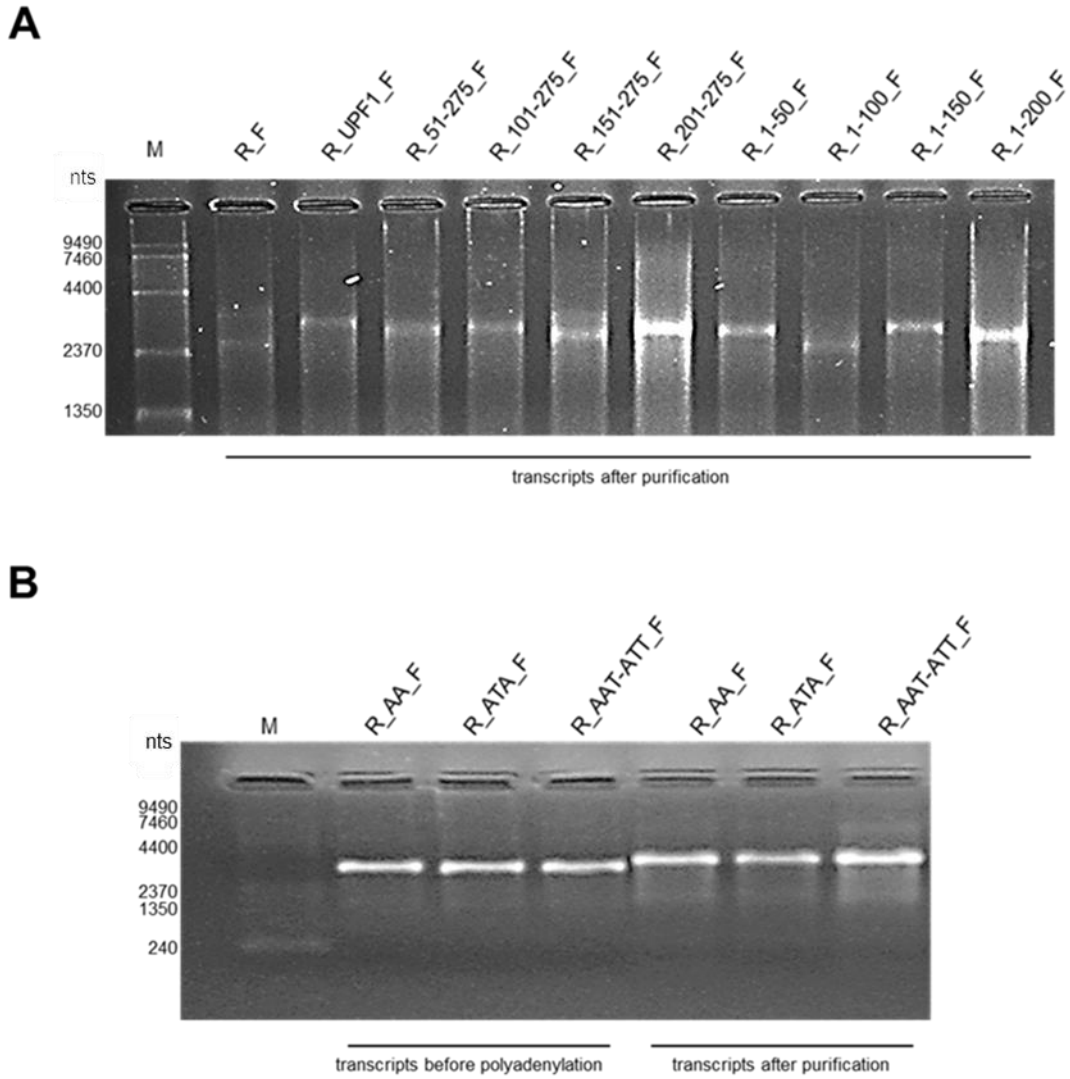


Figure 3.26 — Denaturing agarose–formaldehyde gels showing the integrity of the *in vitro* transcribed, capped (m^7G) and polyadenylated (A_n) mRNA used to identify the minimal sequence of *UPF1* 5'UTR required for cap-independent translation activity. *R_F* is the empty transcript (3004bp); *R_UPF1_F*, the full-length *UPF1* 5'UTR-containing transcript; R_51-275_F, the *UPF1* 5'UTR nucleotides (nt) 51–275-containing transcript; *R_101-275_F*, the *UPF1* 5'UTR nt 101–275-containing transcript; *R_151-275_F*, the *UPF1* 5'UTR nt 151–275-containing transcript; *R_201-275_F*, the *UPF1* 5'UTR nt 201–275-containing transcript; *R_1-50_F*, the *UPF1* 5'UTR nt 1–50-containing transcript; *R-1-100_F*, the *UPF1* 5'UTR nt 1–100-containing transcript; *R-1-150_F*, the *UPF1* 5'UTR nt 1–150-containing transcript; *R-1-200_F*, the *UPF1* 5'UTR nt 1–200-containing transcript; *R_AA_F*, the *UPF1* 5'UTR mutated at nts 39–40-containing transcript; *R_ATA_F*, the *UPF1* 5'UTR mutated at nts 98–100-containing transcript; *R_AAT_ATT_F*, the *UPF1* 5'UTR mutated at nts 161–163 and 209–211-containing transcript. M: 0.24–9.5 Kb RNA Ladder (Invitrogen); nts: weight, in bases of RNA ladder bands. Transcripts before polyadenylation: *in vitro* capped transcripts without poly(A) tail. Transcripts after purification: *in vitro* capped and polyadenylated transcripts after RNA extraction by phenol–chloroform purification.

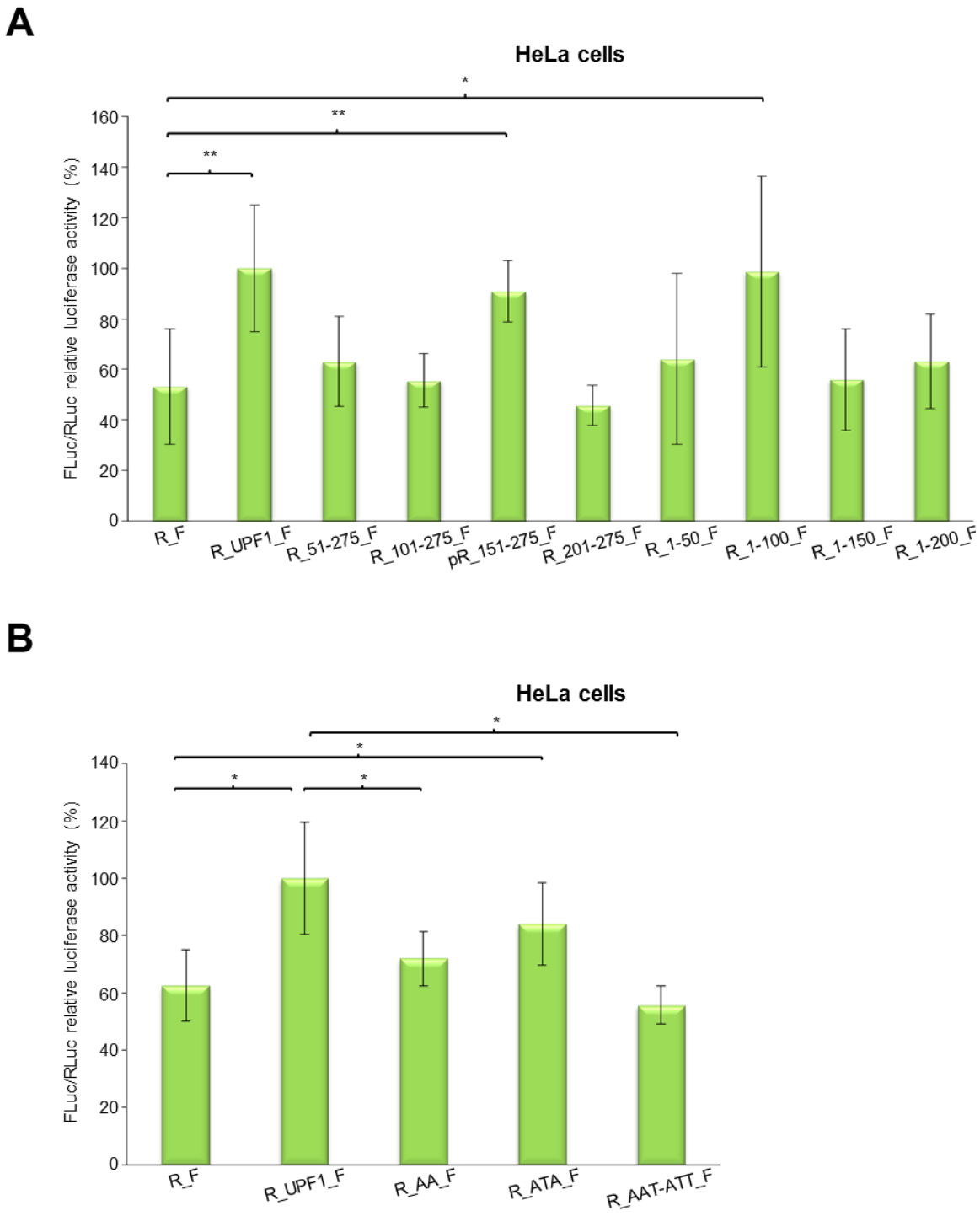


Figure 3.27 — Stem loop I (nts 1–100) and stem loop III (nts 151–275) are required for *UPF1* 5'UTR cap-independent translation activity. (A) Cap-independent translation activity from deletional mutant transcripts as a percentage of that from full-length *UPF1* 5'UTR-containing transcript. (B) Cap-independent activity from point-mutated transcripts as a percentage of that from full-length *UPF1* 5'UTR-containing transcript. The presented results are the outcome of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the indicated transcript. *P<0.05, **P<0.01

When the central sequence is present, both extremities must be present, in order to drive cap-independent translation, otherwise one of those external sequences alone is not enough to overcome the inhibition by the central sequence. That is why both *R_151-275_F* and *R_1-100_F* transcripts allow cap-independent translation activity (because they contain one of the required extremities, but not the central sequence), but *R_51-275_F*, *R_101-275_F*, *R_1-150_F*, and *R_1-200_F* mRNA do not (because they contain either only the 5' or the 3' portion, together with the central sequence). Furthermore, disrupting only SLI (*R_AA_F*) or III (*R_AAT-ATT_F*) leads to a loss of cap-independent translation activity, because, although one of the SL required for full cap-independent translation is present, SLII is also present and inhibits full cap-independent activity, which indicates both sequences, corresponding to SLI and III, must be present. Disrupting only SLII (*R_ATA_F*) does not affect cap-independent activity, as the relative FLuc expression levels from the mutated transcript are similar to those from *R_UPF1_F*. These results corroborate the predictions by mFold software, in which the formation of SL I and III is maintained with a spatial conformation similar to that in the wild-type sequence, whereas SLII formation is less stable. Experimental verification indicates formation of SLI and III is of the utmost importance for cap-independent translation initiation mediated by *UPF1* 5'UTR, whereas formation of SLII is not required to rescue the wild-type phenotype. When SLI or III alone are present, the formation of SLII impairs cap-independent translation activity mediated by each of those loops, suggesting both SLI and III must work together in mediating cap-independent translation initiation, possibly by arranging in a spatial conformation prone to the direct recruitment of the ribosome subunits to the vicinity of the main AUG and subsequence peptide synthesis.

II.6. *UPF1* 5'UTR cap-independent translation activity is maintained under stress conditions

After having identified cap-independent translation activity within *UPF1* 5'UTR and knowing which segments of the sequence are required for its proper functioning, we investigated whether such activity can be maintained under conditions that impair cap-dependent translation initiation. We subjected HeLa, NCM460 and HCT116 cells to several external stimuli and evaluated the relative FLuc expression levels from *R_UPF1_F* in cells under stress conditions and compared them to those from the same transcript in cells in control conditions. The external stimuli applied on cells were: knock-down of eIF4E, the cap-binding protein crucial for cap-dependent translation; hypoxia, which reduces overall protein synthesis as a protective measure for cell metabolism; rapamycin, which targets mTOR kinase and blocks its ability to phosphorylate 4E-BP and S6K proteins; and thapsigargin, which induces endoplasmic reticulum stress and eventually leads to an unfolded protein response.

In order to test the effect of knocking down eIF4E protein in *UPF1* 5'UTR-mediated translation, we transfected HeLa cells with either siRNA against GFP (control conditions) or

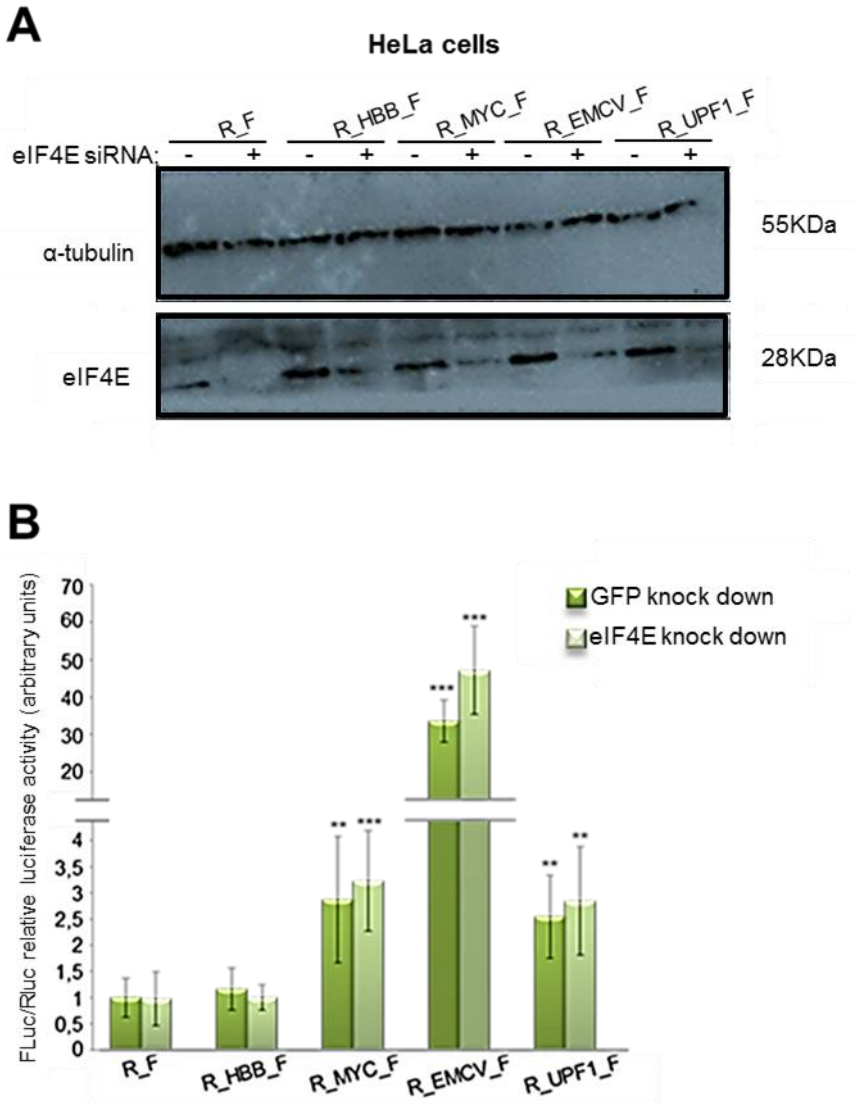


Figure 3.28 — *UPF1* 5'UTR cap-independent translation activity is maintained after knock-down of eIF4E protein. HeLa cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA after knocking down eIF4E protein. (A) Western blot against eIF4E shows knock-down efficiency. (-) indicates GFP siRNA transfection and (+) indicates eIF4E siRNA transfection. α-tubulin was used as a loading control for the amount of protein. (B) Relative luciferase activity measured from *R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F and *R_UPF1_F* transcripts. Dark green bars indicate conditions of GFP siRNA cellular treatment and light green bars indicate conditions of eIF4E siRNA transfection. Asterisks (*) indicate statistical significance in relation to the empty counterpart in control conditions (GFP siRNA). **P*<0.05, ***P*<0.01, ****P*<0.001.

against eIF4E and, 40 h posttransfection, we transfected the same cells with *in vitro* transcribed, capped and polyadenylated mRNA — *R_F*, *R_HBB_F*, *R_MYC_F*, *R_EMCV_F* and *R_UPF1_F*. Western Blot analysis against eIF4E confirmed the profuse reduction on eIF4E availability as observed in figure 3.28.A. Regarding the relative FLuc expression levels from cells in control conditions, we observe a significant increase in those levels from transcripts

containing either *c-Myc* or *EMCV* IRES, or *UPF1* 5'UTR, compared to those measured from cells transfected with *R_F* or *R_HBB_F* (figure 3.28.B). Such levels were similar to those previously observed (figure 3.21). As for the relative FLuc expression levels from cells transfected with the same transcripts in conditions of eIF4E knock-down, we observed they are maintained in levels similar to the counterpart levels in control conditions (figure 3.28.B). These results suggest the cap-independent translation activity mediated by *UPF1* 5'UTR is maintained in conditions of low eIF4E protein expression levels, reflecting the former's ability to function under stress conditions, independent of the eIF4E cap-binding protein.

As for the induction of hypoxic conditions, we subjected HeLa, NMC460 and HCT116 cells to a cobalt chloride ($CoCl_2$) treatment. We transfected cells with *in vitro* transcribed, capped and polyadenylated mRNA and, 2 h posttransfection, supplemented the medium with 200 μM $CoCl_2$ for 6 h (a suitable interval considering the half-time of luciferase protein is 5.3 h). This drug is a chemical hypoxia-mimicking agent and its functioning is observed in all treated cells as it induced the accumulation of hypoxia inducible factor 1 α (HIF1α) protein, as confirmed by Western blot against this protein in every tested cell line (figure 3.29.A, C and E). The relative FLuc expression levels from transcripts containing either *UPF1* 5'UTR or each of the positive and negative controls for cap-independent translation activity, in control conditions (cells treated with vehicle, H_2O) (figure 3.29.B, D and F), they were similar to previously obtained results for every tested cell line (figure 3.21). Regarding *UPF1* 5'UTR-containing transcript, it induced a 2-fold increase in relative FLuc expression levels compared to the empty transcript, whose expression levels were arbitrarily set to 1. Likewise, the relative FLuc expression levels from such transcripts under hypoxic conditions were similar to those observed under control conditions, i.e., approximately 2-fold the relative FLuc expression levels measured from the empty transcript in control conditions. Such results are similar to each of the tested cell lines — HeLa, NCM460 and HCT116. These results indicate *UPF1* 5'UTR is able to mediate cap-independent translation initiation under stress conditions in a manner similar to that used by *c-Myc* IRES in every tested cell line, as the relative FLuc expression from the transcript containing the latter is similar to that from R_*UPF1*_F.

We also tested to what extent cap-independent translation mediated by *UPF1* 5'UTR is affected by mTOR kinase-impaired activity. Thus, we transfected HeLa, NCM460 and HCT116 cells with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, *R_EMCV_F* and *R_UPF1_F*) and, 2 h posttransfection, treated cells with 80 nM rapamycin for 6 h. Western blot analysis of all tested cell lines confirms phosphorylated S6K protein is not detected in cells treated with rapamycin (figure 3.30.A, C, E). Relative FLuc expression levels were assessed by luminometry assays in all cell lines transfected with each transcript. Again, in control conditions, the relative FLuc expression levels measured from each transcript (figure 3.30. B, D, F) were similar to previous results, i.e.relative FLuc expression levels from positive controls, *R_MYC_F* and *R_EMCV_F*, and from *R_UPF1_F* were significantly greater than those from *R_F*, arbitrarily set to 1, whereas those from *R_HBB_F* — negative control for cap-

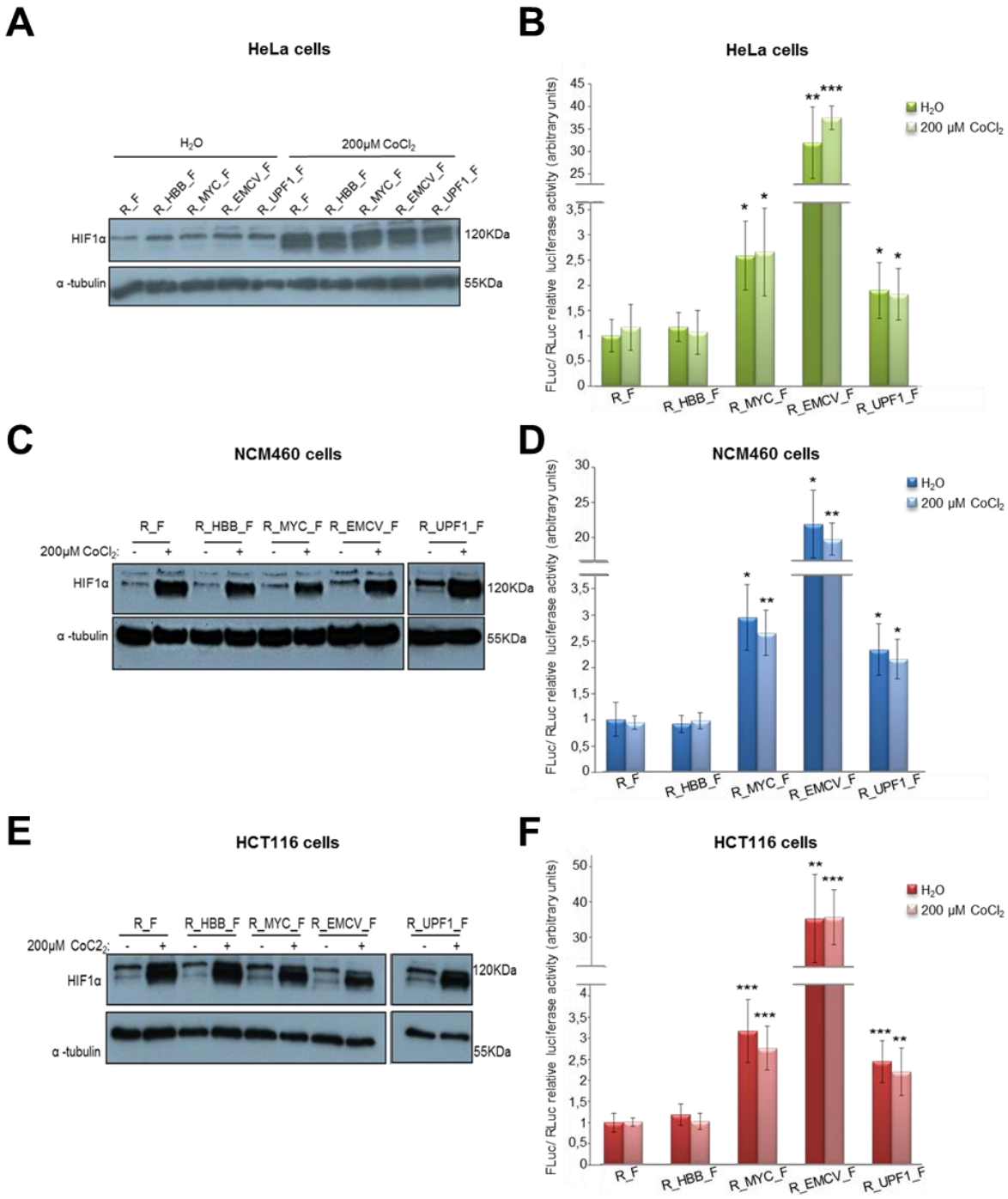


Figure 3.29 — *UPF1* 5'UTR cap-independent translation activity is maintained under hypoxic conditions in HeLa, NCM460 and HCT116 cells. HeLa (A, B), NCM460 (C, D) and HCT116 (E, F) cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F, *R_UPF1_F*) and then subjected to a 6-h-treatment with 200 μM of $CoCl_2$ 2 h posttransfection. (A, C, E) Western blot against HIF1α, whose accumulation reflects a cellular hypoxic status. (-) indicates treatment with H_2O and (+) indicates treatment with 200 μM of $CoCl_2$. α-tubulin was used as a loading control for the amount of protein. (B, D, F) Relative FLuc expression levels from each transcript. Dark green, dark blue and red bars represent relative FLuc expression in control conditions; light green, light blue and pink bars represent relative FLuc expression under hypoxia. Asterisks (*) indicate statistical significance in relation to the empty counterpart transcript in control conditions. **P*<0.05, ***P*<0.01, ****P*<0.001.

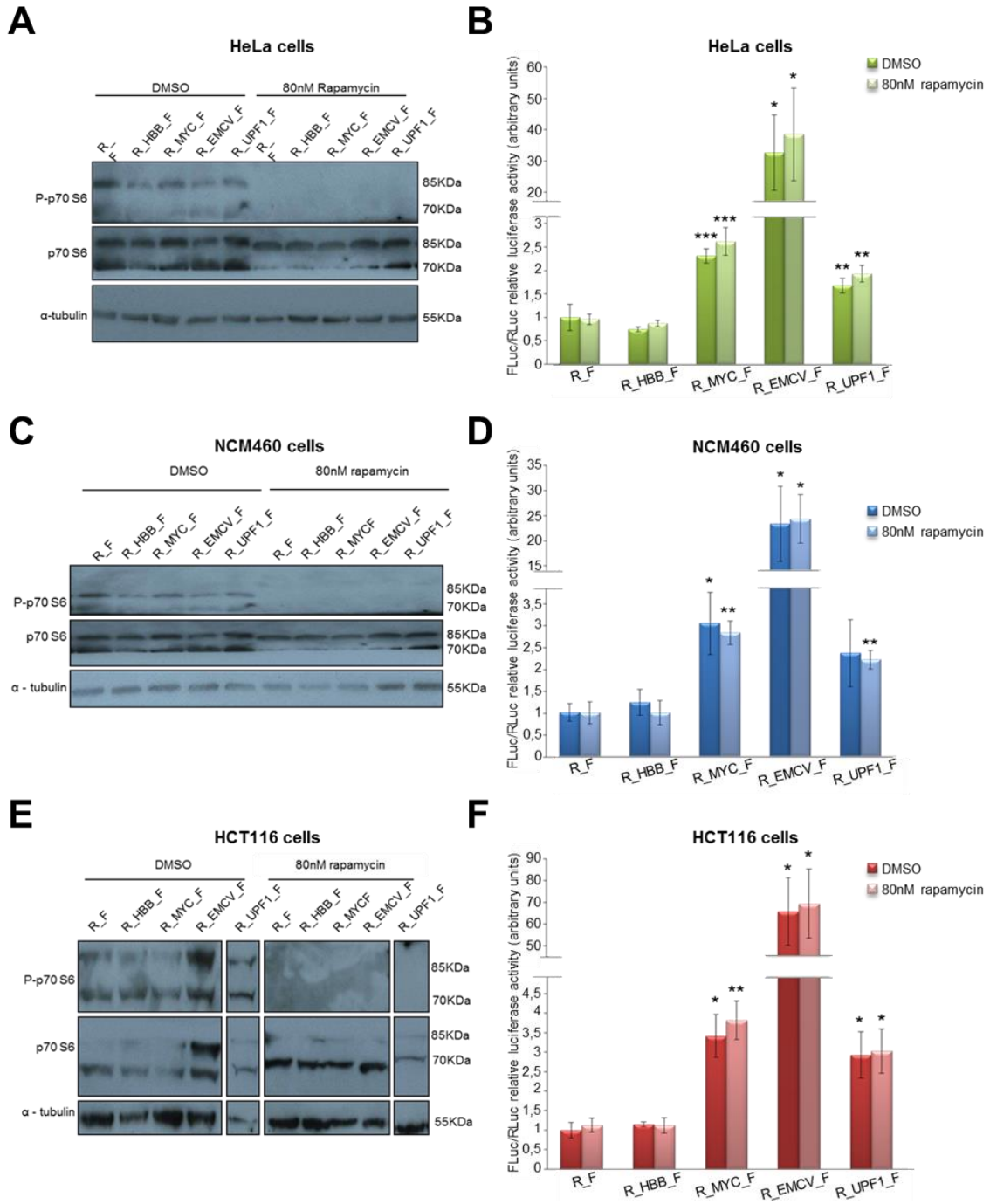


Figure 3.30 — *UPF1* 5'UTR-mediated cap-independent translation is maintained under conditions impairing mTOR kinase activity. HeLa (A, B), NCM460 (C, D) and HCT116 (E, F) cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F, *R_UPF1_F*) and then subjected to a 6-h-treatment with 80 nM rapamycin 2 h posttransfection. (A, C, E) Western blot against phosphorylated and non-phosphorylated S6K protein — absence of the former indicates mTOR kinase activity on its downstream targets is blocked. "DMSO" indicates cells in control conditions and "80 nM rapamycin" indicates cells treated with the drug. α-tubulin was used as a loading control for the amount of protein. (B, D, F) Relative FLuc expression levels from each transcript. Dark green, dark blue and red bars represent relative FLuc expression in control conditions; light green, light blue and pink bars represent relative FLuc expression under treatment with rapamycin. Asterisks (*) indicate statistical significance in relation to the empty counterpart transcript in control conditions. **P*<0.05, ***P*<0.01, ****P*<0.001.

independent translation activity — were similar to those from the empty transcript (figure 3.21). Under stress conditions, in every tested cell lines, we observed that relative FLuc expression levels from each transcript are similar to the corresponding expression levels in control conditions (figure 3.30.B, D, E). Again, relative FLuc expression levels from both positive control and *UPF1* 5'UTR-containing transcript are significantly greater than those from *R_F*, whereas those from *R_HBB_F* are similar to those from the latter — in the case of *UPF1* 5'UTR, approximately 2-fold the expression levels from *R_F*. These results suggest *UPF1* 5'UTR is able to drive cap-independent translation in conditions that compromise the regular functioning of mTOR pathway.

We also tested the effect of endoplasmic reticulum (ER) stress in the cap-independent translation activity mediated by *UPF1* 5'UTR. For that, we transfected HeLa, NCM460 and HCT116 cells with *in vitro* transcribed, capped and polyadenylated mRNA containing either *UPF1* 5'UTR (*R_UPF1_F*) or each of the positive (*R_MYC_F* and *R_EMCV_F*) and negative (*R_HBB_F*) controls for cap-independent translation activity, and, 2 h posttransfection, we treated cells with 1 µM thapsigargin to induce ER stress, or DMSO (vehicle). Western blot analysis of each of the tested cell lines showed an increase in phosphorylated eIF2α protein in cells treated with the drug compared to those treated with DMSO (figure 3.31.A, C, E), indicating cells are experiencing ER stress. In cells treated with DMSO (control conditions), the relative FLuc expression levels from *R_UPF1_F* (figure 3.31.B, D, F) were significantly greater than those from *R_F* in every tested cell lines, as previously seen (figure 3.21). Relative FLuc expression levels from *R_MYC_F* and *R_EMCV_F* were also significantly greater than those from the empty transcript, whereas those from *R_HBB_F* were similar to those from *R_F*, corroborating our previous results that *UPF1* 5'UTR is capable of mediating cap-independent translation initiation in a manner similar to that of *c-Myc* IRES. In cells treated with 1 µM thapsigargin, we observed that the relative expression levels from transcripts capable of mediating cap-independent translation initiation were similar to those obtained in control conditions, suggesting they can maintain the ability of driving cap-independent under ER stress-inducing conditions. The results obtained under ER stress are in line with the results obtained under the other tested external cellular stimuli — knock-down of eIF4E, hypoxia, and mTOR kinase-impaired activity. Altogether, these results suggest *UPF1* 5'UTR is able to mediate cap-independent translation initiation under stress conditions, when canonical translation initiation is impaired by an external cellular stimuli.

II.7. *UPF1* 5'UTR can mediate cap-independent translation in monocistronic transcripts lacking the cap structure

To further confirm that *UPF1* 5'UTR is able to mediate cap-independent translation initiation, we produced *in vitro* monocistronic polyadenylated transcripts encoding FLuc and lacking the cap structure required for mediating cap-dependent translation initiation. Since transcripts with

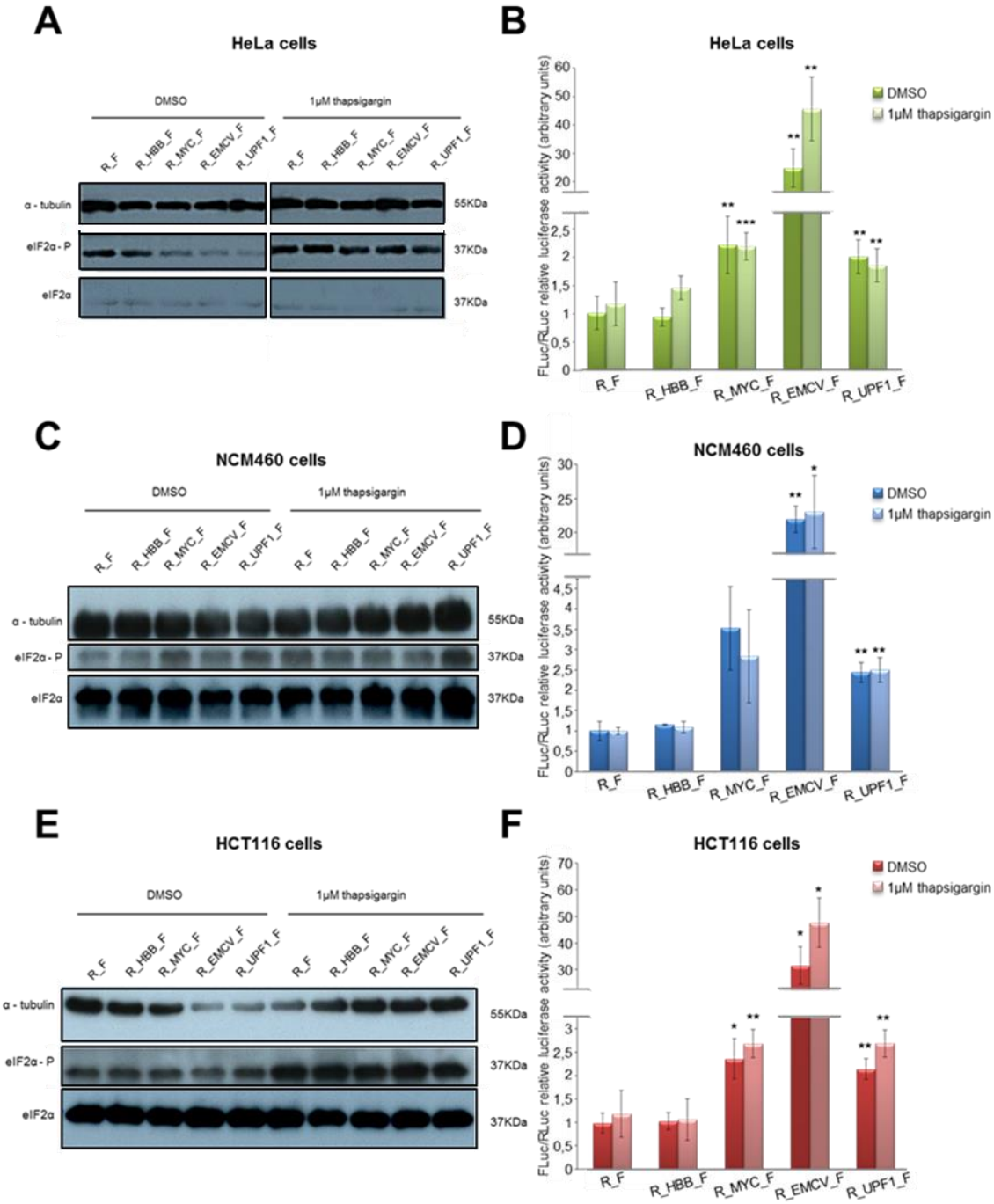


Figure 3.31 — *UPF1* 5'UTR cap-independent translation activity is maintained under endoplasmic reticulum (ER) stress conditions. HeLa (A, B), NCM460 (C, D) and HCT116 (E, F) cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F, *R_UPF1_F*) and then subjected to a 6-h-treatment with 1 μM thapsigargin 2 h posttransfection. (A, C, E) Western blot against phosphorylated and non-phosphorylated eIF2α protein, whose increased expression in treated cells reflects ER stress. "DMSO" indicates cells in control conditions; "1 μM thapsigargin" indicates cells treated with the drug. α-tubulin was used as a loading control for the amount of protein. (B, D, F) Relative FLuc expression levels from each transcript. Dark green, dark blue and red bars represent relative FLuc expression in control conditions; light green, light blue and pink bars represent relative FLuc expression under ER stress. Asterisks (*) indicate statistical significance in relation to the empty counterpart transcript in control conditions. **P*<0.05, ***P*<0.01, ****P*<0.001.

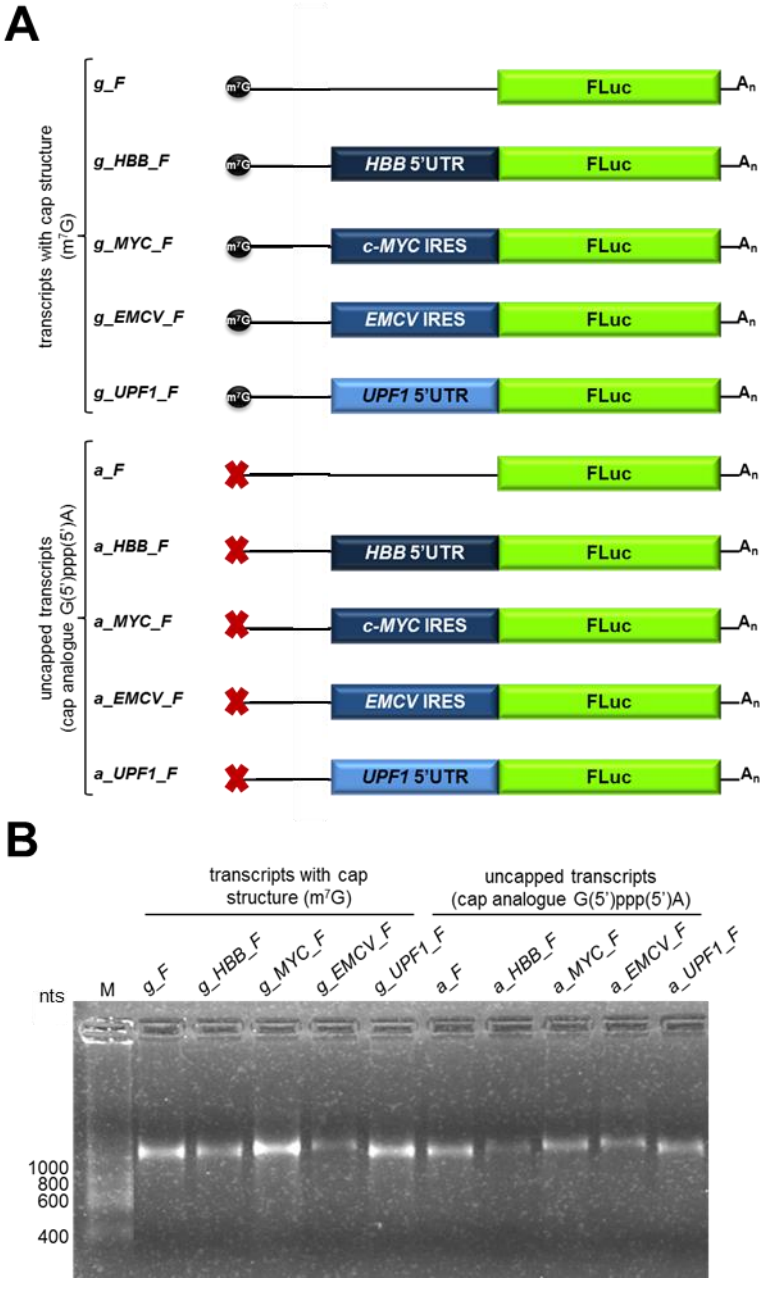


Figure 3.32 — *In vitro* transcribed and polyadenylated monocistronic transcripts used to evaluate *UPF1* 5'UTR cap-independent translation activity. (A) Schematic representation of the *in vitro* transcribed and polyadenylated (A_n) monocistronic transcripts. *g_F* (empty transcript), *g_HBB_F* (*HBB* 5'UTR-containing transcript), *g_MYC_F* (*c-Myc* IRES-containing transcript), *g_EMCV_F* (*EMCV* IRES-containing transcript) and *g_UPF1_F* (*UPF1* 5'UTR-containing transcript) are the capped (m^7G, black circles) transcripts, and *a_F*, *a_HBB_F*, *a_MYC_F*, *a_EMCV_F* and *a_UPF1_F* are the counterpart uncapped [cap analogue G(5')ppp(5')A, red crosses] transcripts. FLuc is the firefly luciferase-enconding cistron (green box). Boxes in different shades of blue represent different sequences cloned upstream FLuc AUG. (B) Denaturing agarose–formaldehyde gel showing the integrity of the *in vitro* transcribed and polyadenylated monocistronic transcripts. M: RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific); nts: weight, in bases, of RNA ladder bands.

unprotected 5' ends are unstable and prone to degradation, we used a cap structure analogue unable to mediate cap-dependent translation [G(5')ppp(5')A, by New England Biolabs] to protect transcripts from degradation. Thus, we produced two sets of *in vitro* transcribed and polyadenylated monocistronic mRNA — capped (m^7G) mRNA, or uncapped [cap structure analogue G(5')ppp(5')A] mRNA (figure 3.32.A) and checked their integrity (figure 3.32.B). We transfected HeLa cells with each of the capped (5'G-capped) or uncapped (5'A-capped) *in vitro* transcribed and polyadenylated monocistronic transcripts (empty transcript, *HBB* 5'UTR-containing transcript, *c-Myc* IRES-containing transcript, *EMCV* IRES-containing transcript and *UPF1* 5'UTR-containing transcript) and co-transfected them with with β-galactosidase-encoding plasmid (pSV-β-Galactosidase control vector), an internal control for transfection. Four hours after transfection, we assessed relative FLuc expression levels of each of the 5'A-capped transcript to its 5'G-capped counterpart, using luminometry assays (figure 3.33). We observed a statistically significant 8.2-fold increase in relative FLuc expression levels from 5'A-capped *UPF1* 5'UTR-containing transcript compared to the counterpart empty transcript, suggesting *UPF1* 5'UTR is able to mediate translation initiation in the absence of the cap structure.

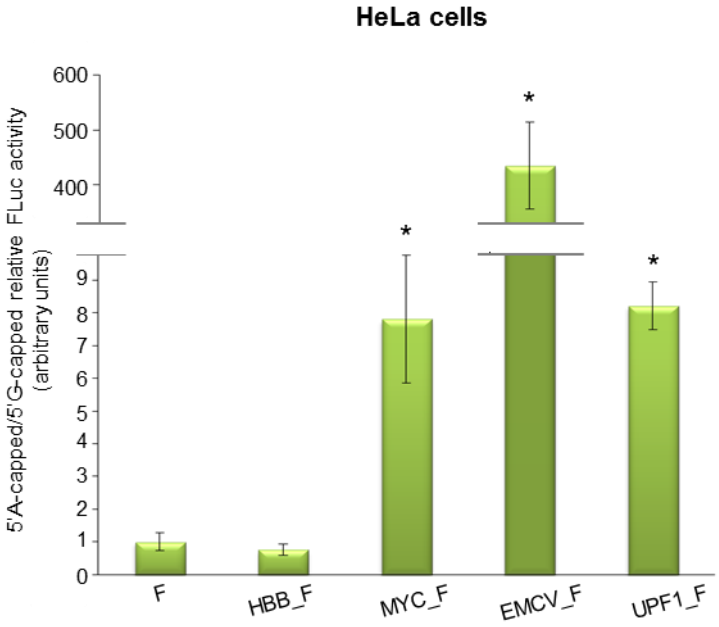


Figure 3.33 — *UPF1* 5'UTR mediates cap-independent translation initiation in monocistronic transcripts lacking cap structure. HeLa cells were transfected with *in vitro* transcribed monocistronic mRNA, either capped (5'G-capped, m^7G) or uncapped [5'A-capped, cap analogue G(5')ppp(5')A], containing either *UPF1* 5'UTR (*UPF1_F*) or one of the controls used in the experiment: *F* (empty transcript), *HBB_F* (*HBB* 5'UTR-containing transcript), *MYC_F* (*c-Myc* IRES-containing transcript), or *EMCV_F* (*EMCV* IRES-containing transcript). Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty transcripts. *P<0.05

Furthermore, relative FLuc expression levels from both 5'A-capped *c-Myc* and *EMCV* IRES-containing transcripts were significantly greater than those from the counterpart empty transcript, i.e., 7.8- and 435-fold, respectively, confirming the ability of such sequences to mediate cap-independent translation initiation. Conversely, relative FLuc expression levels from 5'A-capped *HBB* 5'UTR-containing transcript were similar to those from the empty transcript, demonstrating that this sequence is not able to drive cap-independent translation initiation, as expected. Of note, relative FLuc expression levels from both 5' A-capped *UPF1* 5'UTR- and *c-Myc* IRES-containing transcripts are similar. All in all, these results confirm our previous observations that *UPF1* 5'UTR is able to mediate cap-independent translation initiation of a downstream open reading frame.

III. Human *AGO1* 5'UTR mediates an eIF4G-enhanced but cap-independent mechanism of translation initiation

Our *In silico* analysis indicated AGO1 as a putative candidate to be translated via an alternative mechanism of translation initiation. Experimental validation is therefore required to either confirm or rule out the ability of its 5'UTR to drive cap-independent translation initiation or an alternative cap-dependent mechanism that works in non-canonical conditions.

III.1. *In silico* analysis of *AGO1* 5'UTR characteristics

A thorough analysis of *AGO1* 5'UTR revealed it is 213 nts long and contains an upstream AUG at position -5 from the initiation codon (figure 3.34.A). It contains an overall GC content of 72.3%, distributed over regions of very high GC content (up to 87.5% at regions adjacent to 5' terminus) and other with lower content (minimum 53.8% at regions adjacent to 3' terminus (figure 3.34.B), and tendency to fold into structures of predicted stability — ΔG=-111.95kcal/mol (predicted as before, figure 3.34.C). According to this prediction, the secondary structure formed within *AGO1* 5'UTR includes four stem loops (SL) — I, II, III and IV. SLI corresponds to the region containing the highest GC content, suggesting a great stability for the predicted stem loop. The formation of such structures may allow ribosome recruitment to the vicinity of the main AUG and thus may help in facilitating cap-independent translation initiation via internal entry of the ribosome (cap-independent mechanism). On the other hand, these structures may impair the regular scanning of the 5'UTR and, hence, promote mechanism of ribosome shunting that force the ribosome to bypass them and reach the AUG in a non-canonical way. Furthermore, an upstream AUG within this untranslated region, suggests the possibility of a uORF regulating AGO1 protein expression.

Sequence conservation throughout evolution may also provide some clues on the importance of the sequence under analysis and how it may be involved in the process of translation initiation. For that purpose, we compared human *AGO1* 5'UTR to those of other mammalian species to evaluate how conserved among species this sequence is and also to what extent the formation of the predicted stem loops is maintained among different species. Using Bioedit software, we compared the sequence of *AGO1* 5'UTR from different mammals (human, gorilla, white-cheeked gibbon and mouse; figure 3.35.A). All four sequences are almost perfectly aligned from nt 64 onwards. The first 64 nts are extremely conserved between human and white-cheeked gibbon sequences, but only the last 12 nts of this segment are present in the other species. Furthermore, the predicted consensus secondary structure of the set of aligned sequences obtained using RNAalifold webserver (figure 3.35.B) is composed of 4 stem loops exactly as in the mFold prediction (figure 3.34.C). According to the results obtained, formation of SL II, III and IV seems to be conserved throughout evolution, whereas formation of SLI is the less conserved. All in all, these results indicate that the sequence comprising the last 149 nts is highly conserved and predictably able to be folded into stable stem loops, suggesting their being involved in a conserved mechanism of gene expression regulation.

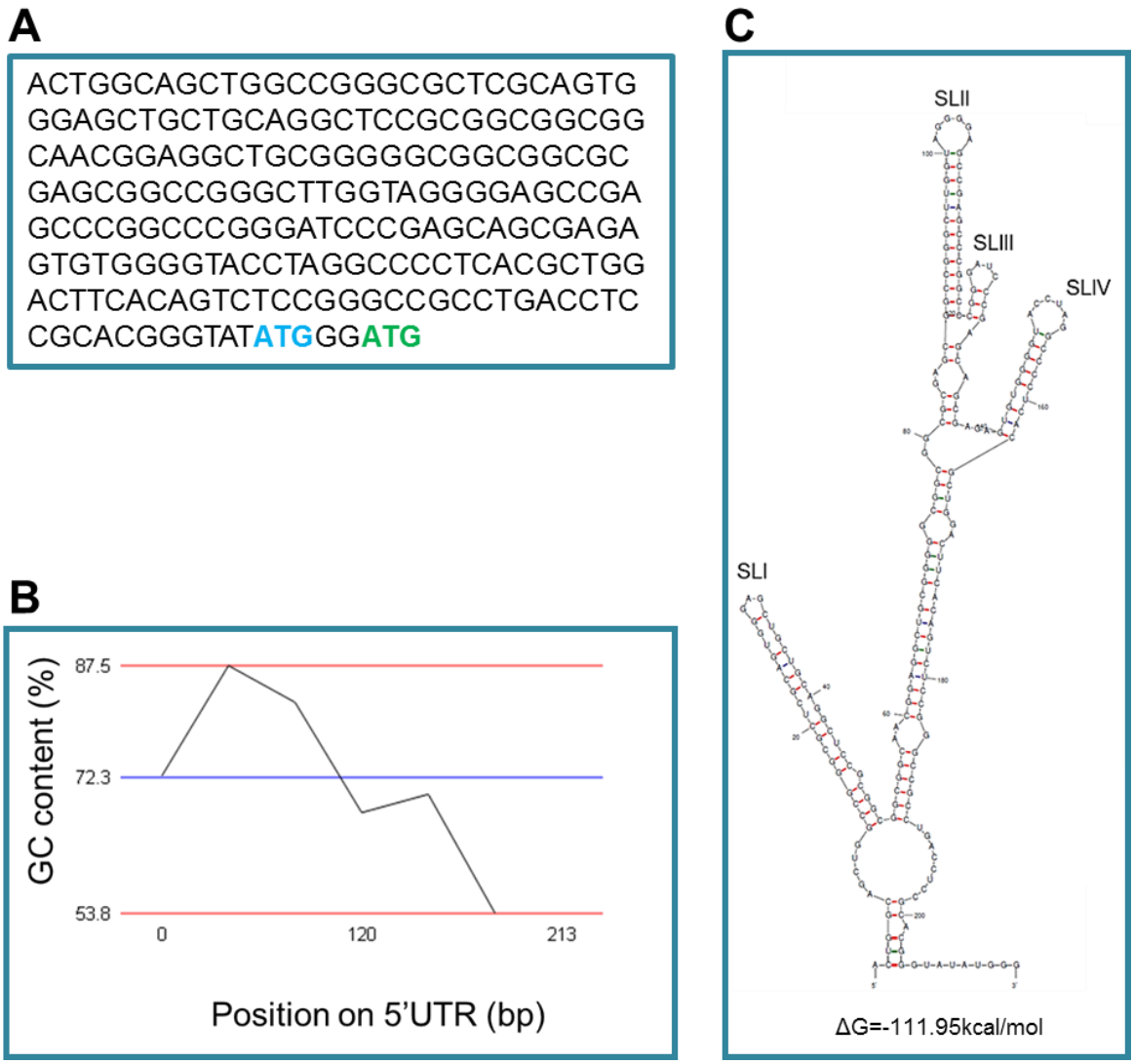


Figure 3.34 — Multiple features of human *AGO1* 5'UTR predicted *in silico*. (A) Nucleotide sequence of human *AGO1* 5'UTR used in this work. **ATG** represents an upstream open reading frame at position -5. **ATG** is the AGO1 translation initiation codon. (B) Calculation of the GC content (%) of different regions of *AGO1* 5'UTR (http://www.endmemo.com). The content (average 72.3%) ranges from 53.8% to 87.5% and the highest percentages tend to localise towards the 5' end of the 5'UTR. (C) RNA secondary structure of AGO*1* 5'UTR predicted by mFold software (http://mfold.rna.albany.edu/?q=mfold). Based on this prediction, four stem loops (SL) — I, II, III and IV — are formed within *AGO1* 5'UTR and the structure is very stable (ΔG = -111.95kJ/mol)**.**

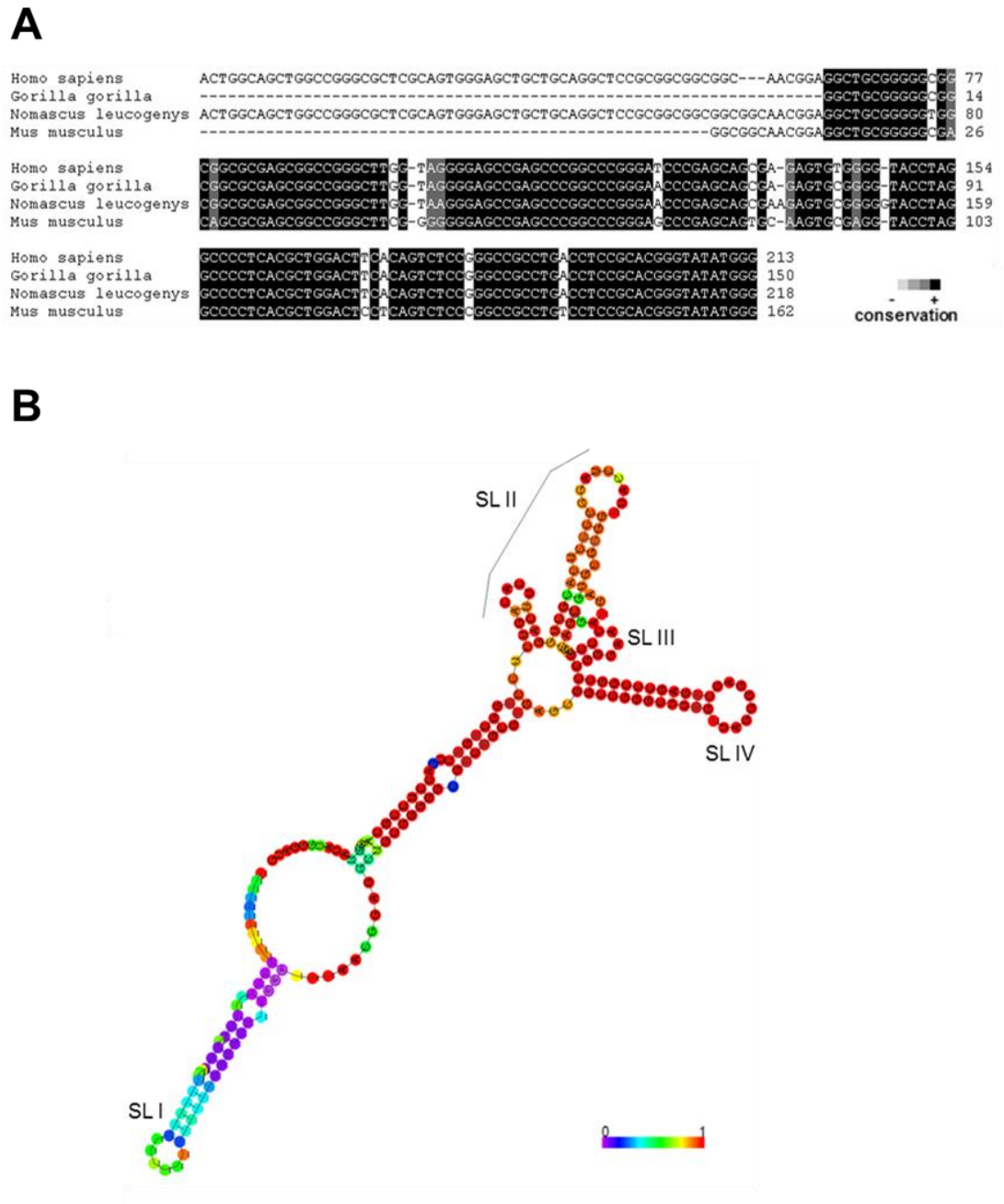


Figure 3.35 — Conservation of *AGO1* 5'UTR sequence among mammalian species. (A) Sequence alignment of *AGO1* 5'UTR among human, gorilla, white-cheeked gibbon and mouse obtained using Bioedit software (http://www.bioinformatics.org/sms2/color_align_cons.html). Grey scale indicates the degree of conservation among species for each nucleotide. White (-) indicates least conserved; black (+) indicates most conserved. (B) Predicted consensus secondary structure of the set of aligned sequences using RNAalifold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi). Coloured scale indicates the degree of conservation of the predicted secondary structure. Purple (0) indicates no secondary structure conservation; Red (1) indicates full secondary structure conservation.

Noteworthy, besides the aforementioned characteristics found in the analysed sequence, we have also found four G-Quadruplex predicted motifs within this sequence (table 3.2). The putative formation of such structures may influence the translation of the downstream ORF, either by inhibiting canonical cap-dependent translation initiation, or by stimulating non-canonical cap-independent translation mechanisms.

Table 3.2 — Quadruplex-forming G-rich sequences (QGRS) found in human *AGO1* 5'UTR*

Position	Length	QGRS	G-Score
4	26	GGCAGCTGGCCGGGCGCTCGCAGTGG	14
52	29	GGCGGCAACGGAGGCTGCGGGGGCGGCGG	21
88	17	GGCCGGGCTTGGTAGGG	21
144	25	GGGGTACCTAGGCCCCTCACGCTGG	10

*The underlined GG represent those putatively involved in the formation of G-quadruplex structures. Position designates the first nucleotide of the QGRS sequences. The putative G-quadruplexes are identified using the motif $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$, where x is the number of guanine tetrads in the G-quadruplex, and y1, y2, y3 are the length of gaps, i.e., the length of the loops connecting the guanine tetrads. The motif consists of four equal length groups of guanines separated by arbitrary nucleotide sequences with at least two tetrads (x≥2) and maximum length of 30 bases. The maximum length of 30 bases restricts G-groups to a maximum size of 6. G-score is a classification attributed by the software that evaluates a QGRS for its likelihood to form a stable G-quadruplex. Higher scoring sequences will make better candidates for G-quadruplex. The scoring method considers the following principles: shorter loops are more common than longer loops; G-quadruplexes tend to have loops roughly equal in size; the greater the number of guanine tetrads the more stable the quadruplex. The highest possible G-score, using the default maximum QGRS length of 30, is 105. (http://bioinformatics.ramapo.edu/QGRS/index.php)

All these features suggest that a putative non-canonical mechanism driven by *AGO1* 5'UTR can be the answer to explain the observed altered expression of AGO1 protein. In this regard, we have conducted experiments to test this hypothesis, as shown below.

III.2. *AGO1* 5'UTR drives FLuc expression in a bicistronic context

AGO1 5'UTR was cloned in the same bicistronic vector (figure 3.36) used for assessing *UPF1* 5'UTR cap-independent translation activity (cf. section 3.2.2). Also, the used positive and negative control sequences to evaluate putative cap-independent translation initiation were the same as previously, i.e., *HBB* 5'UTR as the negative control, *c-Myc* IRES sequence as the positive cellular control, and *EMCV* IRES sequence as the positive viral sequence (figure 3.36). We transfected HeLa cells with each of the plasmids mentioned above and performed luminometry assays to measure relative FLuc activity from each construct. The obtained results

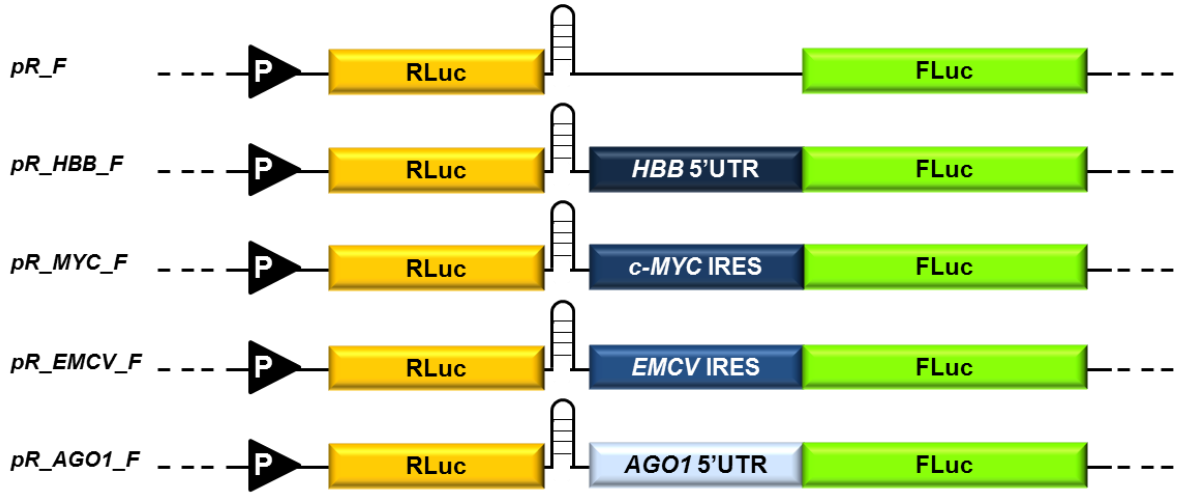


Figure 3.36 — Schematic representation of the constructs used to check whether *AGO1* 5'UTR is able to drive cap-independent translation initiation. RLuc is the *Renilla* luciferase cap-dependent-translated cistron (yellow box) and FLuc the firefly cap-independent-translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Boxes in different shades of blue represent the different sequences cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_HBB_F,* the human β-globin (*HBB*) 5'UTR-containing vector, is the negative control for cap-independent activity; *pR_MYC_F,* the *c-Myc* IRES-containing vector, is the cellular positive control for cap-independent activity; *pR_EMCV_F,* the *EMCV* IRES-containing vector, is the viral positive control for IRES activity; *pR_AGO1_F,* the *AGO1* 5'UTR-containing vector, is the sequence under study. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

are depicted in figure 3.37. Relative FLuc expression levels from *pR_AGO1_F* were 2.8-fold those from the empty plasmid, *pR_F*. As for the relative FLuc expression levels from *pR_HBB_F*, they were similar to those from *pR_F*, whereas those from *pR_MYC_F* and *pR_EMCV_F* were significantly greater than those from the empty plasmid — 5.8- and 13.4-fold, respectively (figure 3.37). These results indicate both positive controls are driving FLuc expression via a cap-independent mechanism, as expected, whereas the negative control is not. From this experiment, we can also conclude *AGO1* 5'UTR is able to drive FLuc expression in a bicistronic context, which suggests a putative non-canonical mechanism of translation initiation may be responsible for such expression.

III.3. FLuc expression driven by *AGO1* 5'UTR is not a consequence of either alternative splicing or cryptic promoter activity

The observed significant increase in relative FLuc expression levels from *pR_AGO1_F* compared to the counterpart empty vector may not be the outcome of a non-canonical mechanism of translation initiation mediated by *AGO1* 5'UTR, but, instead, the result of an event leading to false-positive occurrences, e.g.: cryptic promoters or alternative splicing.

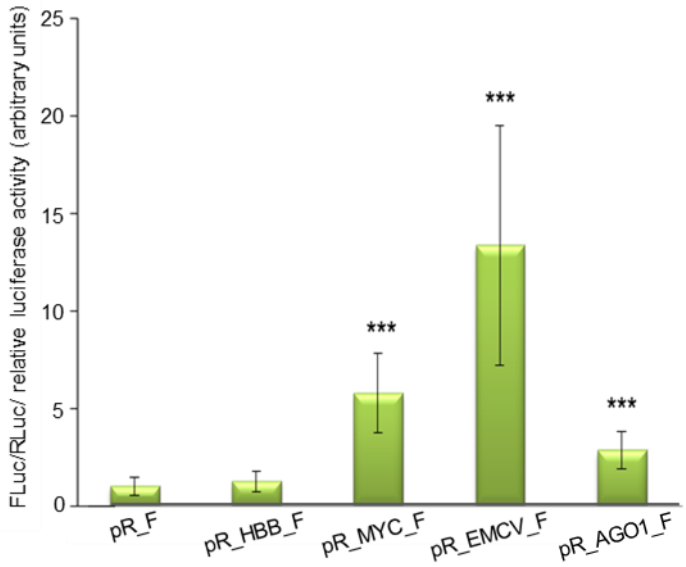


Figure 3.37 — Expression of FLuc reporter protein is mediated by *AGO1* 5'UTR in a bicistronic context. HeLa cells were transfected either with *AGO1* 5'UTR-containing plasmid (*pR_AGO1_F*) or with one of the controls used in the experiment: *pR_F* (empty plasmid), *pR_HBB_F* (*HBB* 5'UTR-containing plasmid), *pR_MYC_F* (*c-Myc* IRES-containing plasmid), or *pR_EMCV_F* (*EMCV* IRES-containing plasmid). Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. *P<0.05, **P<0.01, ***P<0.001

In order to rule out the presence of putative cryptic promoter sequences within *AGO1* 5'UTR sequence, we produced a promoterless plasmid containing *AGO1* 5'UTR and one containing *MLH1* 5'UTR, a positive control for the presence of cryptic promoter sequences (figure 3.38). We transfected HeLa cells with each of the constructs depicted in figure 3.38 and measured relative RLuc and FLuc expression levels (figure 3.39). Regarding relative RLuc expression, its levels were expected to be much greater in constructs containing the SV40 promoter than in those it has been removed. Accordingly, RLuc expression levels from promoterless constructs are actually virtually inexistent. *MLH1* 5'UTR-containing plasmid is the one presenting the least transfection efficiency in HeLa cells, as the relative RLuc expression levels measured from cells transfected with such construct are significantly lower than those from the empty plasmid (figure 3.39). Relative FLuc expression levels from *pR_MLH1_F* were also significantly greater than those from the empty vector, as expected, due to the presence of the promoter (Ito *et al.*, 1999; Arita *et al.*, 2003). Accordingly, the relative FLuc expression levels from promoterless *MLH1* 5'UTR-containing plasmid were also significantly greater than those from p-R_F, the empty promoterless plasmid. Contrariwise, promoterless p-R_AGO1_F did not express FLuc reporter protein. This result suggests that *AGO1* 5'UTR sequence does not contain any sequence capable of originating monocistronic mRNA.

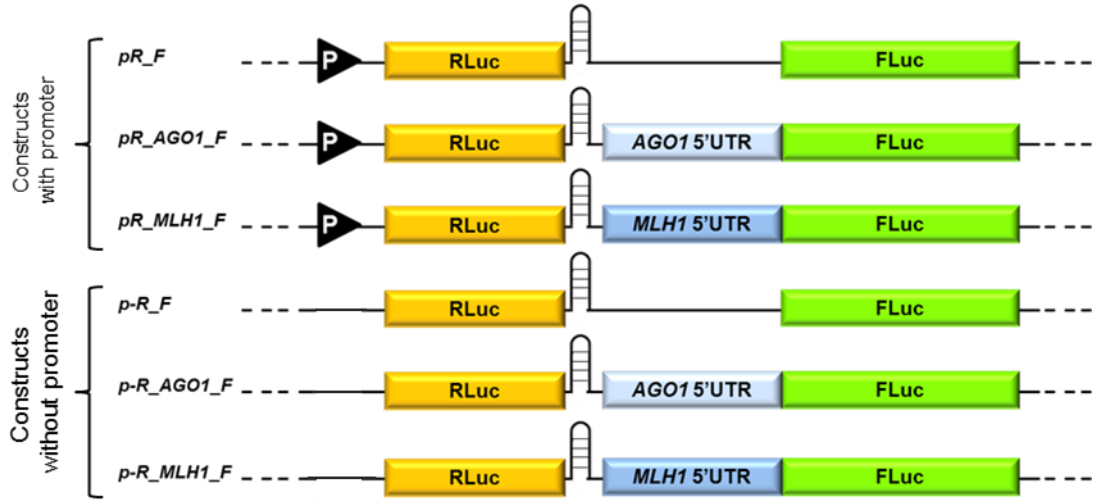


Figure 3.38 — Schematic representation of the constructs used to check whether *AGO1* 5'UTR contains a cryptic promoter. RLuc is the *Renilla* luciferase cap-dependent-translated cistron (yellow box) and FLuc the firefly cap-independent-translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Boxes in different shades of blue represent the different sequences cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_AGO1_F,* the *AGO1* 5'UTR-containing vector, is the sequence under study; *pR_MLH1_F*, the *MLH1* 5'UTR-containing vector, is the positive control for the presence of cryptic promoters. *p-R_F*, *p-R_AGO1_F* and *p-R_MLH1_F* are the counterpart promoterless plasmids. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

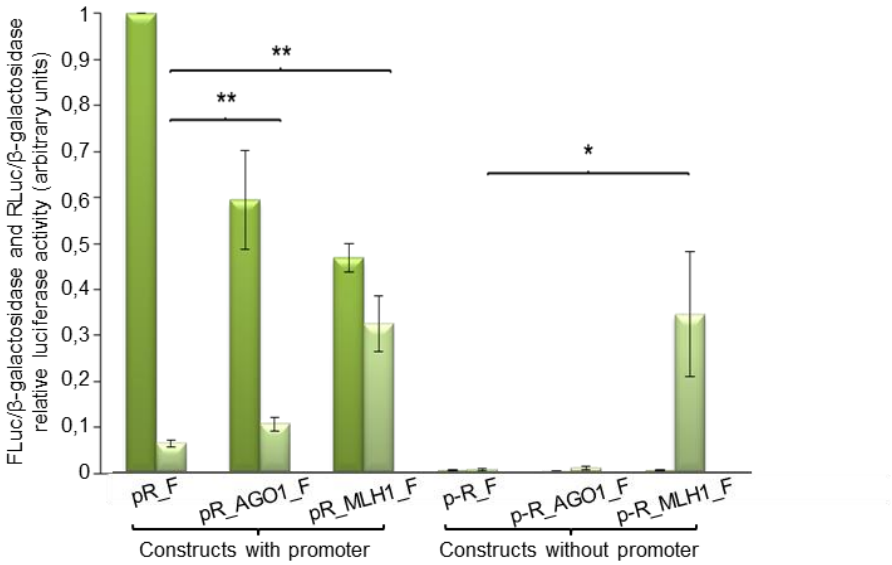


Figure 3.39 — *AGO1* 5'UTR does not contain a cryptic promoter. HeLa cells were transfected with promoter-containing constructs (*pR_F*, *pR_AGO1_F* and *pR_MLH1_F*) or promoterless constructs (*p-R_F*, *p-R_AGO1_F* and *p-R_MLH1_F*), and co-transfected with *β-galactosidase*-encoding plasmid (*pSV-β-Galactosidase* Control Vector), an internal control for transfection efficiency in mammalian cells. Relative RLuc (dark green bars) and FLuc (light green bars) expression levels were obtained by normalising each of them to those from *β-galactosidase*-expressing plasmid, all measured by luminometry assays. Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. * $P<0.05$, ** $P<0.01$.

Then, we evaluated the existence of alternative splicing events that may contribute to the observed increased levels in relative FLuc expression. For that, we analysed the integrity of the transcript by RT–PCR. The resulting complementary DNA (cDNA) was amplified with two pairs of primers spanning the whole sequence, according to the positions depicted in figure 3.40.A. Each pair of primers originated a single DNA fragment and with the expected size — 1583bp for amplified fragment I and 2100bp for amplified fragment II. Plasmid DNA fragment I is bigger than the corresponding cDNA because it includes a chimeric intron that is removed during mRNA processing. The whole sequence of both amplified fragments was confirmed by Sanger sequencing (data not shown). Both *RLuc* and *AGO1* 5'UTR sequences are intact proving no splicing has occurred within these sequences.

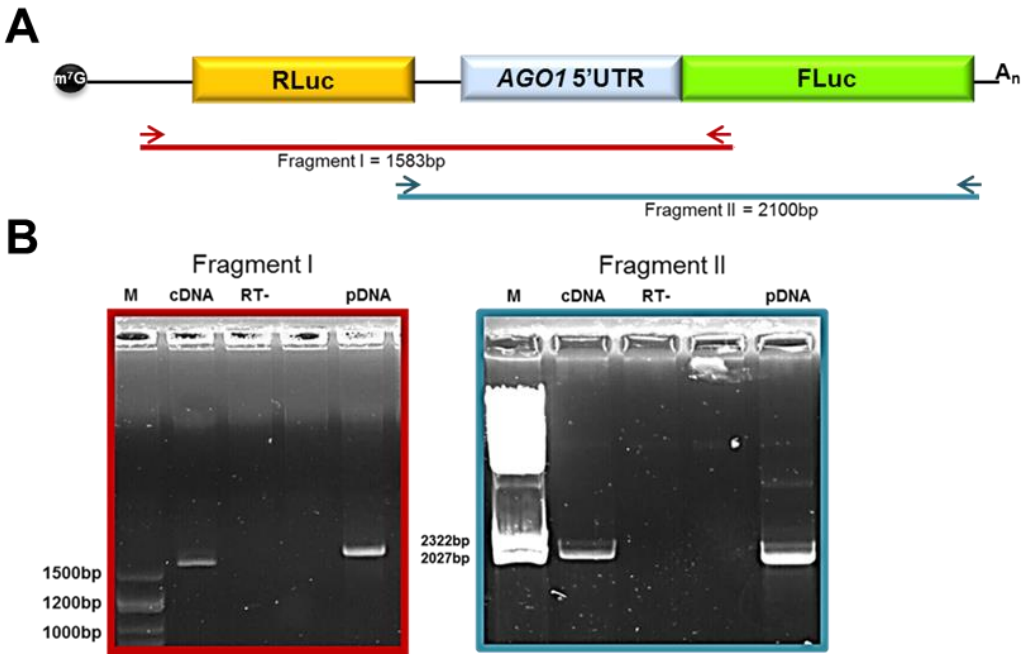


Figure 3.40 — RT–PCR analysis of the transcribed mRNA containing *AGO1* 5'UTR confirmed the integrity of the transcript. (A) Schematic representation of the putative mRNA transcribed from the equivalent transfected plasmid DNA. Arrows indicate the location of the primers used to amplify the corresponding complementary DNA (cDNA). (B) Agarose gels showing the amplified fragments. Each pair of primers originates one fragment only. cDNA fragment I is shorter than the corresponding amplified plasmid DNA (pDNA), because the latter includes a chimeric intron that has been removed during splicing. Fragment II is alike in both cases. M is the NZYLadderVI DNA molecular weight ladder (NZYTech); RT- indicates the PCR amplification reaction without cDNA synthesis step, proving no DNA contamination occurred in the cDNA sample. The blank lanes contain the PCR negative controls.

Monitoring the integrity of the produced mRNA gives a good indication that only one bicistronic transcript is being produced. However, obtaining a single fragment from each pair of primers does not completely rule out alternative splicing events. Thus, we performed a knock-down of RLuc and FLuc to understand whether reducing expression of RLuc mRNA would concomitantly and equally reduce FLuc expression and *vice versa*, proving both ORF are

originated from the same transcript. In this regard, we co-transfected HeLa cells with each of the plasmids used to assess cap-independent translation activity (p*R_F*, p*R_HBB_F*, p*R_MYC_F*, p*R_EMCV_F* and p*R_AGO1_F*) or cryptic promoter activity (p*R_MLH1_F*) and co-transfected them with siRNA against each of the proteins encoded by the transcript (RLuc and FLuc) or GFP (scrumble siRNA for control conditions). We evaluated the effect of each knock-down condition by luminometry assays and the results are depicted in figure 3.41. Figure 3.41.A shows the effect of each knock-down on the relative RLuc expression levels. Regarding the control condition (knock-down of GFP), *pR_F*, *pR_HBB_F* and *pR_MYC_F* expressed similar relative levels of RLuc, whereas *pR_EMCV_F* expressed much greater relative RLuc levels and *pR_AGO1_F* and *pR_MLH1_F* lower relative expression levels, observed before. After RLuc knock-down, relative RLuc expression levels from each transfected plasmid significantly decreased. As predicted, the same occurred after FLuc knock-down, which indicates both ORF are actually in the same bicistronic transcript. Oddly, the levels of RLuc from *pR_MYC_F* after FLuc knock-down remained similar to those obtained after GFP knock-down. Figure 3.41.B shows the effect of all performed knock-downs on relative FLuc expression levels from each plasmid. The levels of FLuc after GFP knock-down were in line with our previous results: relative FLuc expression levels from p*R_HBB_F* were similar to those from the empty plasmid, whereas those from p*R_MYC_F*, p*R_EMCV_F* and p*R_AGO1_F* were significantly greater than those from p*R_F* (figure 3.37). Still, after RLuc knock-down, relative FLuc expression levels from p*R_F* and p*R_HBB_F* were maintained. Such results corroborate the fact that no sequence upstream FLuc ORF is able to drive its transcription or translation. Hence, in both cases, FLuc expression measured after GFP and RLuc knock-down is similar. After FLuc knock-down, however, that residual FLuc expression is completely abolished, which is, as predicted, common to all other plasmids. Regarding relative FLuc expression levels from the other plasmids as a consequence of RLuc knock-down, we observed that those measured from *pR_MYC_F*, *pR_EMCV_F* and *pR_AGO1_F* were significantly lower than those observed in control conditions (GFP knock-down), indicating FLuc expression levels decrease concomitantly with the reduction in RLuc-containing mRNA, meaning both cistrons are included in the same bicistronic mRNA. On the contrary, in the case of *pR_MLH1_F*, the relative FLuc expression levels after RLuc knock-down did not decrease concomitantly, which is explained by the cryptic promoter included within the *MLH1* 5'UTR. Thus, the relative FLuc expression levels from the bicistronic mRNA were drastically reduced to virtually inexistent levels, but those from the monocistronic transcript produced from the cryptic promoter were not, which explains why FLuc expression levels remained high.

Altogether, these results confirm that the *pR_AGO1_F* plasmid, when expressed in HeLa cells, originates only a single bicistronic transcript. Furthermore, these results confirm *AGO1* 5'UTR is able to mediate FLuc expression via a non-canonical mechanism of translation initiation.

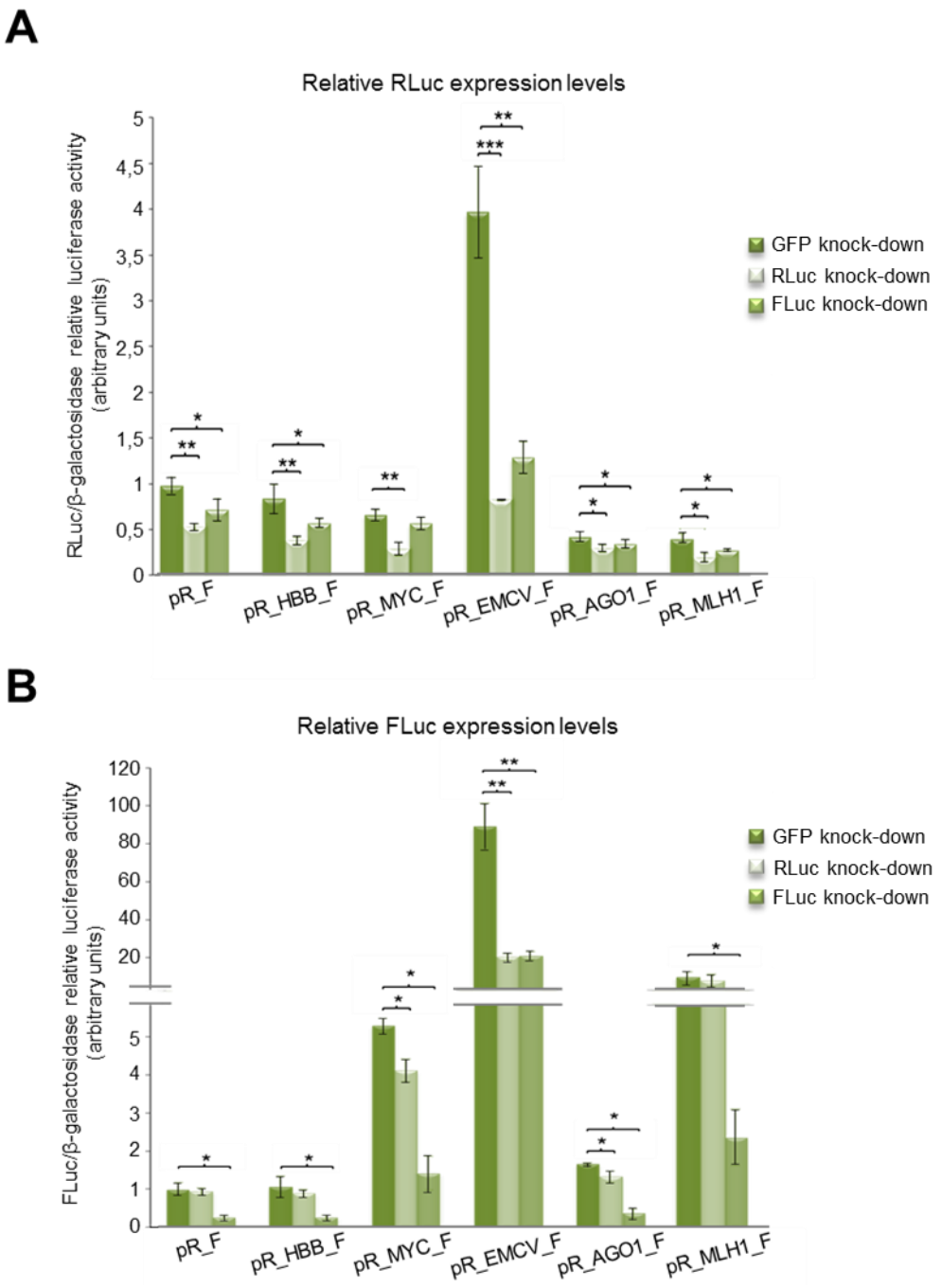


Figure 3.41 — Knock-down of RLuc and FLuc proved both proteins are produced from the same transcript. HeLa cells were co-transfected with siRNA against GFP (control conditions) RLuc or FLuc, and each of the constructs *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F*, *pR_AGO1_F* and *pR_MLH1_F*. (A) Relative RLuc expression levels from each plasmid after knock-down of either GFP (dark green bars), RLuc (greyed green bars) or FLuc (light green bars). (B) FLuc expression levels from each plasmid in the same conditions. Asterisks (*) indicate statistical significance in relation to the indicated counterpart. **P*<0.05, ***P*<0.01, ****P*<0.001.

III.4. *AGO1* 5'UTR-mediated FLuc expression is maintained under stress conditions

After ruling out false-positive events contributing to altered FLuc expression, we checked whether the identified cap-independent translation activity would be maintained under stress conditions. For that purpose, we subjected cells to several external stimuli known for reducing cap-dependent translation initiation. HeLa cells were transfected with either siRNA agaist GFP (control conditions) or against eIF4E, and, 24 h later, transfected with each of the plasmids used to assess *AGO1* 5'UTR-mediated cap-independent translation activity — *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F* and *pR_AGO1_F*. Figure 3.42.A shows a severe decrease in eIF4E amount of protein in cells transfected with siRNA against eIF4E compared to those in control conditions. Figure 3.42.B shows relative FLuc expression levels from each plasmid either in control conditions or under eIF4E knock-down conditions. In control conditions, the relative FLuc expression levels from each of the transfected plasmids is similar to those previously obtained (figure 3.37), that is to say *HBB* 5'UTR-containing plasmid cannot mediate FLuc cap-independent translation initiation, whereas *c-Myc* IRES-, *EMCV* IRES-, and *AGO1* 5'UTR-containing plasmids can. Under conditions of eIF4E knock-down, we observed that relative FLuc expression levels from *pR_MYC_F* and *pR_EMCV_F* were also significantly greater than those from *pR_F* in control conditions, but those from *pR_HBB_F* were not, confirming all positive controls are able to mediate cap-independent translation initiation in cells with reduced levels of eIF4E. Furthermore, relative FLuc expression levels from *pR_AGO1_F* in eIF4E knocked-down cells, not only is significantly greater than those from *pR_F* in control conditions, but is also significantly greater than those observed from such plasmid in control conditions — from 2.2-fold in control conditions to 4.5-fold in cells with reduced levels of eIF4E (figure 3.42.B). These results suggest *AGO1* 5'UTR is able to mediate cap-independent translation initiation activity under conditions impairing canonical cap-dependent translation initiation, and, further, that such mechanism is enhanced by reduced levels of available eIF4E, the cap-binding protein.

Another mechanism that prevents eIF4E from binding to the cap structure and hence impairs cap-dependent translation is treating cells with 4EGI-1 compound, an inhibitor of the eIF4E–eIF4G interaction — a required reaction for cap-dependent translation initiation to occur (Moerke *et al.*, 2007). This compound mimics the activity of naturally occurring molecules known as 4E-BPs (eIF4E-binding proteins). These proteins bind to eIF4E, preventing its association with eIF4G and therefore inhibit canonical translation. Under normal conditions, 4E-BPs are phosphorylated by mTOR kinase and are not able to bind eIF4E, leaving those molecules free to bind eIF4G and properly initiate cap-dependent translation. However, if 4E-BPs are not phosphorylated, they sequester eIF4E, impairing its binding to eIF4G (cf. section 1.2.1 and Showkat *et al.*, 2014). 4EGI-1, like 4E-BPs, associates with a binding site on eIF4E, displacing eIF4G and the subsequent formation of the eIF4F complex (Moerke *et al.*, 2007). To

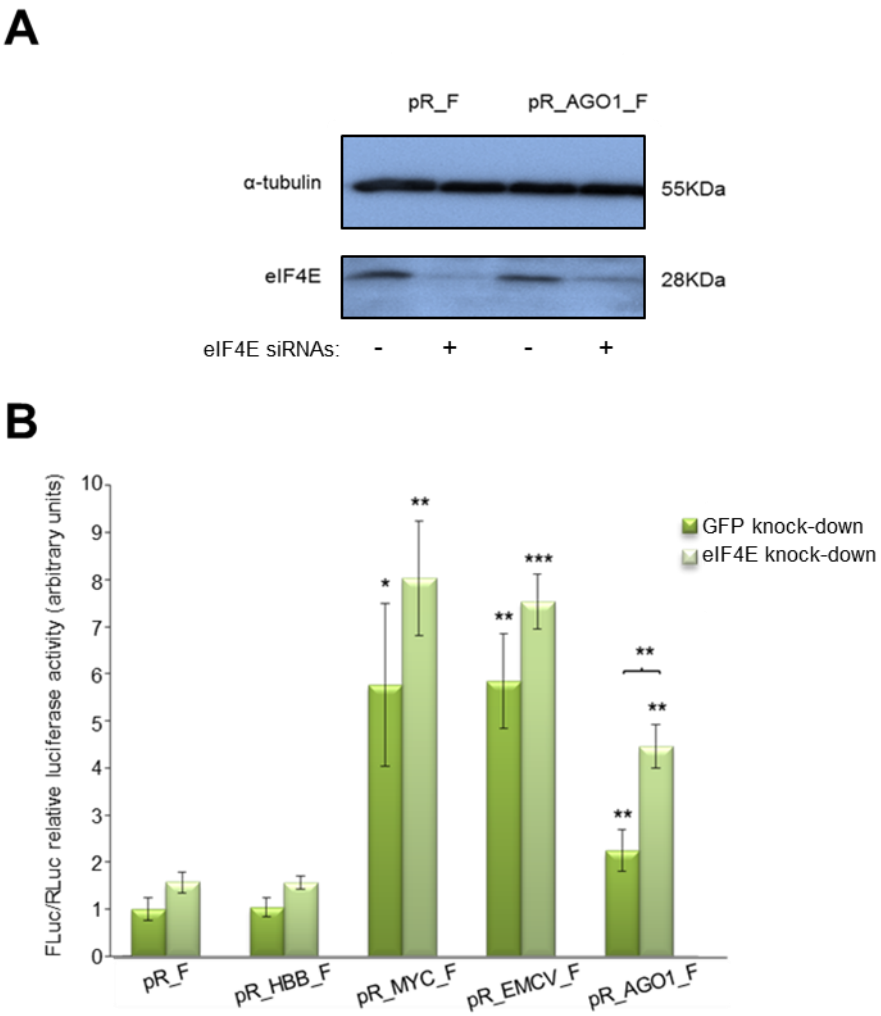


Figure 3.42 — *AGO1* 5'UTR mediates a more efficient translation of FLuc in HeLa cells under eIF4E knock-down conditions. HeLa cells were transfected with siRNA against eIF4E or GFP (control conditions) and, 24 h later, with plasmids *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F* and *pR_AGO1_F*. (A) Western blot against eIF4E showing knock-down efficiency. (-) indicates GFP siRNA trasnsfection and (+) indicates eIF4E siRNA transfection. α-tubulin was used as a loading control for the amount of protein. (B) Relative luciferase activity measured from each plasmid. Dark green bars indicate conditions of GFP siRNA cellular treatment and light green bars indicate eIF4E siRNA transfection conditions. Asterisks (*) indicate statistical significance in relation to the empty counterpart in control conditions or in relation to the indicated plasmid. *P<0.05, **P<0.01, ***P<0.001.

understand the effect of such drug on *AGO1* 5'UTR-mediated non-canonical translation initiation, we treated HeLa cells transfected with each of the plasmids used to evaluate cap-independent translation activity with either 200 μM 4EGI-1 or DMSO (vehicle). Figure 3.43.A shows the Western blot analysis of a co-immunoprecipitation (co-IP) reaction between eIF4E and eIF4G to verify whether the drug treatment conditions prevent or reduce eIF4E–eIF4G interaction. Results show that in the Pre-IP lysate (the pure lysate prior to any antibody pull-

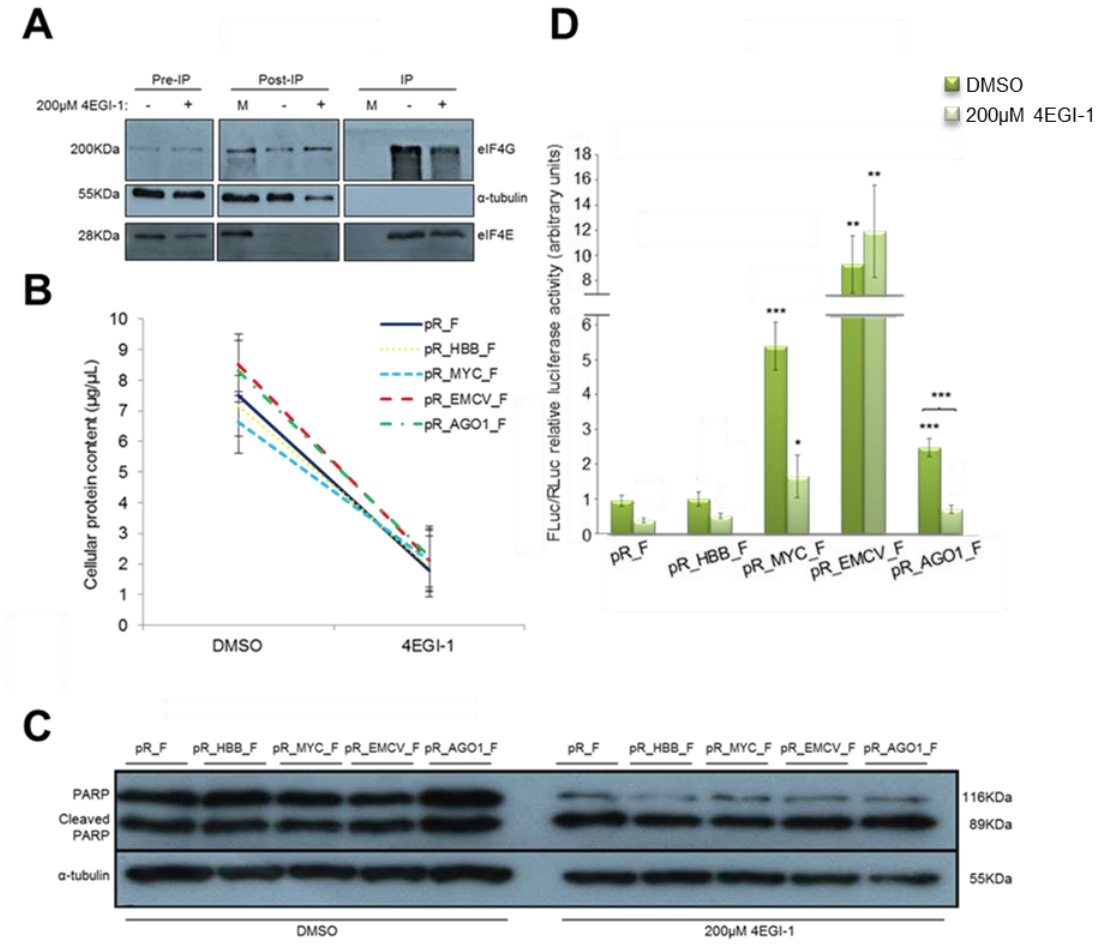


Figure 3.43 — Treatment with 4EGI-1 inhibits interaction between eIF4G and eIF4E, drastically reduces global protein synthesis and inhibits *AGO1* 5'UTR-mediated internal translation initiation, in HeLa cells. (A) Western blot analysis of co-immunoprecipitation between eIF4E and eIF4G after treatment of HeLa cells with eIF4E–eIF4G interaction-inhibiting drug. (-) indicates cells were treated with DMSO (vehicle); (+) indicates cells were treated with 200 μM of 4EGI-1; (M) indicates no treatment was applied and no agarose beads were added; "Pre-IP" represents the pre-immunopreciptation lysate; "Post-IP" represents the lysate that did not bind the agarose beads; "IP" represents the actual immunoprecipitated lysate, i.e., everything to which the agarose beads have bound after immunoprecipitating lysate with anti-eIF4E antibody, and everything that has co-immunoprecipitated with it. α-tubulin was used as a loading control for the amount of protein. (B) Variation of the cellular protein content (expressed as μg of protein per μl of lysate) between control conditions and treatment with 4EGI-1 for 20 h. (C) Western blot against Poly ADP ribose polymerase (PARP) protein and its cleaved fragment under treatment with either DMSO or 4EGI-1. α-tubulin was used as a loading control for the amount of protein. (D) Relative luciferase activity measured from HeLa cells transfected with *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F* and *pR_AGO1_F* and then treated with 200 μM 4EGI-1. Dark green bars represent treatment with DMSO and light green bars represent treatment with 200 μM of 4EGI-1. Asterisks (*) indicate statistical significance in relation to the empty plasmid in control conditions or in relation to the indicated counterpart. **P*<0.05, ***P*<0.01, ****P*<0.001

down) both subunits are present in the cells treated with DMSO or in those treated with 4EGI-1, as expected. After pulling down eIF4E subunit (IP lysate), we expected to see a decrease in the amount of eIF4G in the lysate of the cells treated with drug, because its interaction with eIF4E would be impaired and hence it would not have bound to the bead-bound eIF4E subunits as under normal conditions. In fact, we observed that the amount of eIF4G decreases in the lysate of cells treated with the drug, but in the lysate of cells treated with DMSO it does not. Adding to this, in the Post-IP lysate (lysate containing everything that has not been bound to the beads), we can see the detection of eIF4G (greater amount of protein in the lysate of cells treated with the drug, confirming less was bound to eIF4E). The effects of the drug were also confirmed by the great decrease in total protein content, reflecting the inability of cells to perform regular levels of cap-dependent translation initiation, after 20 h of drug treatment, as depicted in figure 3.43.B. Furthermore, an indirect measure of its effect was the cleavage of Poly ADP ribose polymerase (PARP), which occurs as a response to the induction of apoptosis caused by treatment with 4EGI-1 (Fan *et al.*, 2010; Descamps *et al.*, 2012). Western blot analysis indicates that the amount of full-length PARP protein is drastically reduced after treatment with 4EGI-1 (figure 3.43.C). As far as relative FLuc expression levels from HeLa cells transfected with each of the plasmids are concerned, we observed that, in control conditions, there is a 2.5-fold increase in relative FLuc expression levels from *AGO1* 5'UTR-containing plasmid compared to that from the empty counterpart, similar to previously obtained results. Relative FLuc expression levels from *pR_MYC_F* and *pR_EMCV_F* were also similar to those previously obtained — 5.4- and 9.4-fold those from the empty counterpart, respectively, whereas those from *pR_HBB_F* were similar to those from *R_F*. In cells treated with 200 µM 4EGI-1, relative FLuc expression levels from *pR_EMCV_F* were similar to those in control conditions, and those from *pR_MYC_F*, although lower than in control conditions, were significantly greater than those from the empty plasmid in cells treated with DMSO. The results obtained from *pR_MYC_F* and *pR_EMCV_F* are in accordance with the presence of a functional IRES element, the *c-Myc* and the *EMCV* IRES, respectively (Stoneley *et al.*, 2000a; Bochkov and Palmenberg, 2006). However, relative FLuc expression levels measured from *pR_AGO1_F* in 4EGI-1-treated cells are significantly lower than those in control conditions, and similar to those from *pR_F* and *pR_HBB_F*, indicating no cap-independent activity mediated by *AGO1* 5'UTR occurs when the interaction between eIF4E and eIF4G is blocked. These results suggest that the putative *AGO1* 5'UTR-mediated cap-independent translation initiation is dependent upon eIF4G. In order to confirm whether FLuc expression under the control of *AGO1* 5'UTR would be maintained under other conditions that impair cap-dependent translation initiation, we subjected cells to several external stimuli. For that, we transfected HeLa cells with either *pR_AGO1_F* or one of the controls — *pR_F*, *pR_HBB_F*, *pR_MYC_F* and *pR_EMCV_F* — and treated them with 80 nM rapamycin, 200 µM $CoCl_2$ or 1 µM thapsigargin, or the corresponding vehicles (H_2O for $CoCl_2$, and DMSO for both rapamycin and thapsigargin). Western blot analysis of transfected cells revealed that S6K protein is absent in cells treated with rapamycin but not in cells treated with

DMSO, indicating rapamycin is blocking mTOR kinase activity (figure 3.44.A). Relative FLuc expression levels from *R_AGO1_F* in cells treated with rapamycin were similar to those from the same plasmid in control conditions, suggesting *AGO1* 5'UTR is able to mediate cap-independent translation initiation under impaired mTOR kinase activity conditions (figure 3.44.B). Additionally, both positive controls mediate FLuc translation under stress conditions, (figure 3.44.B). Treatment with $CoCl_2$ led to a cellular hypoxic status, as confirmed by Western blot analysis of cells treated with such compound, in which there is an accumulation of HIF1α protein compared to cells treated with H_2O (figure 3.44.C). Relative FLuc expression levels from *pR_AGO1_F* in cells subjected to hypoxia were similar to those from the same vector in cells treated with vehicle and significantly greater than those from the negative controls in control conditions. Again, relative FLuc expression levels from both positive controls were maintained. Western blot analysis of transfected cells shows an increased amount of phosphorylated eIF2α protein in cells treated with thapsigargin (figure 3.44. E). Relative FLuc expression levels from pR_AGO1_F in the same cells are also significantly greater than those from the negative controls in control conditions.

Altogether, these results show that *AGO1* 5'UTR is able to maintain protein synthesis under cap-dependent translation initiation-impairing conditions. Furthermore, both positive controls (*pR_MYC_F* and *pR_EMCV_F*) behave similarly to *pR_AGO1_F* under all tested treatments, whereas no relative FLuc expression was driven by *pR_HBB_F* (negative control) or *pR_F*, supporting the evidence of cap-independent activity mediated by *AGO1* 5'UTR.

III.5. *AGO1* 5'UTR-mediated cap-independent translation requires a free 5' end

To further understand the mechanism employed by *AGO1* 5'UTR to mediate non-canonical translation, we analysed its behaviour in cells transfected with *in vitro* transcribed and polyadenylated bicistronic or monocistronic mRNA. For this purpose, we produced *in vitro* transcribed, capped and polyadenylated mRNA containing each of the sequences required for evaluating cap-independent translation activity in a bicistronic context (figure 3.45.A). Prior to transfection, we assessed the integrity of the produced transcripts with an agarose–folmaldehyde denaturing gel electrophoresis (figure 3.45.B). HeLa cells were transfected with each of the indicated transcripts — *R_F*, *R_HBB_F*, *R_MYC_F* and *R_AGO1_F* — and relative FLuc expression levels from each transcript were measured by luminometry assays. The outcome of such experiment is depicted in figure 3.46. As expected, relative FLuc expression levels from *R_HBB_F* were similar to those from *R_F*, whereas those from *R_MYC_F* were significantly greater than those from the empty plasmid, confirming *c-Myc* IRES is able to mediate cap-independent translation initiation in a bicistronic context (figure 3.46).

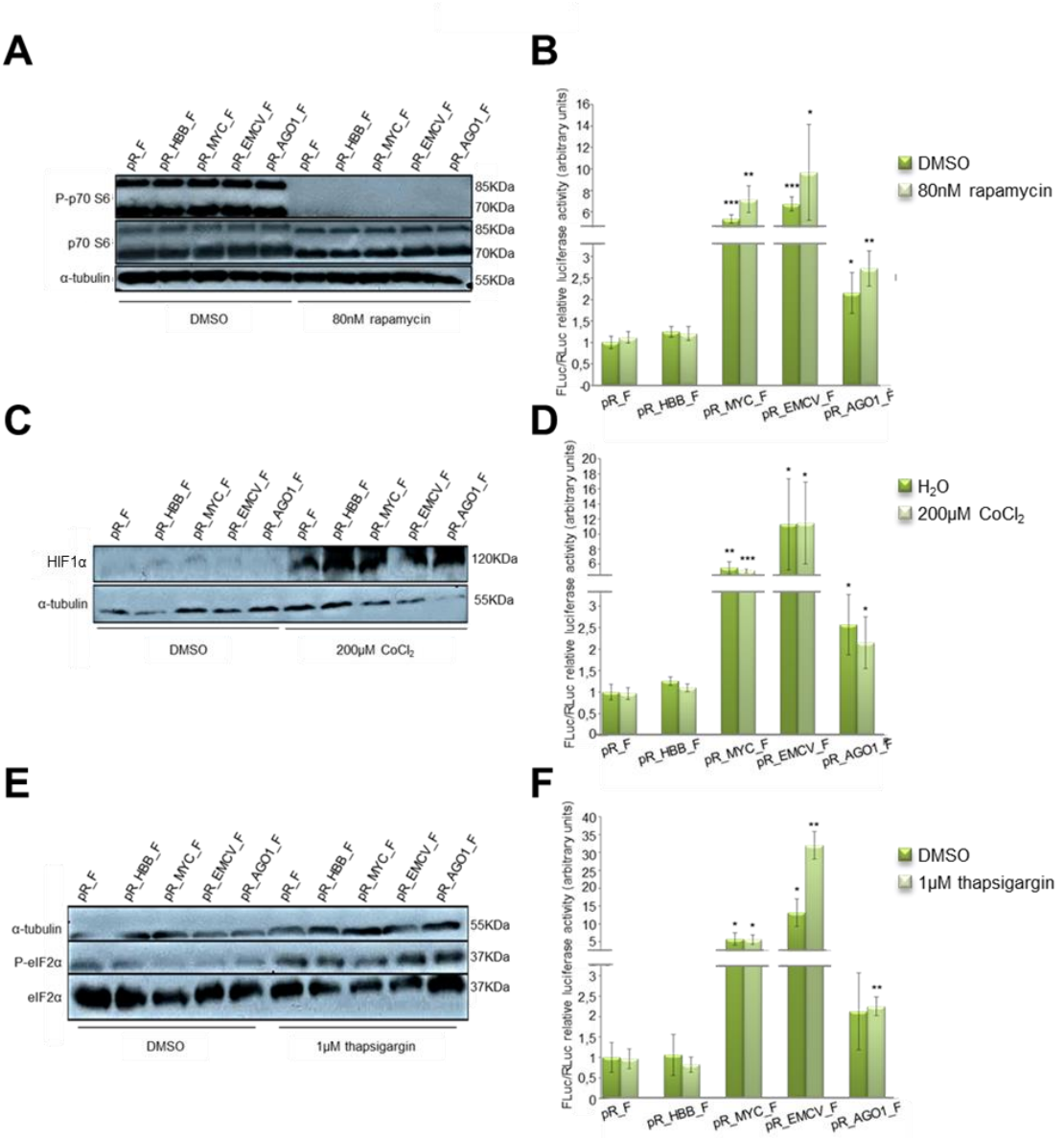


Figure 3.44 — Treatment of HeLa cells with rapamycin, $CoCl_2$ and thapsigargin does not affect relative FLuc expression mediated by *AGO1* 5'UTR in a bicistronic context. HeLa cells were transfected with *pR_F*, *pR_HBB_F*, *pR_MYC_F*, *pR_EMCV_F* and *pR_AGO1_F* plasmids and treated with 80 nM of rapamycin (A, B), 200 μM of $CoCl_2$ (C, D), or 1 μM of thapsigargin (E, F). A, C and E: Western blot against phosphorylated and non-phosphorylated S6K proteins (A) — in which no expression of the former indicates rapamycin is impairing mTOR kinase activity —, HIF1α protein (B) — whose increased expression indicates a cellular hypoxic status —, and phosphorylated and non-phosphorylated eIF2α proteins (C) — in which increased expression of the former is a consequence of endoplasmic reticulum stress. α-tubulin was used as a loading control for the amount of protein. B, D and F: relative FLuc expression levels in cells under treatment with rapamycin (B), $CoCl_2$ (D) and thapsigargin (F).. Dark green bars indicate relative FLuc expression levels in cells treated with vehicle (DMSO for rapamycin and thapsigargin, or H_2O for $CoCl_2$) and light green bars each of the aforementioned stimuli. Presented results are the outcome of, at least, three independent experiments. Asterisks (*) indicate statistical significance in relation to the empty counterpart in control conditions.*P<0.05, **P<0.01, ***P<0.001

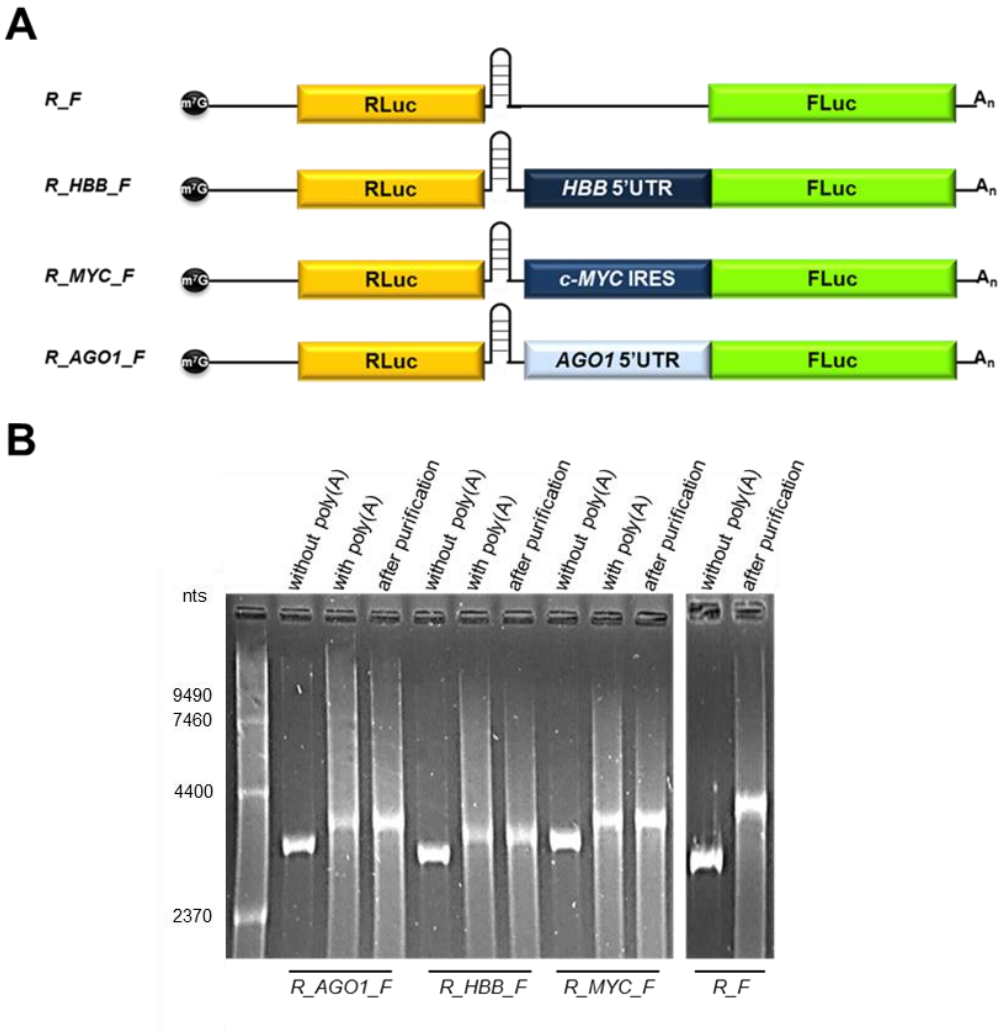


Figure 3.45 — *In vitro* transcribed mRNA used to detect *AGO1* 5'UTR-mediated cap-independent translation activity in a bicistronic context. (A) Schematic representation of the capped (m^7G) and polyadenylated (A_n) *in vitro* transcribed bicistronic mRNA produced from the corresponding plasmids (cf. figure 3.36). RLuc is the *Renilla* luciferase cap-dependent-translated cistron (yellow box) and FLuc the firefly luciferase cap-independent-translated cistron (green box). Boxes in different shades of blue represent the different sequences cloned upstream FLuc AUG. *R_F* is the empty transcript; *R_HBB_F,* the human *β-globin* (*HBB*) 5'UTR-containing transcript, is the negative control for cap-independent activity; *R_MYC_F,* the *c-MYC* IRES-containing transcript, is the cellular positive control for cap-independent activity; *R_AGO1_F,* the *AGO1* 5'UTR-containing transcript, is the sequence under study. All transcripts contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation. (B) Denaturing agarose–formaldehyde gel showing the integrity of the produced mRNA. Expected transcript sizes before polyadenylation: 3004bp (*R_F*); 3054bp (*R_HBB_F*); 3344bp (*R_MYC_F*); 3217bp (*R_AGO1_F*). M: 0.24-9.5 Kb RNA Ladder (Invitrogen); nts: weight, in bases, of RNA ladder fragments. Without poly(A): *in vitro* capped transcripts before polyadenylation. With poly(A): *in vitro* capped transcripts after polyadenylation. After purification: *in vitro* capped and polyadenylated transcripts after RNA purification by phenol–chloroform extraction.

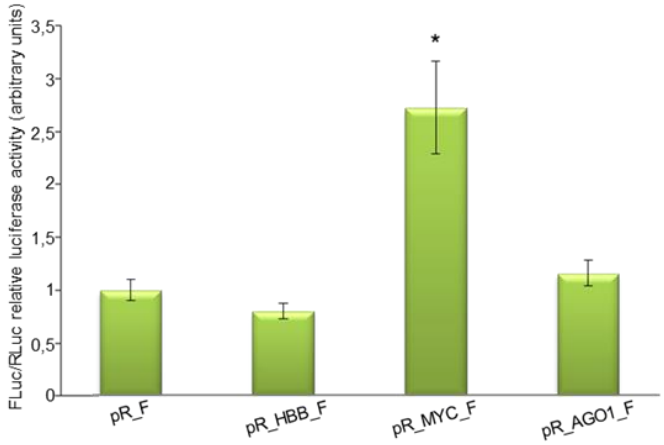


Figure 3.46 — Expression of FLuc reporter protein is not mediated by *AGO1* 5'UTR in a bicistronic context without nuclear experience. HeLa cells were transfected either with *AGO1* 5'UTR-containing plasmid (*R_AGO1_F*) or with one of the controls used in the experiment: *R_F* (empty transcript), *R_HBB_F* (*HBB* 5'UTR-containing transcript) or *R_MYC_F* (*c-Myc* IRES-containing transcript). Presented data are the result of, at least, three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. **P*<0.05

However, relative FLuc expression levels from *R_AGO1_F* were similar to those from the empty plasmid, indicating this sequence is not able to mediate cap-independent translation initiation of the downstream cistron in a bicistronic *in vitro* transcribed and polyadenylated mRNA (figure 3.46). This result indicates *AGO1* 5'UTR sequence is not able to mediate internal cap-independent translation initiation in conditions in which it does not go through a nuclear experience. This suggests the existence of nuclear proteins that need to bind *AGO1* 5'UTR in order for it to have a role in mediating cap-independent translation initiation.

In order to understand whether *AGO1* 5'UTR is able to mediate cap-independent translation initiation in a 5' end-free system, we produced *in vitro* monocistronic polyadenylated transcripts encoding FLuc and lacking the cap structure. Since transcripts with unprotected 5' ends are unstable and prone to degradation, we used a cap structure analogue unable to mediate cap-dependent translation [G(5')ppp(5')A, by New England Biolabs] to protect transcripts from degradation. Thus, we produced two sets of *in vitro* transcribed and polyadenylated monocistronic mRNA: capped (m^7G) mRNA, or uncapped [cap structure analogue G(5')ppp(5')A] mRNA (figure 3.47.A) and checked their integrity in a denaturing agarose–formaldehyde electrophoresis gel (figure 3.47.B). We transfected HeLa cells with each of the capped (5'G-capped) or uncapped (5'A-capped) *in vitro* transcribed and polyadenylated monocistronic transcripts: empty transcript (*g_F* and *a_F*), *HBB* 5'UTR-containing transcript

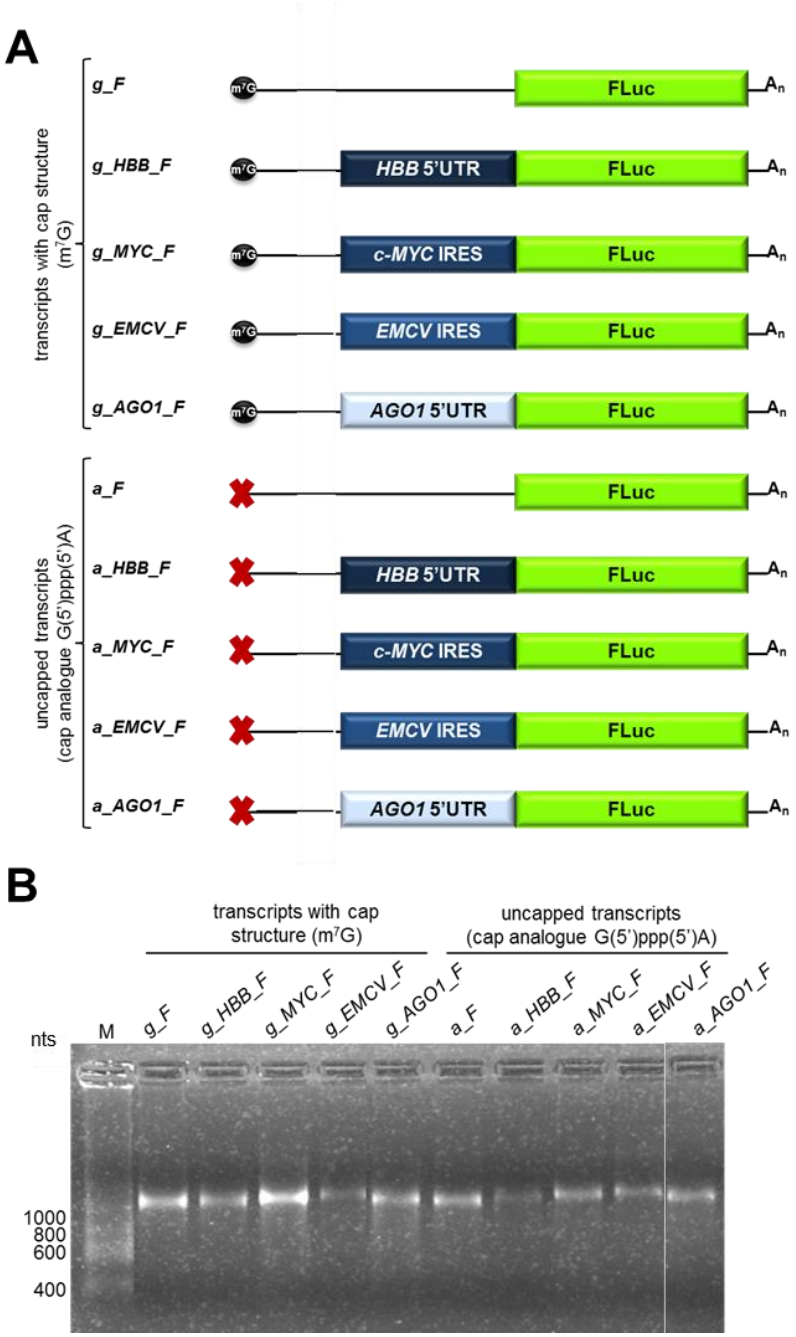


Figure 3.47 — *In vitro* transcribed and polyadenylated monocistronic transcripts used to evaluate *AGO1* 5'UTR-mediated translation dependency on a free 5' end. (A) Schematic representation of the *in vitro* transcribed and polyadenylated (A_n) monocistronic transcripts. *g_F* (empty transcript), *g_HBB_F* (*HBB* 5'UTR-containing transcript), *g_MYC_F* (*c-Myc* IRES-containing transcript), *g_EMCV_F* (*EMCV* IRES-containing transcript) and *g_AGO1_F* (AGO1 5'UTR-containing transcript) are the capped (m^7G, black circles) transcripts, and *a_F*, *a_HBB_F*, *a_MYC_F*, *a_EMCV_F* and *a_MLH1_F* are the counterpart uncapped [cap analogue G(5')ppp(5')A, red "X"] transcripts. FLuc is the firefly luciferase enconding cistron (green box). Boxes in different shades of blue represent different sequences cloned upstream FLuc AUG. (B) Denaturing agarose–formaldehyde gel showing the integrity of the *in vitro* transcribed and polyadenylated monocistronic transcripts. M: RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific); nts: molecular weight, in bases, of RNA ladder bands.

(*g_HBB_F* and *a_HBB_F*), *c-Myc* IRES-containing transcript (*g_MYC_F* and *a_MYC_F*), *EMCV* IRES-containing transcript (*g_EMCV_F* and a*_EMCV_F*) and *AGO1* 5'UTR-containing transcript (*g_AGO1_F* and a*_AGO1_F*) (figure 3.47) — and co-transfected them with β-galactosidase-encoding plasmid (pSV-β-Galactosidase control vector), an internal control for transfection efficiency in mammalian cells. Four hours after transfection, we assessed relative FLuc expression levels of each of the 5'A-capped transcript in relation to its 5'G-capped counterpart, using luminometry assays (figure 3.48).

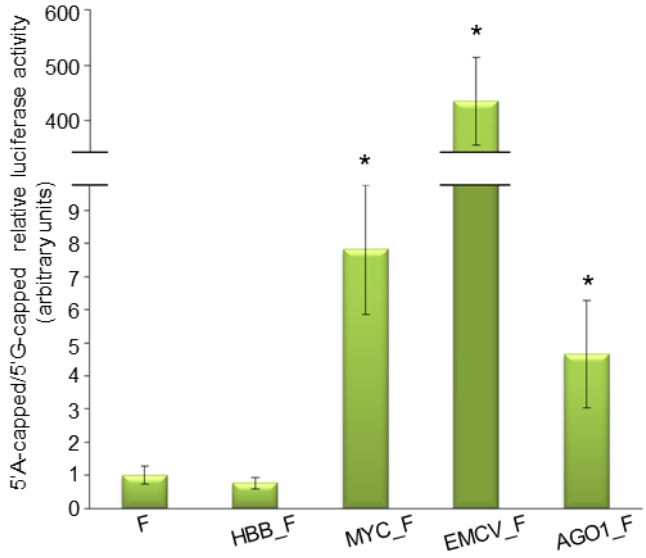


Figure 3.48 — AGO*1* 5'UTR mediates cap-independent translation in monocistronic transcripts lacking cap structure. HeLa cells were transfected with *in vitro* transcribed monocistronic mRNA, either capped (5'G-capped, m^7G) or uncapped [5'A-capped, cap analogue G(5')ppp(5')A], containing either *AGO1* 5'UTR (*AGO1_F*) or one of the controls used in the experiment: *F* (empty transcript), *HBB_F* (*HBB* 5'UTR-containing transcript), *MYC_F* (*c-Myc* IRES-containing transcript), or *EMCV_F* (*EMCV* IRES-containing transcript). Presented data are the result of, at least, three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty transcripts. **P*<0.05

We observed a 4.7-fold significant increase in relative FLuc expression levels from 5'A-capped *AGO1* 5'UTR-containing transcript compared to those from the counterpart empty transcript, arbitrarily set to 1. This result suggests *AGO1* 5'UTR is able to mediate translation initiation in the absence of the cap-structure. Furthermore, relative FLuc expression levels from both 5'A-capped *c-Myc* and *EMCV* IRES-containing transcripts were also significantly greater than those from the counterpart empty transcript, i.e., 7.8- and 435-fold, respectively, confirming the ability of such sequences to mediate cap-independent translation initiation. Conversely, relative FLuc expression levels from 5'A-capped *HBB* 5'UTR-containing transcript were similar to those from the empty plasmid, demonstrating this sequence is not able to drive cap-independent translation

initiation, and proving this system is robust to the detection of cap-independent translation initiation mechanisms. Relative FLuc expression levels from *AGO1_F* are lower than those from the cellular positive control, *MYC_F*, indicating the mode of action of each sequence in mediating cap-independent translation initiation may be different.

All in all, these results indicate *AGO1* 5'UTR is able to mediate translation initiation in the absence of the cap structure, when the mRNA 5' end is free, but not via an internal ribosome entry site in a transcript that does not go through a nuclear experience, as is the case of *c-Myc* IRES and *UPF1* 5'UTR.

IV. *MLH1* 5'UTR regulates gene expression at transcription and translation level

In silico analysis of protein and mRNA expression levels indicate this protein as a putative candidate to be translated via an alternative mechanism of translation initiation. Experimental validation is therefore required to either confirm or rule out the ability of its 5'UTR to drive cap-independent translation initiation or an alternative cap-dependent mechanism that works in non-canonical conditions.

IV.1. *In silico* analysis of *MLH1* 5'UTR

A thorough analysis of *MLH1* 5'UTR revealed it is 198 nts long and contains an upstream AUG at position -111 from the initiation codon in frame with a stop codon, producing a uORF with two codons (figure 3.49.A). It contains an overall GC content of 53%, distributed over regions of higher GC content (up to 57.3% at nt -98) and other with lower content (minimum 47.5% at regions adjacent to 5' terminus; figure 3.49.B), and tendency to fold into structures of predicted stability — ΔG=-58.22kcal/mol (prediction with mFold software, figure 3.49.C). According to this prediction, the secondary structure formed within *MLH1* 5'UTR includes four stem loops — SL I, II, III and IV. SL II and III correspond to the region of highest GC content, suggesting that the stability of this predicted structure may be greater than that of the remaining predicted secondary structure. As previously mentioned these structures may impair the regular scanning of the 5'UTR and facilitate the recruitment of the ribosome to the vicinity of the main AUG, enhancing non-canonical mechanisms of translation initiation.

Sequence conservation throughout evolution may also provide some clues on the importance of the sequence under analysis among different mammalian species and therefore its putative relevance in non-canonical translation initiation. Thus, we compared human *MLH1* 5'UTR to those of other mammalian species to evaluate how conserved among species this sequence is and also to what extent the formation of the predicted stem loops is maintained among different species. Using Bioedit software, we compared *MLH1* 5'UTR sequences from different mammals (human, chimpanzee, rat and mouse, figure 3.50.A). From this analysis, we observe that human and chimpanzee *MLH1* 5'UTR are identical, whereas those of rat and mouse contain more 68 or 80 nts, respectively, at the 5' end, and less 52 or 31 nts, respectively at 3' end. Regarding structure conservation, according to RNAalifold software prediction, although the spatial organisation of the predicted structure differs from that predicted by mFold, we can identify the four stem loops predicted by the latter (figure 3.50.B), which are well preserved among species. The observed characteristics of *MLH1* 5'UTR indicate this is a well-conserved sequence capable of forming a complex and conserved secondary structure, which may be helpful for its putative role in mediating non-canonical translation initiation. Additionally, we evaluated the predicted formation of G-quadrupex structures within this sequence (table 3.3). Two quadruplex-forming G-rich sequences were found within *MLH1* 5'UTR initiating at nts 76 and 127. These predicted structures may influence translation initiation either by inhibiting cap-dependent translation or stimulating cap-independent mechanisms of translation initiation, namely IRES-mediated translation (Bugaut and Balasubramanian, 2012).

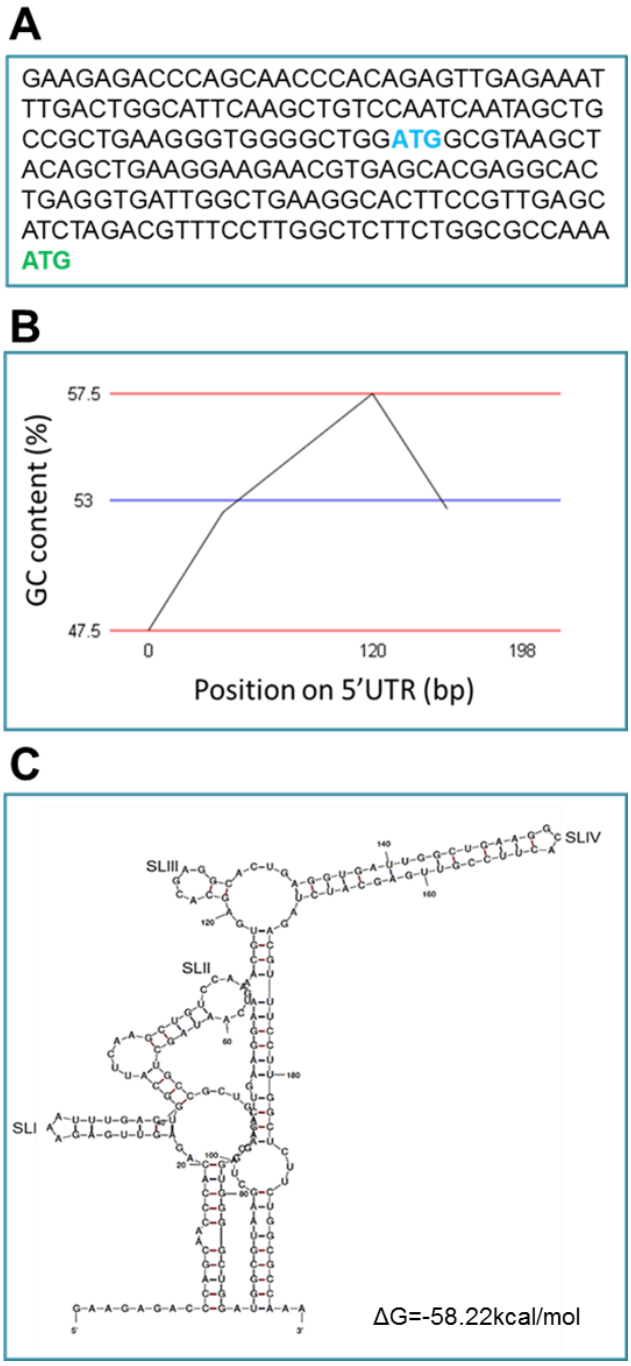


Figure 3.49 — Multiple features of human *MLH1* 5'UTR predicted *in silico*. (A) *MLH1* 5'UTR nucleotide sequence used in this work. **ATG** represents an upstream open reading frame at position -111 in frame with a stop codon (TAA) two codons downstream. **ATG** is the MLH1 translation initiation codon. (B) GC content (%) of different regions of *MLH1* 5'UTR (http://www.endmemo.com). The content (average 53%) ranges from 47.5% to 57.5% and the highest percentages tend to localise around nt -98 of the 5'UTR. (C) RNA secondary structure of *MLH1* 5'UTR predicted by mFold software (http://mfold.rna.albany.edu/?q=mfold). Based on this prediction, four stem loops (SL) — I, II, III and IV — are formed within *MLH1* 5'UTR and the structure is predictably stable (ΔG = -58.22kcal/mol).

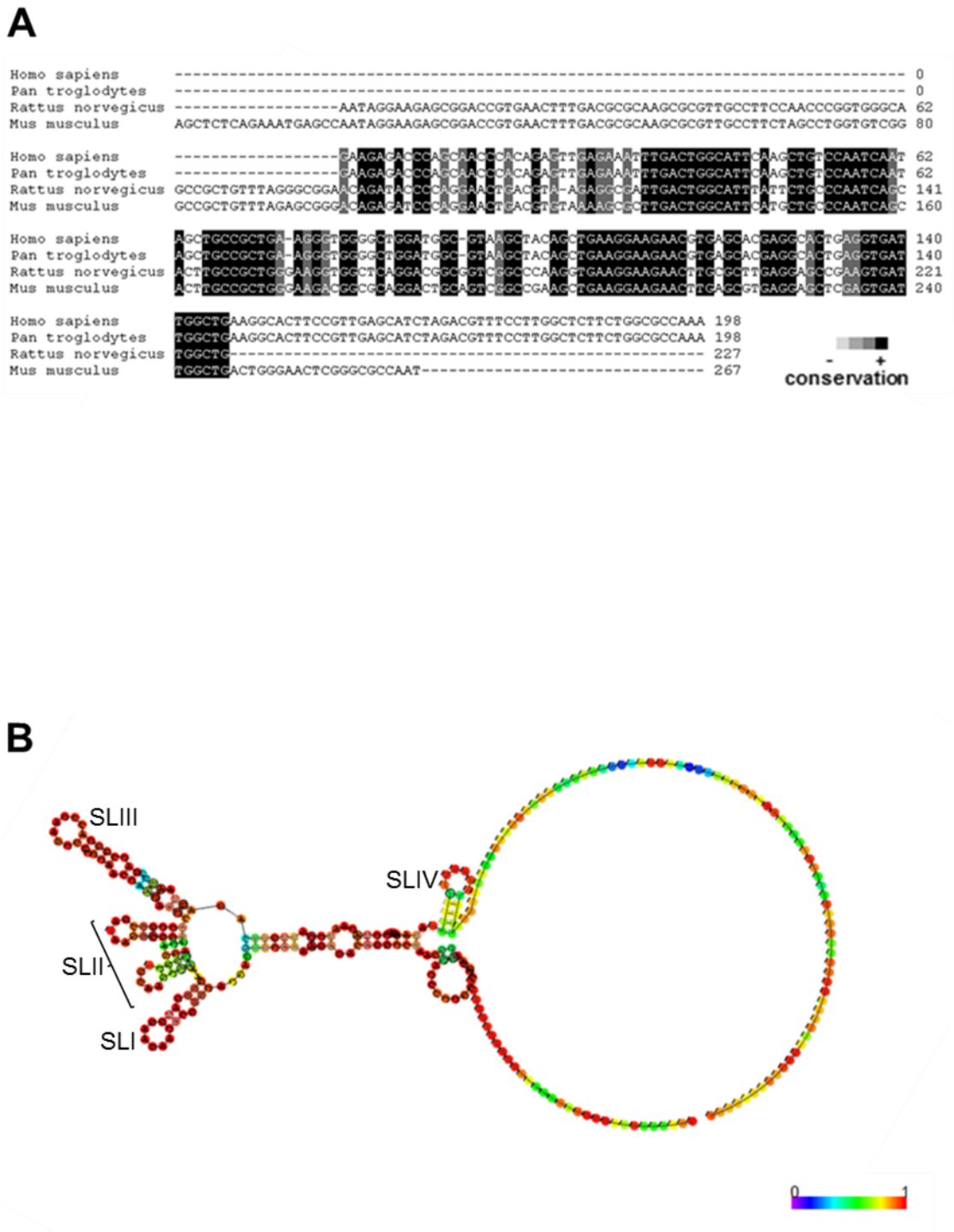


Figure 3.50 — Conservation of *MLH1* 5'UTR sequence among mammalian species. (A) Sequence alignment of *MLH1* 5'UTR among human, chimpanzee, rat and mouse obtained using Bioedit software (http://www.bioinformatics.org/sms2/color_align_cons.html). Grey scale indicates the degree of conservation among species for each nucleotide. White (-) indicates less conserved; black (+) indicates most conserved. (B) Predicted consensus secondary structure of the set of aligned sequences using RNAalifold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi). Coloured scale indicates the degree of conservation of the predicted secondary structure. Purple (0) indicates no secondary structure conservation; Red (1) indicates full secondary structure conservation.

Table 3.3 — Quadruplex-forming G-rich sequences (QGRS) found in human *MLH1* 5'UTR*

Position	Length	QGRS	G-score
76	16	GGGTGGGGCTGGATGG	20
127	24	GGCACTGAGGTGATTGGCTGAAGG	20

*The underlined GG represent those putatively involved in the formation of G-quadruplex structures. Position designates the first nucleotide of the QGRS sequences. The putative G-quadruplexes are identified using the motif $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$, where x is the number of guanine tetrads in the G-quadruplex, and y1, y2, y3 are the length of gaps, i.e., the length of the loops connecting the guanine tetrads. The motif consists of four equal length groups of guanines, separated by arbitrary nucleotide sequences with at least two tetrads ($x \geq 2$) and maximum length of 30 bases. The maximum length of 30 bases restricts G-groups to a maximum size of 6. G-score is a classification attributed by the software that evaluates a QGRS for its likelihood to form a stable G-quadruplex. Higher scoring sequences will make better candidates for G-quadruplex. The scoring method considers the following principles: shorter loops are more common than longer loops; G-quadruplexes tend to have loops roughly equal in size; the greater the number of guanine tetrads the more stable the quadruplex. The highest possible G-score, using the default maximum QGRS length of 30, is 105. (http://bioinformatics.ramapo.edu/QGRS/index.php)

IV.2. *MLH1* 5'UTR cryptic promoter seems to be tissue-specific

Previously in this work, we have used *MLH1* 5'UTR as a positive control for the presence of cryptic promoters within *UPF1* and *AGO1* 5'UTR, because this sequence includes a core promoter that is able to mediate transcription and subsequent translation of a downstream ORF (Ito *et al.*, 1999; Arita *et al.*, 2003). Our results presented so far, confirmed this evidence, as FLuc translation was mediated by *MLH1* 5'UTR in HeLa (figures 3.17 and 3.39), NCM460 and HCT116 cells transfected with promoterless constructs containing that sequence. In order to understand how such promoter behaves in different tissues, we transfected every tested cell line with promoterless plasmids (empty or containing *MLH1* 5'UTR) or the counterpart plasmids with promoter, as depicted in figure 3.51. We co-transfected cells with β-galactosidase-encoding plasmid (pSV-β-Galactosidase control vector) and, 24 h posttransfection, measured the relative RLuc and FLuc expression levels from each plasmid in relation to the internal control. Figure 3.52 shows the outcome of such experiment. Relative RLuc expression levels from all promoterless constructs decreased to background levels, as expected, because there was no promoter to drive transcription and subsequent translation of RLuc (figure 3.52.A). As for the promoter-containing constructs, RLuc expression levels from *pR_MLH1_F* in NCM460 cells is similar to that from *pR_F* in the same cells. However, the levels of RLuc expression from the *MLH1* 5'UTR-containing vector in HeLa and HCT116 cells are significantly lower than the corresponding levels from the empty vector, (figure 3.52.A). Regarding relative FLuc expression levels (figure 3.52.B), the levels measured from *pR_MLH1_F* in all cell lines are significantly greater than those from *pR_F* (7.4-, 161.0- and 30.0-fold the levels from the empty vector in HeLa, NCM460 and HCT116 cells, respectively).

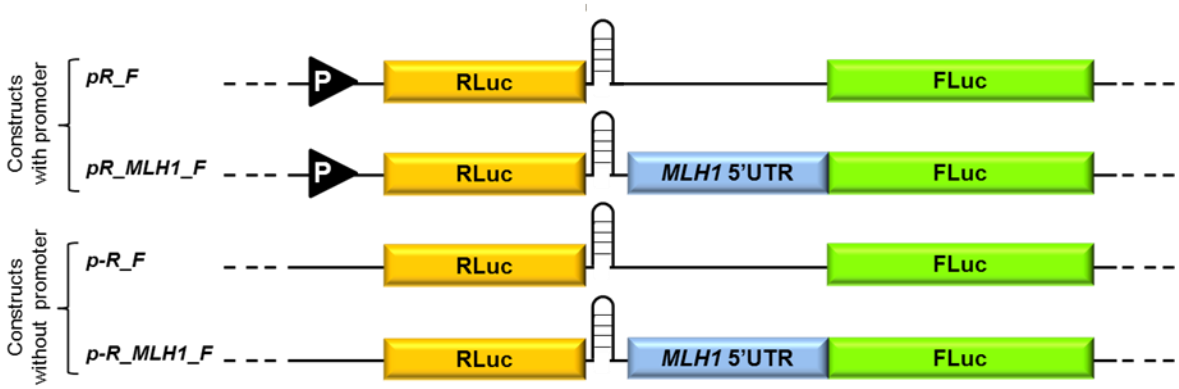


Figure 3.51 — Schematic representation of the constructs used to evaluate *MLH1* 5'UTR cryptic promoter activity in different cell lines. RLuc is the *Renilla* luciferase cap-dependent translated cistron (yellow box) and FLuc the firefly cap-independent translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Blue boxes represent the *MLH1* 5'UTR sequences cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_MLH1_F*, the *MLH1* 5'UTR-containing vector, is the sequence under study. *p-R_F* and *p-R_MLH1_F* are the counterpart promoterless plasmids. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

Furthermore, relative FLuc expression levels from *pR_MLH1_F* in NCM460 cells were also significantly greater than those in cancer cell lines. Out of these, relative FLuc expression levels from such plasmid expressed in HCT116 cells are significantly greater than those measured in HeLa cells, although they are in the same order of magnitude (figure 3.52.B). The pattern of FLuc expression from promoterless plasmids is similar to that from promoter-containing plasmids (20.8-, 234.5- and 29.1-fold the levels from the empty vector in HeLa, NCM460 and HCT116 cells, respectively) and hence the differences registered among cell lines might be a consequence of different levels of promoter activity in each cell line (figure 3.52.B). These results indicate that the cryptic promoter included in *MLH1* 5' flanking region is more active in normal than in cancer cells, and in colorectal cancer-derived cells is more active than in cervical cancer-derived cells. This suggests that the cryptic promoter activity observed within *MLH1* 5'UTR may vary depending on the tissue in which it is being expressed.

IV.3. Cryptic promoter activity is reduced in the presence of colorectal cancer-related mutations within *MLH1* 5'UTR in cancer cells but not in normal colon mucosa-derived cells

There are several evidence in the literature concerning mutations or polymorphisms within *MLH1* 5'UTR proven to be associated with a colorectal cancer phenotype. In order to understand whether such sequence modifications would alter cryptic promoter activity, we mutated the *MLH1* 5'UTR wild-type sequence at nt -28 (mutation c.-28A>T, Isidro *et al.*, 2003),

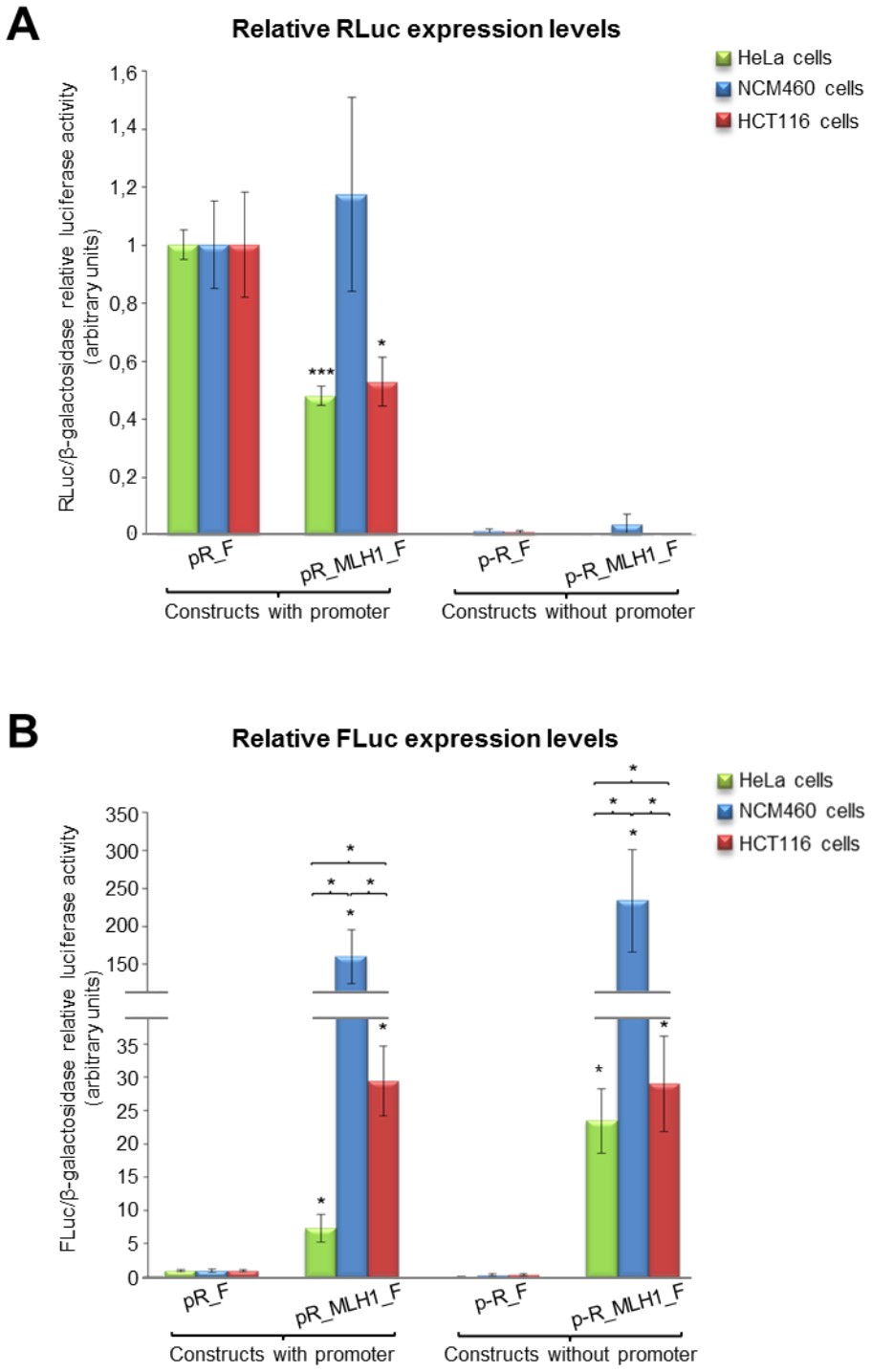


Figure 3.52 — Cryptic promoter within *MLH1* 5'UTR is more active in NCM460 cells than in HeLa or HCT116 cells. HeLa (green bars), NCM460 (blue bars) and HCT116 (red bars) cells were transfected with promoter-containing constructs (*pR_F* and *pR_MLH1_F*) or promoterless constructs (*p-R_F* and *p-R_MLH1_F*), and co-transfected with *β-galactosidase*-encoding plasmid (*pSV-β-Galactosidase* control vector), an internal control for transfection efficiency in mammalian cells. Relative RLuc (A) and FLuc (B) expression levels were obtained by normalising each of them to those from β-galactosidase-expressing plasmid, all measured by luminometry assays. Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector or the indicated construct. *P<0.05, **P<0.01, ***P<0.001

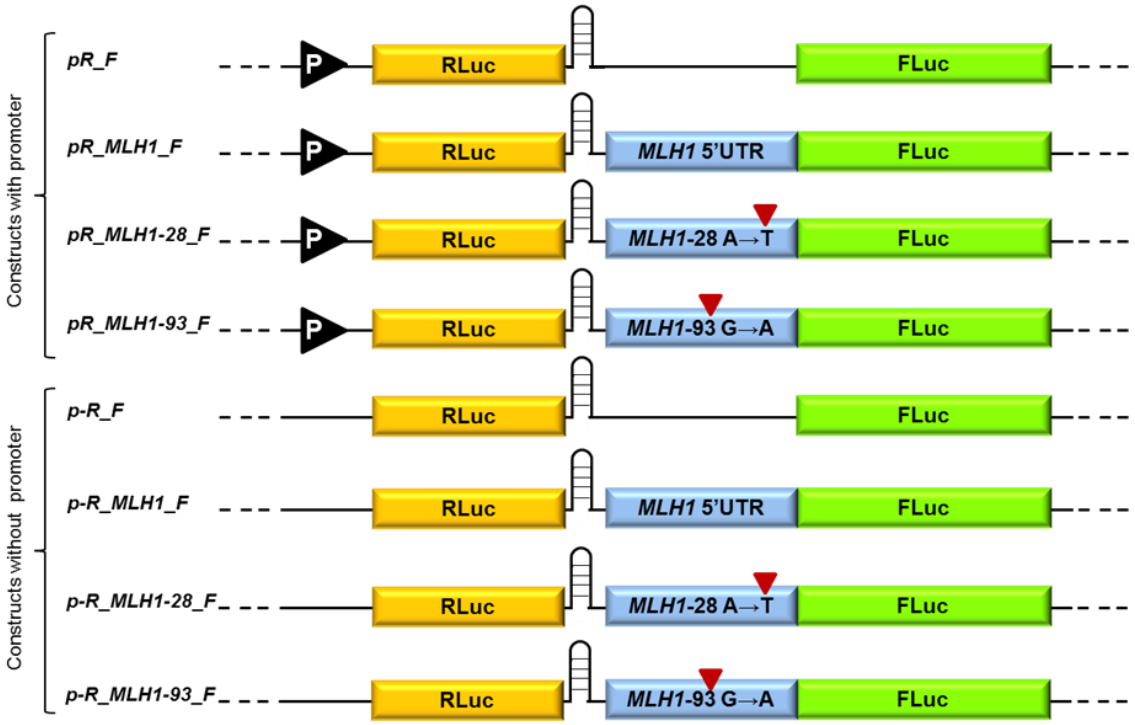


Figure 3.53 — Schematic representation of the constructs used to evaluate the effect of c.-28A>T mutation and c.-93G>A single nucleotide polymorphism within *MLH1* 5'UTR in cryptic promoter activity in different cell lines. RLuc is the *Renilla* luciferase cap-dependent translated cistron (yellow box) and FLuc the firefly cap-independent translated cistron (green box). Black triangles with white "P" symbolise the SV40 promoter. Blue boxes represent the *MLH1* 5'UTR sequences cloned upstream FLuc ATG. *pR_F* is the empty vector; *pR_MLH1_F*, *MLH1* 5'UTR-containing vector, is the sequence under study; *pR_MLH1-28_F*, *MLH1* 5'UTR-containing vector mutated at nucleotide -28 of the 5'UTR, with the mutation c.-28A>T; and *pR_MLH1-93_F*, *MLH1* 5'UTR-containing vector mutated at nucleotide -93 of the 5'UTR, is the sequence containing the polymorphism c.-93G>A. *p-R_F*, *p-R_MLH1_F*, *p-R_MLH1-28_F* and *p-R_MLH1_-93_F* are the counterpart promoterless plasmids. Red triangles indicate the relative position of the mutated nucleotide within *MLH1* 5'UTR. All constructs contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation.

or at nt -93 (single nucleotide polymorphism c.-93G>A; Mei *et al.*, 2010) in both bicistronic promoter-containing and promoterless plasmids (figure 3.53). As before we transfected HeLa, NCM460 and HCT116 cells with each of the referred constructs and co-transfected them with *β-galactosidase*-encoding plasmid. The expression levels from the latter plasmid (in absolute light units, data not shown) measured from different transfected cells were similar regardless of the co-transfected bicistronic construct, indicating that the observed difference in relative FLuc or RLuc expression levels were due to variations in their expression and not to variation in the expression of the internal control. Figures 3.54 and 3.55 show both relative RLuc and FLuc expression levels from *pR_MLH1_F* and *p-R_MLH1_F* and from the plasmids containing mutated *MLH1* 5'UTR sequences.

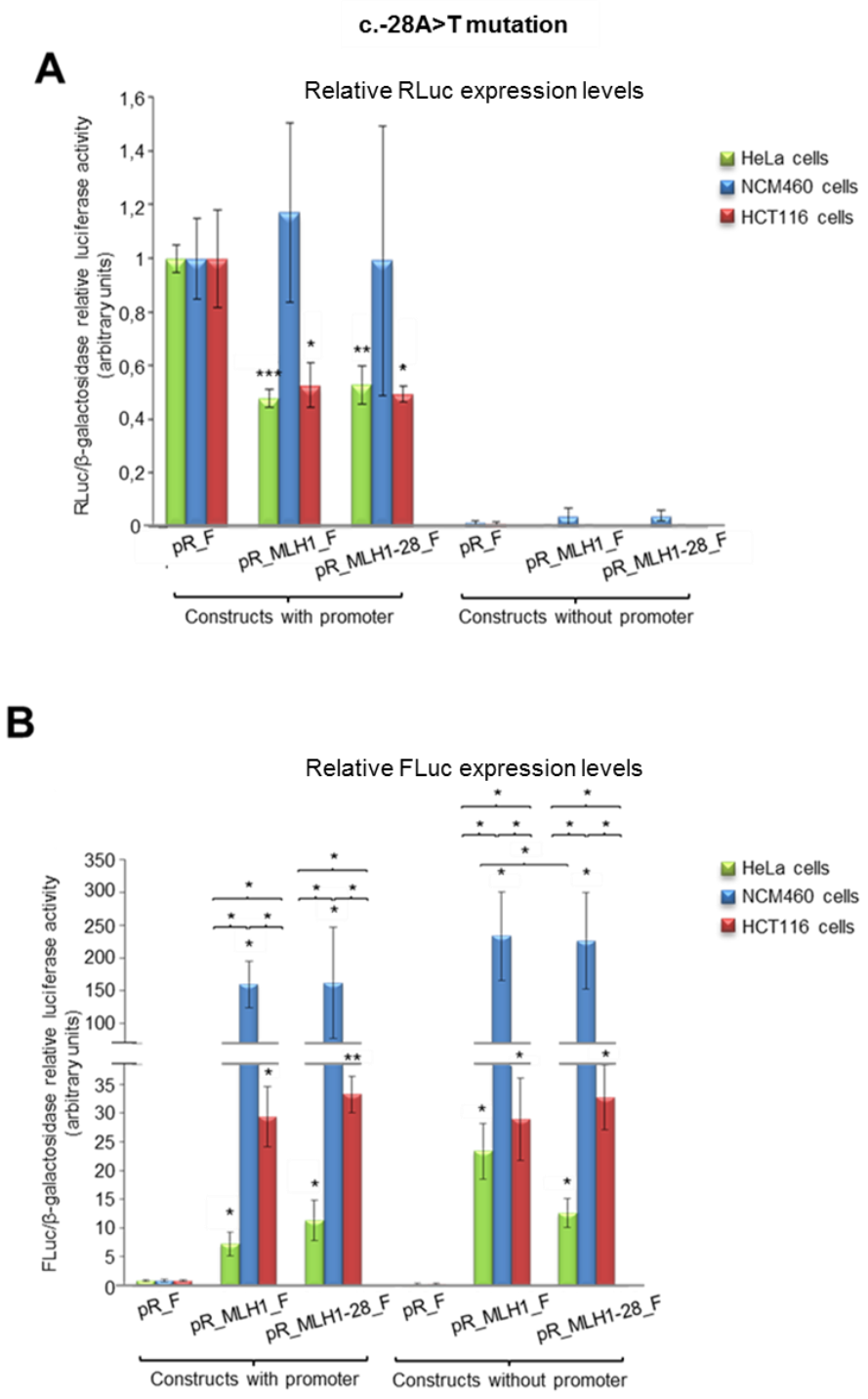


Figure 3.54 — c.-28A>T mutation within *MLH1* 5'UTR does not abolish cryptic promoter activity. HeLa (green bars), NCM460 (blue bars) and HCT116 (red bars) cells were transfected with promoter-containing constructs (*pR_F*, *pR_MLH1_F* and *pR_MLH1-28_F*) or promoterless constructs (*p-R_F*, *p-R_MLH1_F* and *p-R_MLH1-28_F*), and co-transfected with *β-galactosidase*-encoding plasmid (*pSV-β-Galactosidase* control vector), an internal control for transfection efficiency in mammalian cells. Relative RLuc (A) and FLuc (B) expression levels were obtained by normalising each of them to those from *β-galactosidase*-expressing plasmid, all measured by luminometry assays. Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector or the indicated construct. **P*<0.05, ***P*<0.01.

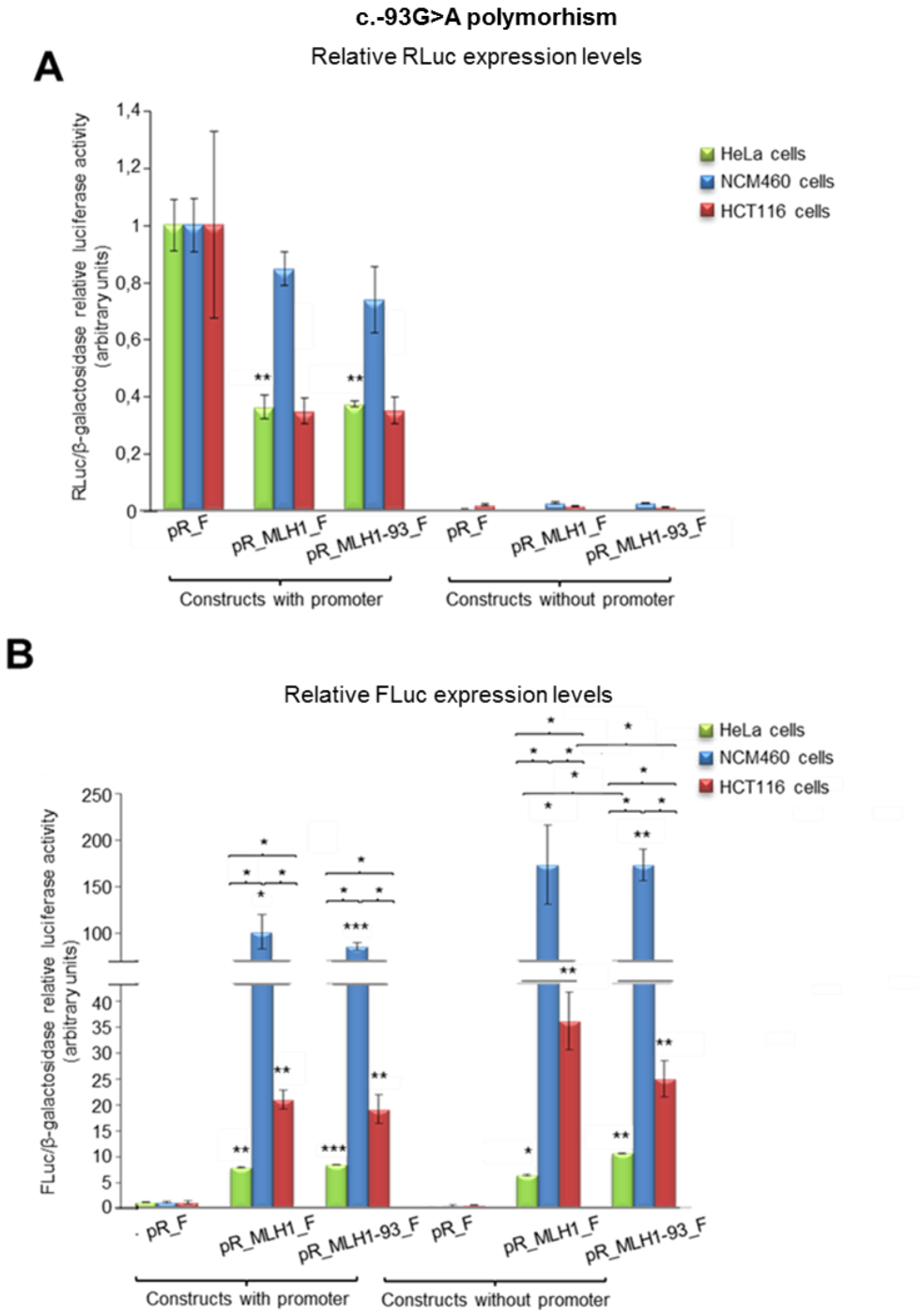


Figure 3.55 — c.-93A>T single nucleotide polymorphism within *MLH1* 5'UTR does not abolish cryptic promoter activity. HeLa (green bars), NCM460 (blue bars) and HCT116 (red bars) cells were transfected with promoter-containing constructs (*pR_F*, *pR_MLH1_F* and *pR_MLH1-93_F*) or promoterless constructs (*p-R_F*, *p-R_MLH1_F* and *p-R_MLH1-93_F*), and co-transfected with *β-galactosidase*-encoding plasmid (*pSV-β-Galactosidase* control vector), an internal control for transfection efficiency in mammalian cells. Relative RLuc (A) and FLuc (B) expression levels were obtained by normalising each of them to those from *β-galactosidase*-expressing plasmid, all measured by luminometry assays. Presented data are the result of at least three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector or the indicated construct. **P*<0.05, ***P*<0.01, ****P*<0.001

Relative luciferase expression levels from *pR_MLH1_F* and *p-R_MLH1_F* were compared to those from the counterpart promoter-containing empty plasmid and the obtained results were similar to our previous results (figure 3.52). As for relative RLuc expression levels from promoter-containing plasmids containing *MLH1* 5'UTR mutated at nt -28 or -93 (figures 3.54.A and 3.55.A, respectively), they were similar to those from *pR_MLH1_F*, that is to say similar to those from *pR_F* in NCM460 cells, but significantly lower than those from *pR_F* in HeLa and HCT116 cells. The relative RLuc expression levels from the counterpart promoterless plasmids, as expected, were significantly reduced to background levels, because the absence of the promoter sequence does not allow transcription and subsequent canonical translation of the bicistronic plasmid. Regarding relative FLuc expression levels, we observed that the levels measured from all *MLH1* 5'UTR-containing plasmids — either promoter-containing or promoterless, and either containing wild-type or mutated sequences — are significantly greater than those from *pR_F* in every tested cell line, and that, in NCM460 cells, these levels are significantly greater than those from HeLa and HCT116 cells (figures 3.54.B and 3.55.B), similar to the previously obtained results shown in figure 3.52. In colon-derived cells (NCM460 and HCT116), the presence of the mutation at nt -28 of *MLH1* 5'UTR does not alter relative FLuc expression levels from the respective plasmids. However, in HeLa cells, we observed a significant decrease in relative FLuc expression levels from *p-R_MLH1-28_F* compared to those from *pR_MLH1_F*, which does not occur from constructs with promoter (figure 3.54.B). Conversely, in the same cell line, relative FLuc expression levels from *p-R_MLH1-93_F* are significantly greater than those from *p-R_MLH1_F*. Furthermore, we observed a decrease in relative FLuc expression levels from *p-MLH1_F* compared to those from *p-MLH1_F*, in HCT116 cells. However, such differences were not observed in the relative FLuc expression levels from promoter-containing constructs (figure 3.55.B).

In sum, these results indicate that the presence of such colorectal cancer-related mutation and polymorphism within *MLH1* 5'UTR may alter gene expression in cancer cells (HeLa and HCT116) but not in normal ones (NCM460). The differences in relative FLuc expression levels observed between promoterless constructs but not between constructs with promoter may be explained by some contribution from a non-canonical translation initiation mechanism occurring in the latter, as will be further analysed.

IV.4. *MLH1* 5'UTR seems to mediate a non-canonical translation initiation mechanism

In order to detect a putative non-canonical translation initiation mechanism mediated by *MLH1* 5'UTR, we produced an *in vitro* transcribed, capped and polyadenylated bicistronic mRNA containing *MLH1* 5'UTR and transfected HeLa, NCM460 and HCT116 cells with either this transcript or each of the control counterparts (figure 3.56.A), as in previous experiments (cf. section II).

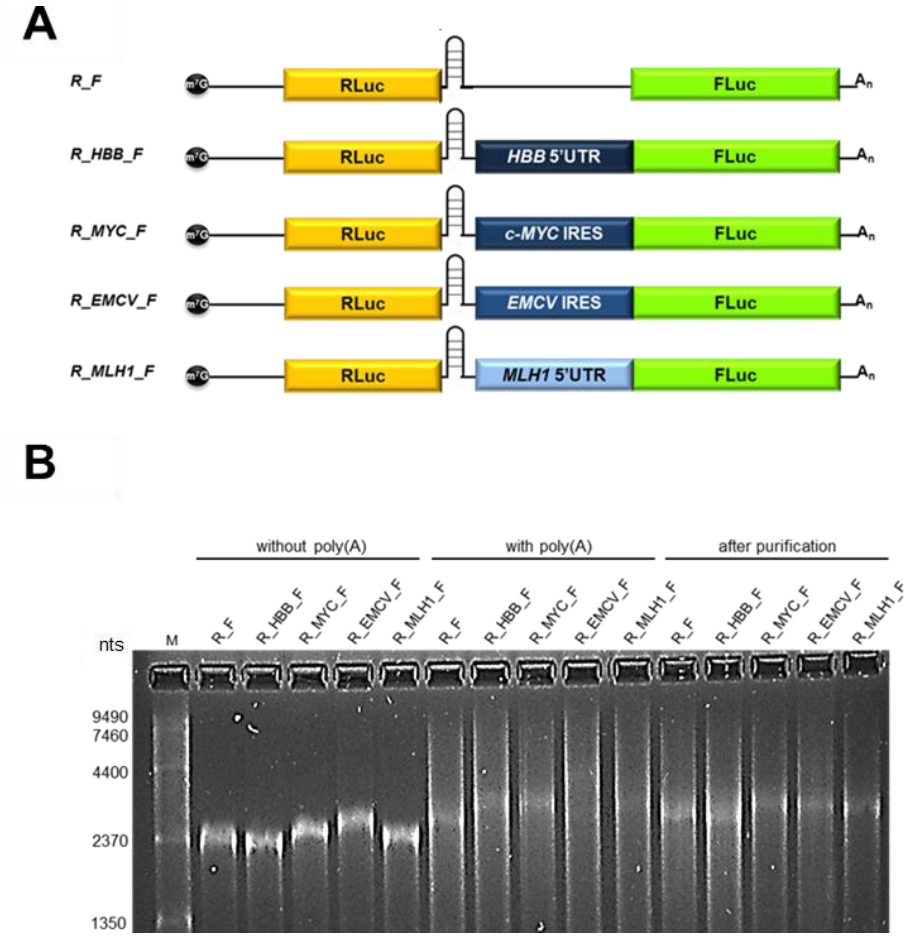


Figure 3.56 — *In vitro* transcribed mRNA used to detect *MLH1* 5'UTR-mediated cap-independent translation activity in a bicistronic context. (A) Schematic representation of the capped (m^7G) and polyadenylated (A_n) *in vitro* transcribed bicistronic mRNA produced from the corresponding plasmids. RLuc is the *Renilla* luciferase cap-dependent translated cistron (yellow box) and FLuc the firefly luciferase cap-independent translated cistron (green box). Boxes in different shades of blue represent the different sequences cloned upstream FLuc AUG. *R_F* is the empty transcript; *R_HBB_F,* the human β-globin (*HBB*) 5'UTR-containing transcript, is the negative control for cap-independent activity; *R_MYC_F,* the *c-MYC* IRES-containing transcript, is the cellular positive control for cap-independent activity; *R_EMCV_F,* the EMCV IRES-containing transcript, is the viral positive control for cap-independent activity; *R_MLH1_F,* the *MLH1* 5'UTR-containing transcript, is the sequence under study. All transcripts contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation. (B) Denaturing agarose–formaldehyde gel showing the integrity of the produced mRNA. Expected transcript sizes before polyadenylation: 3004nts (*R_F*); 3054nts (*R_HBB_F*); 3344nts (*R_MYC_F*); 3585nts (R_*EMCV*_F); 3202nts (*R_MLH1_F*). M: 0.24-9.5 Kb RNA Ladder (Invitrogen); nts: weight, in bases, of RNA ladder bands. Without poly(A): *in vitro* capped transcripts before polyadenylation. With poly(A): *in vitro* capped transcripts after polyadenylation. After purification: *in vitro* capped and polyadenylated transcripts after RNA purification by phenol–chloroform extraction.

Prior to transfection, we evaluated the integrity of the produced transcripts in a denaturing agarose–formaldehyde gel electrophoresis (figure 3.56.B). Four hours after transfection we assessed the relative FLuc expression levels in every tested cell line (figure 3.57). We observed

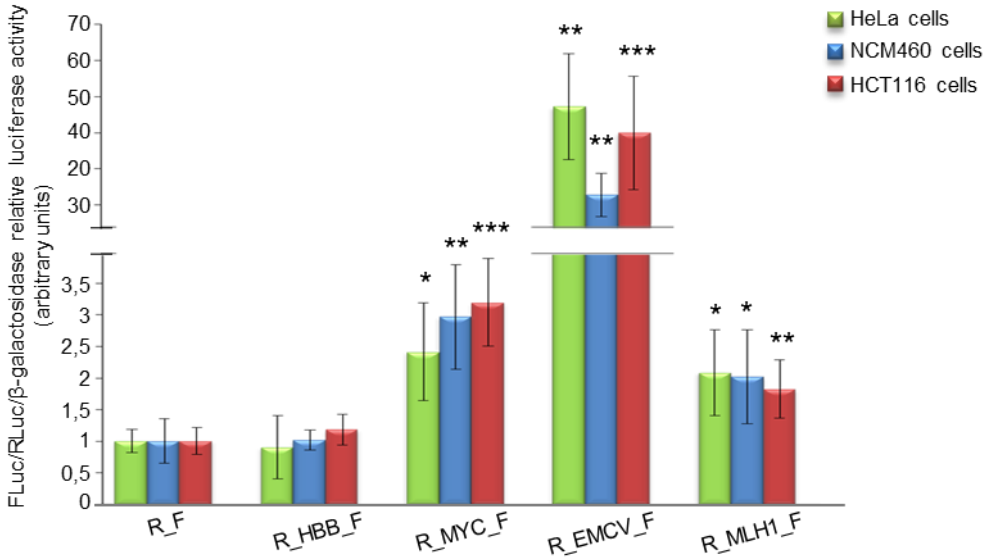


Figure 3.57 — *MLH1* 5'UTR mediates non-canonical translation initiation in HeLa, NCM460 and HCT116 cell lines transfected with bicistronic transcripts. HeLa (green bars), NCM460 (blue bars), and HCT116 (red bars) cells were transfected with *in vitro* transcribed mRNA containing either *MLH1* 5'UTR (*R_MLH1_F*) or one of the controls used in the experiment: *R_F* (empty transcript), *R_HBB_F* (*HBB* 5'UTR-containing transcript), *R_MYC_F* (*c-Myc* IRES-containing transcript), or *R_EMCV_F* (*EMCV* IRES-containing transcript). Presented data are the result of, at least, three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector. **P*<0.05, ***P*<0.01, ****P*<0.001

a significant increase in relative FLuc expression levels from *MLH1* 5'UTR-containing transcripts compared to those from the empty transcript — 2.1-, 2.0- and 1.8-fold in HeLa, NCM460 and HCT116 cells, respectively. Relative FLuc expression levels from all the controls used in this experiement were in agreement with previously obtained results (cf. section II). These results suggest therefore that *MLH1* 5'UTR is able to mediate a non-canonical cap-independent mechanism of translation initiation used for driving FLuc translation in a bicistronic context.

IV.5. *MLH1* 5'UTR-mediated FLuc translation seems to be maintained in HeLa and HCT116 cells under some stress conditions

To evaluate whether *MLH1* 5'UTR is able to mediate non-canonical translation initiation under stress conditions impairing canonical cap-dependent translation initiation, we transfected HeLa, NCM460 and HCT116 cells with *MLH1* 5'UTR-containing transcripts — or each of the counterpart control transcripts — and, 2 h posttransfection, subjected cells to several external stimuli known to impair cap-dependent translation initiation. Thus, similar to what has been previously done for testing *UPF1* and *AGO1* 5'UTR activity under stress conditions, we treated cells for 6 h with rapamycin (impairs mTOR kinase activity), $CoCl_2$ (induces chemical hypoxia) or thapsigargin (induces ER stress).

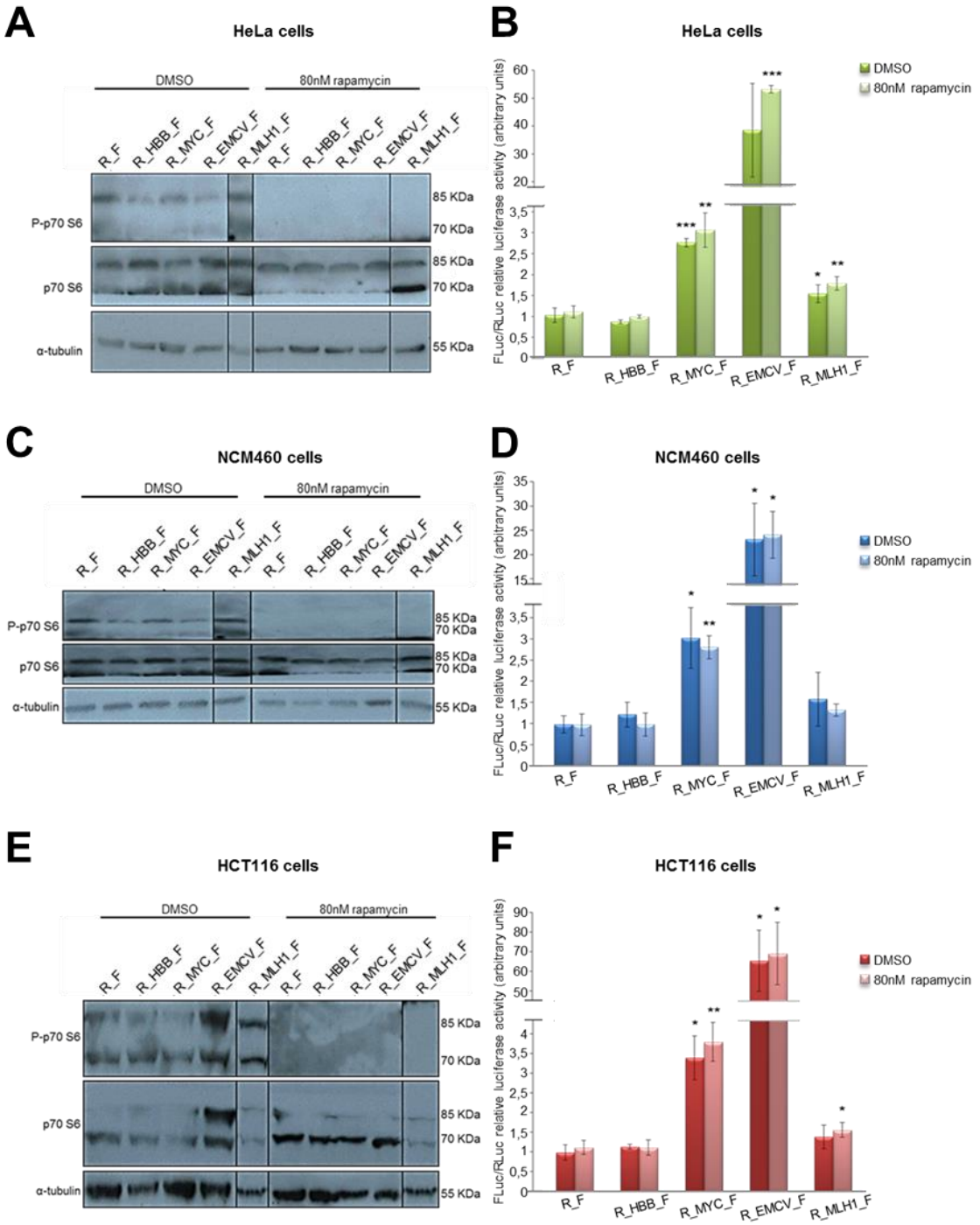


Figure 3.58 — *MLH1 5'UTR*-mediated non-canonical translation initiation is maintained under conditions impairing mTOR kinase activity in cancer but not in normal cells. HeLa (A, B), NCM460 (C, D) and HCT116 (E, F) cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F, *R_MLH1_F*) and then subjected to a 6-h-treatment with 80 nM rapamycin 2 h posttransfection. (A, C, E) Western blot against phosphorylated and non-phosphorylated S6K protein — no expression of the former indicates mTOR kinase activity on its downstream targets is blocked. "DMSO" indicates cells in control conditions and "80nM rapamycin" indicates cells treated with the drug. α-tubulin was used as a loading control for the amount of protein. (B, D, F) Relative FLuc expression levels from each transcript. Dark green, dark blue and red bars represent relative FLuc expression in control conditions; light green, light blue and pink bars represent relative FLuc expression under treatment with rapamycin. Asterisks (*) indicate statistical significance in relation to the empty counterpart transcript in control conditions. **P*<0.05, ***P*<0.01, ****P*<0.001.

In figure 3.58.A, C and E, we observed that treatment with 80 nM rapamycin blocks phosphorylation of S6K protein in every tested cell lines, as no phosphorylated protein was detected by Western blot analysis. Regarding relative luciferase activity, in HeLa cells (figure 3.58.B), we observed a significant increase in relative FLuc expression levels from *R_MLH1_F* compared to *R_F*, arbitrarily set to 1, both in control conditions (1.5-fold) and in cells treated with rapamycin (1.8-fold). On the other hand, in NCM460 cells (figure 3.58.D), the relative FLuc expression levels from transcripts both in normal and stress conditions did not significantly increase (1.6-fold and 1.3-fold, respectively), compared to those from *R_F* in control conditions. Additionally, in HCT116 cells (figure 3.58.F), relative FLuc expression levels from *R_MLH1_F* in cells treated with DMSO (control conditions) were not significantly greater than those from the counterpart empty transcript (1.4-fold), whereas those in cells treated with rapamycin were significantly greater than those from the empty transcript (1.6-fold). These results are inconsistent among them and do not reflect the outcome from previous experiments (figure 3.57), in which we observe a significant 2-fold increase in relative FLuc expression levels from *R_MLH1_F* compared to *R_F*). This discrepancy may be explained by the fact that, in order to evaluate the effect of the drug, cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA for at least 8 h (6-h-long treatment, 2 h posttransfection), which may account for some degradation of the *in vitro* produced transcript due to a long transfection period. Also, the *MLH1* 5'UTR sequence in the bicistronic context may make the transcript more susceptible to degration. However, these results sustain the hypothesis that this sequence is mediating non-canonical translation initiation under conditions impairing mTOR kinase activity, in cancer cells but not in normal mucosa-derived cells.

As for the treatment with $CoCl_2$, we observed an accumulation of HIF1α protein in cells treated with the drug compared to cells treated with vehicle (H_2O), indicating a cellular hypoxic status. This was observed in every tested cell lines (figure 3.59.A, C, E). Regarding relative FLuc expression levels from *R_MLH1_F* in control conditions, we observed a significant increase in those levels in every tested cell line (1.6-fold in HeLa cells, 2.1-fold in NCM460 cells and 1.9-fold in HCT116 cells, figure 3.59.B, D and F, respectively), a result concordant with previously obtained results (figure 3.57). Under hypoxia, the relative FLuc expression levels were also significantly greater than those from *R_F* in control conditions, suggesting *MLH1* 5'UTR is able to mediate non-canonical translation initiation in every tested cell line, under hypoxia.

Regarding treatment with thapsigargin, we observed an accumulation in phosphorylated eIF2α protein in HeLa and HCT116 cell lines (figure 3.60.A and C), but not in NCM460 cells (figure 3.60.B). To further confirm the effect of thapsigargin on impairing protein synthesis, we measured the total protein content in every tested cell line treated with DMSO and thapsigargin and observed a significant decrease in cellular protein content from control conditions to stress conditions after a 6-h-treatment (figure 3.60.D, E and F). The relative FLuc expression from *R_MLH1_F* compared to those from *R_F* in cells treated with vehicle (DMSO), were similar to

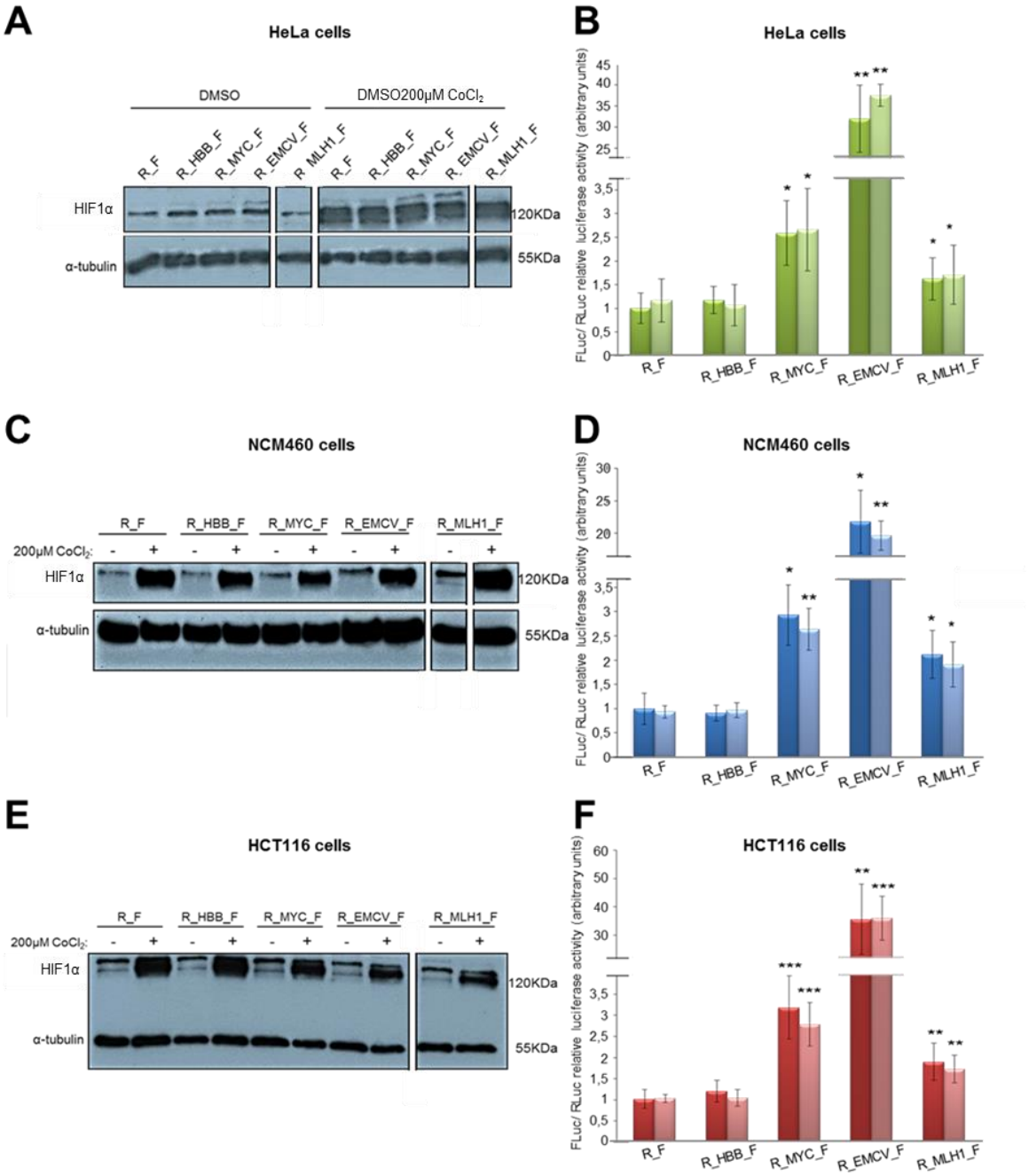


Figure 3.59 — *MLH1* 5'UTR mediates non-canonical translation initiation under hypoxia in HeLa, NCM460 and HCT116 cells. HeLa (A, B), NCM460 (C, D) and HCT116 (E, F) cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F, *R_MLH1_F*) and then subjected to a 6-h-treatment with 200 μM of $CoCl_2$ 2 h posttransfection. (A, C, E) Western blot against HIF1α, whose increased expression reflects a cellular hypoxic status. (-) indicates treatment with H_2O and (+) indicates treatment with 200 μM $CoCl_2$. α-tubulin was used as a loading control for the amount of protein. (B, D, F) Relative FLuc expression levels from each transcript. Dark green, dark blue and red bars represent relative FLuc expression in control conditions; light green, light blue and pink bars represent relative FLuc expression under hypoxia. Asterisks (*) indicate statistical significance in relation to the empty counterpart transcript in control conditions. **P*<0.05, ***P*<0.01, ****P*<0.001.

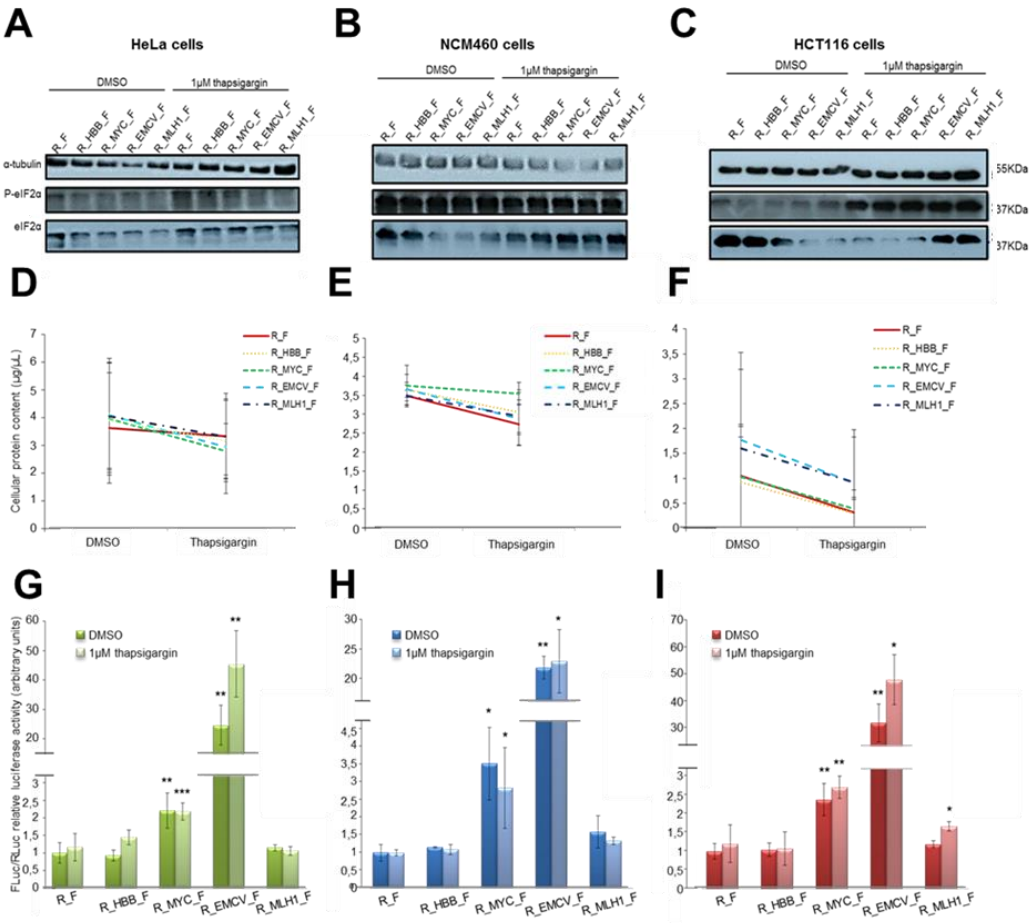


Figure 3.60 — *MLH1* 5'UTR cap-independent activity is not maintained under endoplasmic reticulum (ER) stress conditions. HeLa (A, D, G), NCM460 (B, E, H) and HCT116 (C, F, I) cells were transfected with *in vitro* transcribed, capped and polyadenylated mRNA (*R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F, *R_MLH1_F*) and then subjected to a 6-h-treatment with 1 µM thapsigargin 2 h posttransfection. (A, B, C) Western blot against phosphorylated and non-phosphorylated proteins; increased phosphorylated eIF2α expression in treated cells reflects endoplasmic reticulum (ER) stress. "DMSO" indicates cells in control conditions; "1µM thapsigargin" indicates cells treated with the drug. α-tubulin was used as a loading control for the amount of protein. (D, E, F) Cellular protein content (µg total protein/µl cell lysate) in DMSO- or thapsigargin-treated cells. (G, H, I) Relative FLuc expression levels from each transcripts. Dark green, dark blue and red bars represent relative FLuc expression in control conditions; light green, light blue and pink bars represent relative FLuc expression under ER stress. Asterisks (*) indicate statistical significance in relation to the empty counterpart transcript in control conditions. **P*<0.05, ***P*<0.01, ****P*<0.001.

those from the empty transcript (figure 3.60.G, H and I) in every tested cell line, suggesting no alternative mechanism of translation initiation is mediating FLuc translation, contrary to previously obtained results (figure 3.57).

Also, relative FLuc expression levels from *R_MLH1_F* in cells treated with thapsigargin were similar to those from *R_F* in HeLa and NCM460 cells, but significantly greater than those from *R_F* in HCT116 cells, suggesting a putative role for *MLH1* 5'UTR in mediating non-canonical

translation initiation in HCT116 cells. These results were obtained after long transfection periods (>8 h), which may account for some transcript degradation, as suggested above. The fact that the vehicle used for both thapsigargin and rapamycin is DMSO may also account for the induction of cellular stress in the cells with which the putative mechanism of non-canonical translation initiation mediated by *MLH1* 5'UTR cannot cope with. Another explanation could be that the stimulus to which cells were subjected was not effective (figure 3.60.D and E versus F, and A and B versus C).

In order to understand to what extent the putative non-canonical translation initiation mediated by *MLH1* 5'UTR is independent of eIF4E, the cap-binding protein, we depleted cells of eIF4E by knocking down this protein, and transfected cells with *R_MLH1_F* or the counterpart control transcripts for 4 h. Western blot analysis revealed a reduction in eIF4E protein amount in cells transfected with siRNA against eIF4E, but not in cells transfected with siRNA against GFP (control), indicating an efficient knock-down of eIF4E protein (figure 3.61.A). The relative FLuc expression levels from *R_MLH1_F* in Hela cells (figure 3.61.B) were significantly greater than those from *R_F* in both control conditions and eIF4E depletion (1.9- and 2.3-fold increase, respectively). This result indicates *MLH1* 5'UTR is actually able to mediate internal, non-canonical translation initiation of FLuc in cells depleted of eIF4E protein, suggesting it may work in a cap-independent manner.

To further characterise the putative *MLH1* 5'UTR-mediated non-canonical mechanism of translation initiation dependency on the cap structure, we produced *in vitro* transcribed and polyadenylated monocistronic mRNA with or without a functional cap structure, containing *MLH1* 5'UTR or the counterpart control sequences, as depicted in figure 3.62. We transfected HeLa cells with the aforementioned transcripts and co-transfected them with *β-galactosidase*-enconding plasmid for 4 h. We evaluated the relative FLuc expression levels from 5'A-capped transcripts in relation to their counterpart 5'G-capped transcripts by luminometry assays. From this experiment, we observed a significant 5.1-fold increase in relative FLuc expression in uncapped *MLH1* 5'UTR-containing transcript compared to the counterpart empty transcript (figure 3.63). As before the positive controls presented a significant increase in relative FLuc expression levels in relation to the counterpart empty transcript, whereas the negative controls did not. This result indicates *MLH1* 5'UTR is, in fact, able to mediate efficient cap-independent translation initiation in a free 5' end transcript. This result also helps explaining the inconsistent results under stress conditions and long transfection periods, because it indicates *MLH1* 5'UTR need a free 5' end to mediate cap-independent translation initiation, which does not occur in the bicistronic system. This suggests *MLH1* 5'UTR is able to mediate cap-independent, but free 5' end-dependent translation initiation.

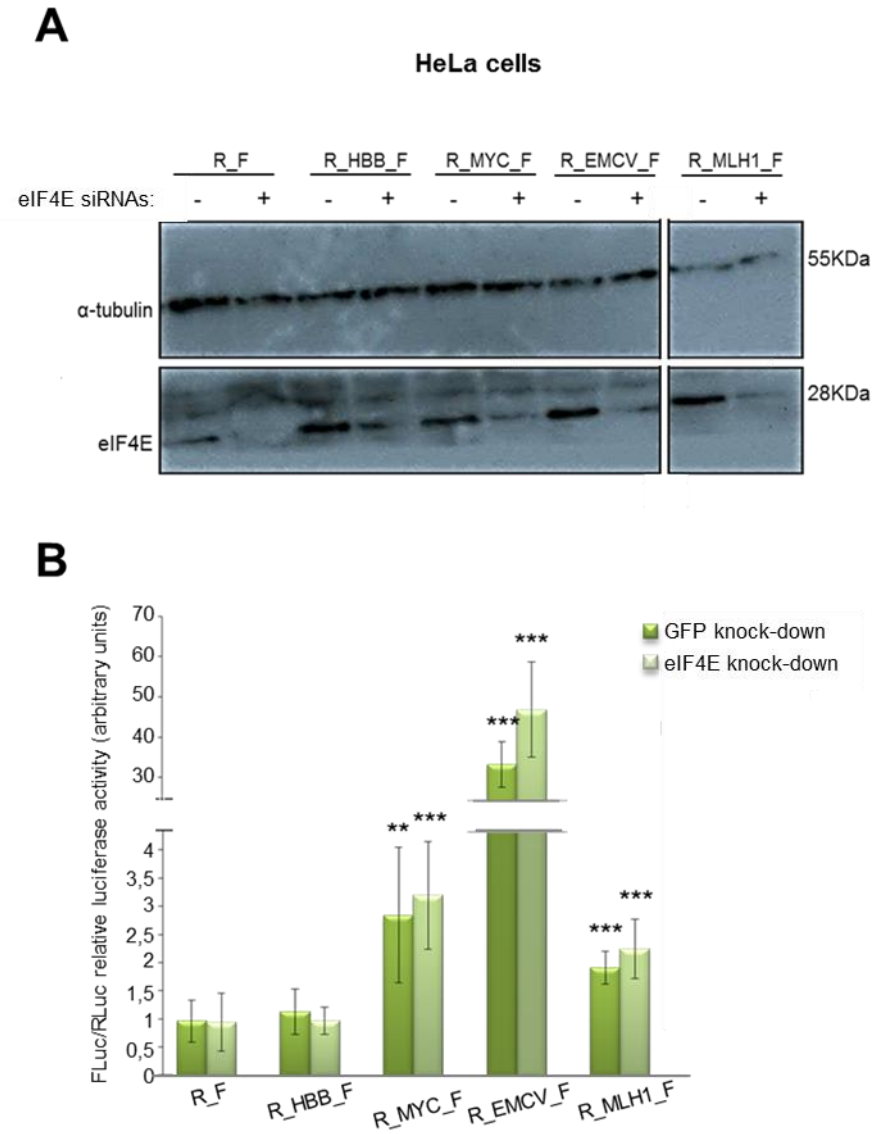


Figure 3.61 — *MLH1* 5'UTR-mediated translation is maintained after knock-down of eIF4E protein. HeLa cells were transfected with *in vitro* transcribed, capped and polyadenylated bicistronic mRNA after knocking down eIF4E subunit. (A) Western blot against eIF4E showing its knock-down efficiency. (-) indicates GFP siRNA transfection and (+) indicates eIF4E siRNA transfection. α-tubulin was used as a loading control for the amount of protein. (B) Relative luciferase activity measured from *R_F*, *R_HBB_F*, *R_MYC_F*, R_*EMCV*_F and *R_UPF1_F* transcripts. Dark green bars indicate conditions of GFP siRNA cellular treatment and light green bars indicate conditions of eIF4E siRNA cellular treatment. Asterisks (*) indicate statistical significance in relation to the empty counterpart in control conditions (GFP siRNA). ***P*<0.01, ****P*<0.001

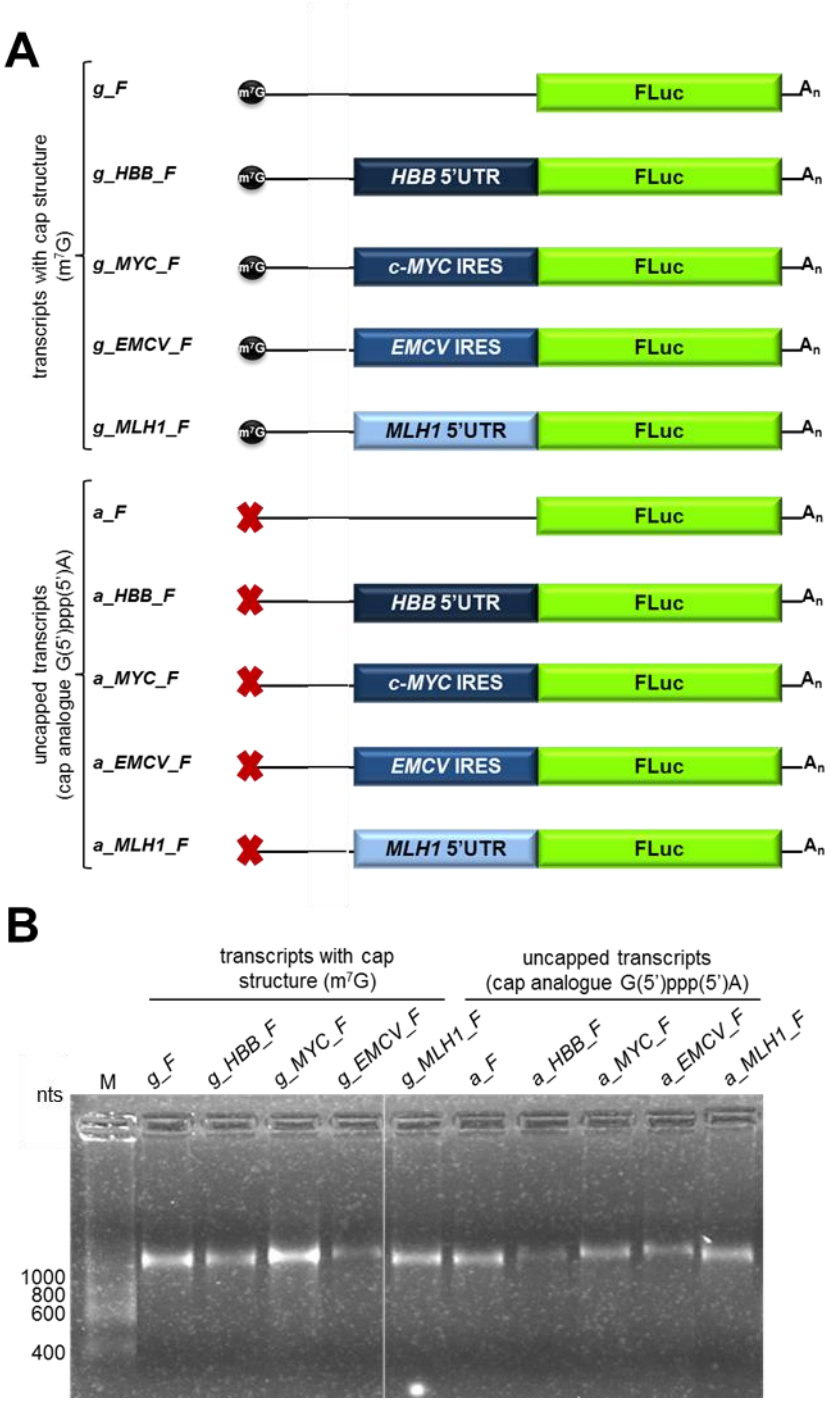


Figure 3.62 — *In vitro* transcribed and polyadenylated monocistronic transcripts used to evaluate *MLH1* 5'UTR-mediated translation dependency on a free 5' end. (A) Schematic representation of the *in vitro* transcribed and polyadenylated (A_n) monocistronic transcripts. *g_F* (empty transcript), *g_HBB_F* (*HBB* 5'UTR-containing transcript), *g_MYC_F* (*c-Myc* IRES-containing transcript), *g_EMCV_F* (EMCV IRES-containing transcript) and *g_MLH1_F* (*MLH1* 5'UTR-containing transcript) are the capped (m^7G, black circles) transcripts, and *a_F*, *a_HBB_F*, *a_MYC_F*, *a_EMCV_F* and *a_MLH1_F* are the counterpart uncapped [cap analogue G(5')ppp(5')A, red crosses] transcripts. FLuc is the firefly luciferase enconding cistron (green box). Boxes in different shades of blue represent different sequences cloned upstream FLuc AUG. (B) Denaturing agarose–formaldehyde gel showing the integrity of the *in vitro* transcribed and polyadenylated monocistronic transcripts. M: RiboRuler Low Range RNA Ladder (Thermo Fisher Scientific); nts: weight, in bases, of RNA ladder bands.

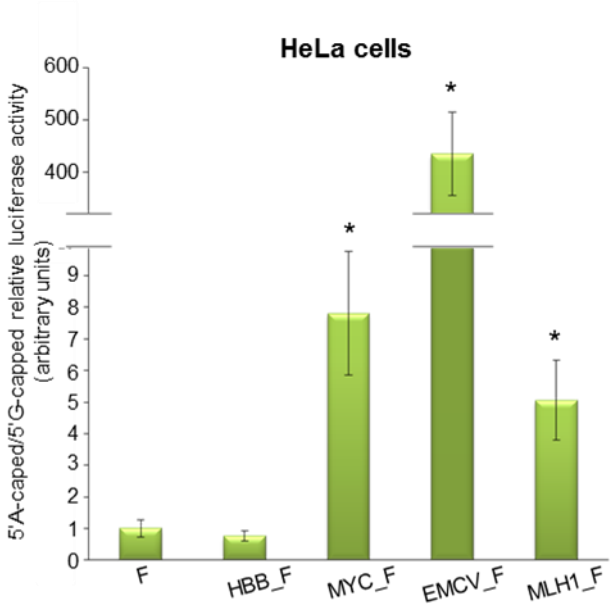


Figure 3.63 — *MLH1* 5'UTR mediates cap-independent translation in monocistronic transcripts lacking the cap structure. HeLa cells were transfected with *in vitro* transcribed monocistronic mRNA, either capped (5'G-capped, m^7G) or uncapped [5'A-capped, cap analogue G(5')ppp(5')A], containing either *MLH1* 5'UTR (*MLH1_F*) or one of the controls used in the experiment: *F* (empty transcript), *HBB_F* (*HBB* 5'UTR-containing transcript), *MYC_F* (*c-Myc* IRES-containing transcript), or *EMCV_F* (*EMCV* IRES-containing transcript). Presented data are the result of, at least, three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty transcripts. **P*<0.05

IV.6. Mutation c.-28A>T and polymorphism c.-93G>A within *MLH1* 5'UTR have different roles in non-canonical translation initiation

We produced *in vitro* transcribed, capped and polyadenylated bicitronic mRNA containing *MLH1* 5'UTR sequences — wild-type, mutated at nt -28 (c.-28A>T), and mutated at nt -93 (c.-93G>A) — to evaluate how translation initiation mediated by *MLH1* 5'UTR may be affected by the presence of such mutations (figure 3.64.A). The integrity of the transcripts was confirmed in a denaturing agarose–formaldehyde gel electrophoresis prior to mRNA transfection (figure 3.64.B). We transfected HeLa, NCM460 and HCT116 cells with the aforementioned mRNA for 4 h and assessed relative FLuc expression levels from each transcript by luminometry assays. In figure 3.65.A, we observed a significant 2-fold increase in relative FLuc expression levels from *MLH1* 5'UTR-containing transcripts compared to those from *R_F*, in every tested cell line. As for the relative FLuc expression levels from *R_MLH1-28_F*, we observed these were similar to those from *R_MLH1_F* in HeLa and HCT116, but significantly lower than the latter in NCM460 cells. This suggests that *MLH1* 5'UTR mutated at nt -28 loses the ability to mediate cap-independent translation initiation in NCM460 but not in cancer cells.

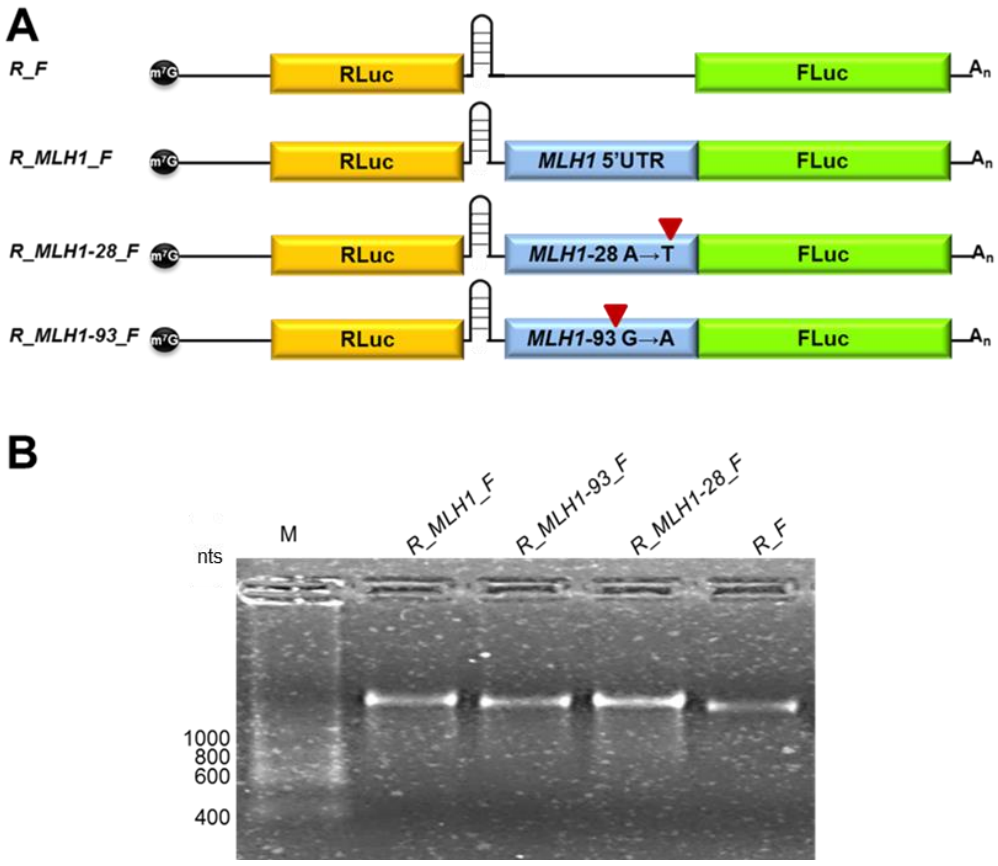


Figure 3.64 — *In vitro* transcribed mRNA used to assess to effect of mutation c.-28A>T and polymorphism c.-93G>A on *MLH1* 5'UTR-mediated cap-independent translation initiation activity in a bicistronic context. (A) Schematic representation of the capped (m^7G) and polyadenylated (A_n) *in vitro* transcribed bicistronic mRNA produced from the corresponding plasmids. RLuc is the *Renilla* luciferase cap-dependent translated cistron (yellow box) and FLuc the firefly luciferase cap-independent translated cistron (green box). Blue boxes represent the different sequences cloned upstream FLuc AUG. *R_F* is the empty transcript, *R_MLH1_F,* is the *MLH1* 5'UTR-containing transcript, *R_MLH1-28_F* is the *MLH1* 5'UTR-containing transcript with mutation c.-28A>T*; R_MLH1-93_F* is the *MLH1* 5'UTR-containinig transcript with polymorphism c.-93G>A. Red triangles indicate the relative position of the mutated nucleotide within *MLH1* 5'UTR. All transcripts contain a stable hairpin downstream RLuc cistron to prevent translation reinitiation. (B) Denaturing agarose**–**formaldehyde gel showing the integrity of the produced capped and polyadenylated mRNA. M: 0.24-9.5 Kb RNA Ladder (Invitrogen); nts: weight, in base pairs, of RNA ladder bands.

Concerning *R_MLH1-93_F*, our results were inconclusive, as no cap-independent activity was detected from the transcript containing the wild-type sequence. This may reflect, again, the inability of *MLH1* 5'UTR to mediate cap-independent translation initiation when the 5' end of the transcript is not free for ribosome binding.

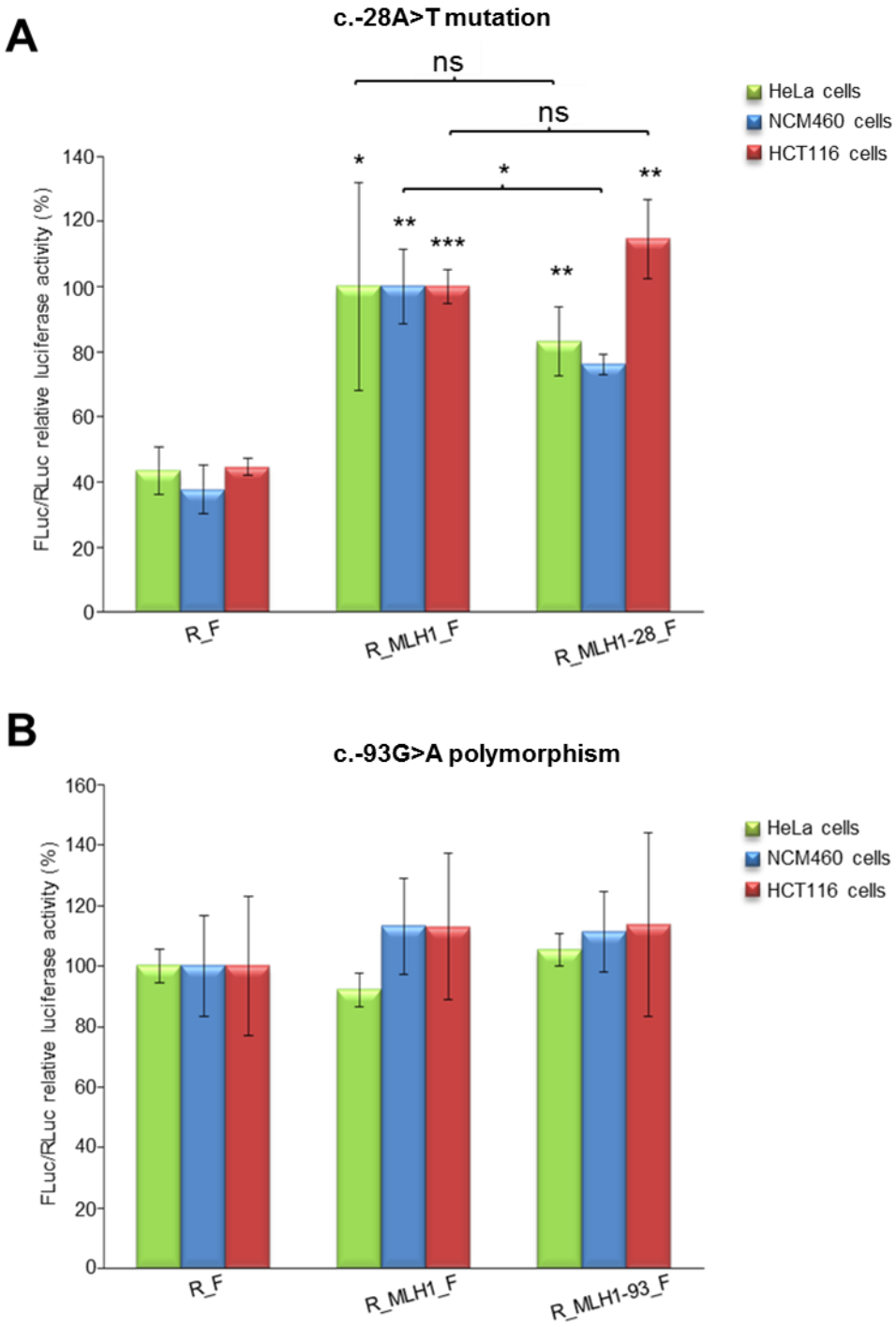


Figure 3.65 — Mutation at nucleotide -28 of *MLH1* 5'UTR reduces translation efficiency in NCM460 cells. HeLa (green bars), NCM460 (blue bars), and HCT116 (red bars) cells were transfected with *in vitro* transcribed mRNA containing either wild-type *MLH1* 5'UTR (*R_MLH1_F*), *R_F*, or one of the altered sequences used in the experiment: *R_MLH1-28_F*, the *MLH1* 5'UTR-containing transcript mutated at nucleotide -28 of *MLH1* 5'UTR (A); *R_MLH1-93_F*, the *MLH1* 5'UTR-containing transcript mutated at nucleotide -93 of *MLH1* 5'UTR (B). Presented data are the result of, at least, three independent experiments. Asterisks (*) indicate statistical significance in relation to the counterpart empty vector or the indicated transcript. *P<0.05, **P<0.01, ***P<0.001. ns — non-singicant.

4.

Discussion & future perspectives

As previously stated, the main goal of this work was to identify proteins that can be translated via a non-canonical mechanism of translation initiation and to understand the nature of such mechanisms. In view of the results produced during the experimental analysis of the selected transcripts, we can say that all tested sequences — human *UPF1*, *AGO1* and *MLH1* 5'UTR — have the ability to mediate translation initiation, although through different mechanisms, as suggested by experiments performed with *in vitro* transcribed bicistronic and monocistronic mRNA. According to published data, sequences that are able to mediate non-canonical translation initiation in a cap-, free 5' end-independent manner are able to internally recruit ribosomal subunits directly to the vicinity of the AUG, promoting an internal translation initiation. It is the case of IRES-mediated translation, as described for proteins like c-Myc (Stoneley *et al.*, 1998; Subkhankulova *et al.*, 2001), XIAP (Holcik *et al.*, 2003; Riley *et al.*, 2010), FGF (Gonzalez-Herrera *et al.*, 2006; Chen *et al.*, 2014), VEGF (Bornes *et al.*, 2007; Morfoisse *et al.*, 2014), etc. Sequences that mediate internal ribosome entry allow translation in a bicistronic context, as is the case of *c-Myc* IRES. The *UPF1* 5'UTR analysed in this study seems to follow a mechanism similar to that of *c-Myc* IRES to mediate cap-independent translation initiation. *In silico* analysis of the sequence shows it forms an intricate secondary structure that may help recruit the ribosome directly to the vicinity of the AUG, thus promoting an internal entry of the ribosome. Also, the 78% GC content of the sequence renders great stability to such secondary structure (ΔG=-141.35kcal/mol), and allows a great conservation of the formed stem loops among species. This suggests the formed structure may be involved in mechanisms that are extremely conserved throughout evolution, of which translation initiation is an accurate example (Mathews *et al.*, 2007). Deletional and mutational analysis of *UPF1* 5'UTR (figure 3.27) showed that predicted SLI and III are of great stability and hardly disrupted, supporting the evidence that their formation is conserved among species. Additionally, experimental verification of the *in silico* data revealed that these two loops are required for *UPF1* 5'UTR-mediated cap-independent translation initiation, and, also, that their disruption causes its elimination. Similar analyses have been performed for other sequences capable of mediating cap-independent translation initiation, such as *c-Myc* IRES (Stoneley *et al.*, 2000b), *HCV* IRES (Buratti *et al.*, 1998) or *VEGF* IRES (Stein *et al.*, 1998), showing that some segments within the corresponding 5'UTR are crucial for cap-independent activity, whereas others do not affect it. For instance, elimination of most (851 out of 1014 nts) of the internal *VEGF* 5'UTR sequence not only maintains full IRES activity but also generates a significantly more potent IRES, whose activity is abrogated by subtitution of a few bases near the 5' terminus and close to the translation start codon (Stein *et al.*, 2008). The sequence of *c-Myc* IRES used in this work as a positive control for cap-independent translation activity is the 340 nt minimal IRES sequence that retrieves full-length cap-independent activity, as described by Stoneley *et al.* (2000b). According to the deletional analysis performed for *UPF1* 5'UTR, we observed that nts 1–100 or nts 151–275 — corresponding, respectively, to the predicted SLI and III — are able to retrieve full-length cap-independent translation activity. However, when the sequence corresponding to SLII is present alongside only one of the referred sequences, there is no cap-independent translation activity, suggesting the SLII sequence may negatively regulate

cap-independent translation activity. This hypothesis is confirmed by mutational analysis. While the mutation of SLI, or III, alone abrogates cap-independent translation activity, the mutation of SLII does not. Further studies would include a mutant containing nts 1–100 and 151–275 but lacking the in-between sequence (corresponding to SLII) to assess whether a combined effect of the two segments would enhance cap-independent translation activity, similarly to what occurs in *VEGF* IRES. A recent study revealed that there are two functional classes of IRES: (i) IRES whose expression is reduced only when a specific position is mutated and (ii) IRES whose mutations in most positions greatly reduce expression (Weingarten-Gabbay *et al.*, 2016). These two classes may point to differences in the underlying mechanism of IRES activity. Either IRES can act through a short sequence motif — such as ITAF binding sites —, in which only mutations in a specific motif reduce activity (local sensitivity), or IRES activity can involve the formation of a secondary structure, in which mutations at various positions can disrupt the overall structure and result in reduced activity (global sensitivity) (Weingarten-Gabbay *et al.*, 2016). The results from deletional and mutational analyses suggest *UPF1* 5'UTR belongs to class II — IRES with global sensitivity —, as mutations in SLI and III, and deletion of a few segments, globally reduce cap-independent translation activity. Furthermore, the relative FLuc expression levels obtained from *R_UPF1_F* under stress conditions are all very robust and in line with the premise that cap-independent translation initiation can occur under canonical cap-dependent translation initiation-impairing conditions, thus working as a back-up mechanism for maintaining protein translation levels in unfavourable conditions for cap recognition and scanning (Graber and Holcik, 2007; Martínez-Salas *et al.*, 2013; Bisio, 2015). Relative FLuc expression levels in *UPF1* 5'UTR-mediated translation initiation are similar to those obtained from *c-Myc* IRES but much lower than those from *EMCV* IRES. By using a high-throughput bicistronic assay, a recent systematic analysis of sequences mediating IRES-dependent translation in human and viral genomes revealed that the fraction of sequences that mediate IRES-dependent translation is higher in viruses than in the human genome and that, in general, viral IRES are more active than human counterparts, as previously predicted (Jackson, 1991; Weingarten-Gabbay *et al.*, 2016). This is in line with our results, as the FLuc expression mediated by *EMCV* IRES, a viral IRES sequence (Bochkov and Palmenberg, 2006), is much greater than that mediated by *c-Myc* IRES, a cellular IRES sequence (Stoneley *et al.*, 1998). Considering the results obtained from *UPF1* 5'UTR, we may say this sequence is capable of mediating cellular IRES-dependent translation initiation in a manner similar to that of *c-Myc* IRES. Furthermore, this recent analysis also revealed that viral 5'UTR with IRES activity have lower GC content and higher minimal free energy in comparison to their human counterparts. On the other hand, the comparison between the GC content and minimal free energy for all active and inactive 5'UTR from both human and viral origins revealed that the active 5'UTR have lower GC content and higher minimal free energy. Due to the high content of GC in *UPF1* 5'UTR, and low minimal free energy, it is logical to assume that this sequence can actively mediate cap-independent translation in a much less efficient manner than *EMCV* IRES.

Results from cells transfected with monocistronic reporter transcripts lacking the cap structure (figure 3.33) confirmed the ability of *UPF1* 5'UTR to mediate translation initiation independent of the cap structure. The relative FLuc expression levels driven by this sequence are similar to those driven by *c-Myc* IRES but much lower than those mediated by the viral IRES, which agrees with the aforementioned comparison between viral and human IRES sequences (Weingarten-Gabbay *et al.*, 2016).

Recent data from a systematic screen for IRES-mediated translation activity have shown that about 10% of human 5'UTR have the potential to be translated by this cap-independent mechanism (Weingarten-Gabbay *et al.*, 2016). The existence of IRES in capped cellular mRNA raised the question of their pathophysiological function and of the advantage of a cap-independent translation (Jackson, 1991). Actually, several reports have demonstrated that cellular IRES function in various physiological processes including spermatogenesis, neuron plasticity, and cell differentiation (Gonzalez-Herrera *et al.*, 2006; Sonenberg and Hinnebusch, 2007; Audigier *et al.*, 2008; Conte *et al.*, 2009). Still, some reports have shown that several IRES are also active during cell cycle mitosis (Cornelis *et al.*, 2000; Pyronnet *et al.*, 2000) and apoptosis (Holcik and Sonenberg, 2005; Hsu *et al.*, 2016), or are aberrantly activated in tumour cells, and are thus involved in deregulation of gene expression in cancer (Sonenberg and Hinnebusch, 2007; Silvera *et al.*, 2010; Topisirovic and Sonenberg, 2011; Marcel *et al.*, 2013; Leprivier *et al.*, 2015). Furthermore, cellular IRES activity is stimulated during various cellular stresses when cap-dependent translation is blocked (Jackson, 1991; Holcik and Sonenberg, 2005; Bornes *et al.*, 2007; Sonenberg and Hinnebusch, 2007; Piccirillo *et al.*, 2014; Morfoisse *et al.*, 2014; Ozretić *et al.*, 2015). Hence, we may assume that cellular IRES exist to play a role at some critical moments of cell life when cap-dependent translation initiation is compromised, in order for the cell to cope with environmental changes affecting its viability. From an evolutionary point of view, it is tempting to speculate that IRES elements have evolved by random genomic events followed by natural selection when a cellular advantage was provided. Therefore, weak IRES may become stronger in the future depending on selective pressure. Similarly to cryptic promoters and alternative splice sites, which constantly evolve in the ever-changing genome (Elroy-Stein and Merrick, 2007), weak cellular IRES elements may represent an additional mechanism used to enhance physiological adaptability. Considering the functional role of UPF1 protein in several cellular mechanisms (NMD, cell cycle progression, telomere homeostasis, and others, cf. section 1.6.1), it is most likely that this protein uses such alternative mechanism of translation initiation to regulate its expression in situations in which the cellular homeostasis is affected. Our future studies will further test this hypothesis.

UPF1 5'UTR-mediated cap-independent translation activity was similar in all tested cell lines (HeLa, NCM460 and HCT116, figure 3.21). This suggests this activity is not tissue-specific but rather a ubiquitous mechanism that is present in every cellular type as a regulatory mechanism to sustain protein synthesis in situations in which the protein is required but the canonical mechanism of translation initiation is impaired. It is the case of S/G2 progression during cell cycle in which UPF1 protein is required and its expression levels are maintained (Azzalin and Lingner,

2006b). Furthermore, this protein is expressed in most tissues in medium-to-high levels (cf. section I), suggesting its regulation must occur in a similar manner in different tissues.

Further investigation on the biological relevance of *UPF1* 5'UTR-mediated cap-independent translation will be of utmost importance to understand how this mechanism may be involved in guaranteeing that crucial cellular functions, such as NMD and cell cycle progression, occur under different environmental stimuli. These include the regulation of translation in conditions, such as tumour onset and development. Also, in the future, it would be of great relevance to assess which proteins may function as ITAFs in facilitating this cap-independent translation initiation mechanism. High-throughput studies would be an asset to the identification of such proteins, particularly, to understand how they regulate UPF1 roles in the cell.

Apart from its role in controlling translation initiation, *UPF1* 5'UTR also contains a cryptic promoter within this sequence, which can initiate transcription of downstream sequences, as confirmed by the experiments performed with promoterless plasmids (figure 3.17). In an attempt to identify the region containing the promoter sequence, we performed a deletional analysis similar to previous studies in the literature. From this analysis, we observed that the removal of any segment of the whole sequence abolishes promoter activity. This suggests that either multiple transcription start sites are required for this process, or that promoter enhancers — e.g. transcription factor binding sites — are scattered throughout this region and are required for transcription initiation, as in *MLH1* (Ito *et al.*, 1999; Arita *et al.*, 2003). The ability of *UPF1* 5'UTR sequence to promote transcription may create an additional layer to its gene expression regulation. Further studies are required to understand which transcripts originate from this promoter sequence and what their biological relevance may be.

The results obtained when testing *AGO1* 5'UTR's ability to mediate non-canonical translation initiation suggest that this sequence is able to successfully drive a cap-independent mechanism of translation initiation, although different from the one described for *UPF1* 5'UTR. Transfection of cells with bicistronic plasmids containing *AGO1* 5'UTR revealed a significant 2.8-fold increase (figure 3.37) in relative FLuc expression levels in HeLa cells compared to those observed in cells transfected with the empty plasmid. Such expression levels were, however, significantly lower than those measured from *c-Myc* IRES-containing plasmid — 5.8-fold compared to the empty plasmid. This suggests that the mechanism through which *AGO1* 5'UTR mediates FLuc translation is less efficient than that used by *c-Myc* IRES. Since false-positive results were ruled out — as *AGO1* 5'UTR sequence neither contains cryptic promoters nor does it foster alternative splicing events able to mask a putative cap-independent translation activity —, we trust this sequence actually mediates a cap-independent mechanism of translation initiation. Such activity is maintained under stress conditions — impaired mTOR kinase activity by rapamycin, chemical hypoxia induced by $CoCl_2$, or ER stress induced by thapsigargin (figure 3.44), and even knock-down of eIF4E, the cap-binding protein (figure 3.42) — but is significantly inhibited in cells treated with 4EGI-1, an eIF4E-eIF4G interaction inhibitor that mimics 4E-BP function (figure 3.43). In the absence of 4EGI-1, both 4E-BP1 and eIF4G bind tightly to eIF4E. Indeed, unphosphorylated 4E-BP1 and eIF4G-I have similar affinities for eIF4E (15 nM and 27 nM, respectively) (Marcotrigiano

et al., 1999), and both ligands are pulled down in cap affinity binding experiments (Moerke *et al.*, 2007). Addition of 4EGI-1 allosterically dissociates eIF4G from eIF4E but does not affect binding of 4E-BP1. Thus, with eIF4G being unable to bind to eIF4E, eIF4G can no longer compete with 4E-BP1 — this leads to the increased binding of 4E-BP1 to eIF4E. Furthermore, if 4E-BP1 dissociates from eIF4E because of its hyperphosphorylation, 4EGI-1 substitutes 4E-BP1 in preventing eIF4G from binding to eIF4E (Sekiyama *et al.*, 2015). eIF4G protein plays a pivotal role in both cap- and IRES-dependent translations, not only for ribosome recruitment, but also for initiation codon selection. eIF4G is a scaffold protein that links the 43S ribosomal complex and mRNA. Moreover, eIF4G is extremely important in IRES-dependent translation of picornaviral mRNA through direct interactions with IRES elements (Kolupaeva *et al.*, 1998; de Breyne *et al.*, 2009). Paek *et al.* (2015) investigated a mechanism by which eIF4G would favour cap-independent translation initiation — i.e. how eIF4G finds the translation initiation codon. They discovered that a modified eIF4G containing the RNA-binding domain of MS2 coat protein can associate with the translational machinery and that tethering of the modified eIF4G at the 3′UTR of mRNA greatly stimulates translation of upstream ORFs. They also found that the eIF4G, tethered to the 3′UTR of bicistronic mRNA, stimulates translation of the second cistron. In addition, insertion of *EMCV* IRES at the 3′UTR of mRNA stimulates translation, much like the tethering of eIF4G to the 3′UTR. This evidence may help explain our results that suggest a cap-independent mechanism of translation initiation dependent on eIF4G.

On the other hand, in cells transfected with *in vitro* transcribed, capped and polyadenylated bicistronic mRNA containing *AGO1* 5'UTR, relative FLuc expression levels were similar to those from cells transfected with either empty or negative control transcripts. This result suggests that *AGO1* 5'UTR sequence is not able to mediate internal cap-independent translation initiation when it does not go through a "nuclear experience". Evidence in the literature states the *XIAP* IRES element is not active in the T7/vaccinia virus system, where the RNA is synthesized in the cytoplasm and does not enter the nucleus, suggesting that the *XIAP* IRES requires "nuclear experience" (G. Belsham, personal communication *in* Holcik *et al.*, 2003). Holcik *et al.* (2003) predicted that this nuclear event could be provided by nuclear RNA binding proteins, such as hnRNPC1 and -C2. These proteins could interact with the *XIAP* IRES RNA in the nucleus and be then transported with the *XIAP* RNA to the cytoplasm, where they would enhance *XIAP* mRNA translation. Alternatively, the binding of hnRNPC1 and -C2 to the *XIAP* IRES in the nucleus could have an impact on the conformational state of the IRES element, which would be essential to the binding of one or more auxiliary proteins involved in the translation of XIAP. The requirement for a nuclear event is not exclusive to *XIAP* IRES. Most cellular IRES elements do not function, or function very inefficiently, in cell-free translation systems or in RNA transfection assays, suggesting that they may require a nuclear event (Jackson, 2000; Stoneley *et al.*, 2000b). Although the nature of this event is yet to be understood, it is plausible that the nuclear experience of at least some IRES elements may be mediated by the hnRNPC1 and -C2 proteins. Further experiments will be necessary to determine the putative nuclear experience required for

AGO1 5'UTR-mediated cap-independent translation initiation and the role of hnRNPC1 and -C2 proteins in this process.

To further address the nature of *AGO1* 5'UTR-mediated cap-independent mechanism of translation initiation, we transfected cells with *in vitro* transcribed monocistronic mRNA lacking cap structure (figure 3.48). From this experiment, we observed that relative FLuc expression levels mediated by the *AGO1* 5'UTR are significantly higher than those from the negative controls, which indicate that this sequence can indeed mediate translation initiation in the absence of a cap structure, when the mRNA 5' end is free. The need for a free 5' end has been proven to be essential for CITE-mediated translation initiation (Shatsky *et al.*, 2010; Andreev *et al.*, 2013; Terenin *et al.*, 2013). Terenin and co-workers (2013) showed that the insertion of an eIF4G-binding element from a viral IRES into 5'UTR of strongly cap-dependent mRNA dramatically reduces their cap requirement in mammalian cells. This mechanism has been proven to be different from the internal entrance because these mRNA fail the bicistronic test, meaning they need a free 5' end for the preinitiation complex to bind. Thus, although this is a cap-independent mechanism, it is 5' end-dependent and involves special elements — CITE (Shatsky *et al.*, 2010). In CITE-mediated translation, some components of the translation apparatus, for example, eIF4G and eIF3, are able to be directly or indirectly recruited onto the 5'UTR via RNA-protein interactions with concomitant recruitment of other components of the scanning apparatus (Andreev *et al.*, 2013; Roberts *et al.*, 2015). In this way, the 5'UTR of an mRNA creates, in its vicinity, a high concentration of translational components. This also helps overcome competition for factors from other cellular mRNA. This mechanism has been described, in the human *Apaf-1* mRNA, as being able to initiate translation with suppression of cap-binding factor eIF4E (Andreev *et al.*, 2013). In the case of *AGO1* 5'UTR, it seems the free 5' end of the transcript enhances *AGO1* 5'UTR-mediated translation initiation, suggesting it mediates an eIF4G-enhanced mechanism of cap-independent translation initiation that is similar to CITE-mediated translation.

All the experiments performed to understand whether and how *AGO1* 5'UTR is able to mediate cap-independent translation initiation were done in HeLa cells. However, since the AGO1 protein has been identified as a potential biomarker in colorectal cancer (Li *et al.*, 2010b), further experiments are required to understand the relevance of this alternative mechanism to the onset and development of such disease, and whether the identified mechanism of translation initiation mediated by *AGO1* 5'UTR is able to be effective in colon-derived cell lines.

The presence of an AUG codon within *AGO1* 5'UTR may also play a role in regulating translation of the downstream ORF if in frame with a stop codon within *AGO1* coding sequence. Thus, further analysis is required to fully characterise the mechanisms that regulate AGO1 protein synthesis and to what extent they influence its role as a component of the RNA silencing complexes. Furthermore, it would be of great relevance to understand whether the other proteins of the Argonaute family can also be translated via a non-canonical mechanism of translation initiation. Although all mammalian AGOs contribute to miRNA silencing, individual AGOs have overlapping functions in this process (Su *et al.*, 2009). This fact may indicate they are all

regulated by similar mechanisms of translation initiation, in order to cope with cellular needs in a similar and concerted fashion.

AGO1 protein, as previously mentioned, plays a central role in RNA silencing complexes and, therefore, is expected to be tightly regulated in the cell. The existence of a cap-independent mechanism able to regulate its expression may be pivotal to maintain its expression levels. This would favour the correct mechanisms of miRNA-related gene expression regulation.

Regarding *MLH1* 5'UTR's ability to regulate gene expression, we observed, on the one hand, that this sequence is able to mediate translation of downstream FLuc ORF from a promoterless plasmid (figure 3.52), which indicates the presence of a cryptic promoter that can initiate transcription and, then, canonical cap-dependent translation of the monocistronic mRNA can take place. The presence of such promoter within this region has already been described (Ito *et al.*, 1999; Arita *et al.*, 2003). Arita *et al.* (2003) identified eight protein-binding sites in the minimal promoter region of the *MLH1* gene, which spans between nts -301 and -76 in relation to the translation start site. This region has been documented to include two hypermethylated regions in *MLH1*-unexpressing colorectal cancer cells (Deng *et al.*, 1999) and a core promoter (nts -184 to -132) determined by a luciferase reporter gene assay with a series of 5' end deletional mutants in NIH3T3 cells (Ito *et al.*, 1999). These results strongly suggest that a transcription-regulatory region of the *MLH1* gene is within 300 bp upstream of the start site. In this minimal promoter region, seven protein-binding sites, initially referred to as FP1–FP6, were determined by *in vivo* methylation footprinting. One additional site, CCAAT-box, was evident in a homology search and electrophorectic mobility shift assay. Three protein-binding sites appeared to be important to fully express a transcriptional activity because single site-disrupted mutants at CCAAT-box (nts -145 to -139), the FP61 site (nts -96 to -93) — which is an upstream part of the FP6 site —, and the FP3 site (nts -163 to -158) showed the lowest luciferase activity (Arita *et al.*, 2003). The *MLH1* 5'UTR sequence referenced in NCBI database, and used in this work, spans from the translation start site up to nt -198, which means it includes the core promoter sequence and all the three protein-binding sites. Hence, our purpose was to understand whether this sequence regulates gene expression differently in colorectal cancer cells compared to the counterpart normal ones. We also evaluated its activity in HeLa cells to analyse its behaviour in different tissues. From the experiments with promoterless plasmids containing *MLH1* 5'UTR, we concluded that the cryptic promoter is much more active in NCM460 cells (normal mucosa-derived cell line) than in cancer-derived cell lines. This agrees with the fact that MLH1 is a mismatch repair protein (NCBI), whose expression is drastically reduced in colorectal cancer due to hypermethylation of the CpG islands in the promoter region (Deaton and Bird, 2011). The fact that the promoter is more active in NCM460 cells reflects its non-hypermethylated status. Additionally, several mutations and polymorphisms have been described in this region as being related to a colorectal cancer phenotype (Isidro *et al.*, 2003; Zhong *et al.*, 2007; Mei *et al.*, 2010). We tested the effect of one mutation — c.-28A>T (Isidro *et al.*, 2003) — and one polymorphism — c.-93G>A (Mei *et al.*, 2010) — on relative FLuc expression levels driven by the promoter within *MLH1* 5'UTR sequence (figures 3.54 and 3.55, respectively). Polymorphism as nt -93 should lead to a reduced gene

expression, as observed by Mei *et al.* (2010). In NCM460 cells, no differences were observed when comparing mutated sequence to the wild-type counterpart. In HeLa cells, the significant increase in relative FLuc expression levels observed in the promoterless *MLH1*-93-containing plasmid does not occur in the counterpart promoter-containing one, which differs from published literature. This result may be explained by some alteration in FLuc translation mediated by a putative *MLH1* 5'UTR-dependent non-canonical mechanism of translation initiation that affects the results from the promoter-containing plasmid. Also, the increase can be explained by the absence of the sequence upstream nt -198, which causes a reduction in gene expression when nt -93 is altered. However, in HCT116 cells, there actually is a reduction in relative expression levels from the promoterless *MLH1-93*-containing plasmid, which is in agreement with the data obtained by Mei *et al.* (2010). This result suggests an opposite effect of the polymorphism at nt -93 in tissues with different origins (HeLa cells and HCT116 cells), whereas in NCM460 cells it has no effect. Regarding mutation at nt -28, there is a decrease in relative FLuc expression levels from the mutated plasmid in HeLa cells, but not in colon-derived cell lines. No data in the literature about the effect of this mutation on gene expression was found; so, in the future, additional experiments will be required to understand how this mutation affects gene expression. However, reduction in gene expression as a consequence of a mutated sequence may lead to reduced levels of functional *MLH1*, which is a characteristic of cervical cancers (cf. section I).

On the other hand, we tried to understand to what extent the *MLH1* 5'UTR sequence is able to mediate a non-canonical mechanism of translation initiation. Transfecting cells with *in vitro* transcribed mRNA resulted in a significant increase in relative FLuc expression levels in all tested cell lines (figure 3.57), which suggests an alternative mechanism of translation initiation. However, in cells that had been subject to stress stimuli (figures 3.58, 3.59 and 3.60), the results obtained were inconsistent with the former, as they did not reflect the existence of a non-canonical mechanism of translation initiation. The differences observed may be due to longer transfection periods that ended up altering the stability of the transfected mRNA and lead to its degradation (Hayashi *et al.*, 2010). Of note, when cells were stimulated with $CoCl_2$, whose vehicle is H_2O, we saw, again, a significant increase in relative FLuc expression levels from *MLH1* 5'UTR-containing plasmids in every tested cell line. This suggests DMSO may be causing some mRNA degradation and we could not therefore gather solid conclusions.

Transfection of cells with a monicistronic transcript containing *MLH1* 5'UTR (figure 3.63), instead of a bicistronic one, revealed that this sequence is able to mediate translation initiation in transcripts without cap structure, and with a free 5' end, similarly to what was observed for *AGO1* 5'UTR.

When cells were transfected with a bicistronic *MLH1-28*-containing transcript, we observed a significant reduction in relative FLuc expression levels from such transcript, compared to those from the wild-type-containing transcript in NCM460 cells, but not in the other cell lines. This suggests this mutation may have an influence in regulating protein expression in those cells, and alterations in the sequence may lead to a reduced MLH1 expression in normal mucosa cells and, consequently, to a deficiency in mismatch repair genes that will eventually originate a transformed

phenotype. Experiments concerning the effect of the polymorphism at nt -93 in the putative non-canonical mechanism of translation initiation proved inconclusive, as none of the transcripts containing either wild-type or altered *MLH1* 5'UTR is able to conduct FLuc translation. Thus, more studies are needed to clarify the actual mechanism through which *MLH1* 5'UTR mediates non-canonical translation initiation. However, this sequence does not seem to be able to mediate an internal cap-independent mechanism of translation initiation, as is the case of *UPF1* 5'UTR, since it failed the bicistronic test in several contexts (Shatsky *et al.*, 2010; Terenin *et al.*, 2013). The results obtained are not consistent, which may be a sign that the presence of *MLH1* 5'UTR makes the transcript more susceptible to degradation.

Overall, the results throughout this work shed light on the mechanisms that govern translation regulation of the selected proteins. Thus, we can conclude that *UPF1* 5'UTR mediates a cap-independent mechanism of translation initiation that works in an IRES-like mode, whereas *AGO1* 5'UTR mediates a cap-independent mechanism of translation initiation that works in a CITE-like mode. As for *MLH1* 5'UTR, no actual mechanism has been identified, although it appears to mediate a non-canonical mechanism of translation initiation. For all sequences, more in-depth studies would need to be performed, either to elucidate their biological function in regulating alternative translation initiation (the case of *UPF1* and *AGO1* 5'UTR), or to clarify the actual mechanism involved in such activity (the case of *MLH1* 5'UTR). Moreover, a more thorough analysis should include an evaluation of the putative, *in vivo* formation of the predicted G-quadruplex structures. Answers to these questions might better explain the regulation of such mechanisms and the development of innovative therapeutic approaches based on the manipulation of the sequences that mediate these alternative mechanisms of translation initiation. Furthermore, controlled stimulation or repression of non-canonical translation initiation of proteins, such as UPF1 and AGO1, might be useful to develop new therapies to fight certain human diseases, including cancer.

5.

Bibliography

Acker, M.G., Shin, B.-S., Dever, T.E., and Lorsch, J.R. (2006). *Interaction between Eukaryotic Initiation Factors 1A and 5B Is Required for Efficient Ribosomal Subunit Joining*. J. Biol. Chem. *281*, 8469–8475.

Agirrezabala, X., and Frank, J. (2009). *Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu*. Q. Rev. Biophys. *42*, 159.

Agirrezabala, X., and Frank, J. (2010). *From DNA to proteins via the ribosome: structural insights into the workings of the translation machinery*. Hum. Genomics *4*, 226.

Ajamian, L., Abrahamyan, L., Milev, M., Ivanov, P.V., Kulozik, A.E., Gehring, N.H., and Mouland, A.J. (2008). *Unexpected roles for UPF1 in HIV-1 RNA metabolism and translation.* RNA *14*, 914–927.

Ajamian, L., Abel, K., Rao, S., Vyboh, K., García-de-Gracia, F., Soto-Rifo, R., Kulozik, A.E., Gehring, N.H., and Mouland, A.J. (2015). *HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA*. Biomolecules *5*, 2808–2839.

Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L., and Pestova, T.V. (2006). *In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3*. Cell *125*, 1125–1136.

Alotte, C., Martin, A., Caldarelli, S., Digiorgio, A., Condom, R., Zoulim, F., Durantel, D., and Hantz, O. (2008). *Short peptide nucleic acids (PNA) inhibit hepatitis C virus internal ribosome entry site (IRES) dependent translation in vitro*. Antiviral Res. *80*, 280–287.

Andersen, C.B.F. (2006). *Structure of the Exon Junction Core Complex with a Trapped DEAD-Box ATPase Bound to RNA*. Science *313*, 1968–1972.

Andreev, D.E., Dmitriev, S.E., Zinovkin, R., Terenin, I.M., and Shatsky, I.N. (2012). *The 5′ untranslated region of Apaf-1 mRNA directs translation under apoptosis conditions via a 5′ end-dependent scanning mechanism*. FEBS Lett. *586*, 4139–4143.

Andreev, D.E., Dmitriev, S.E., Terenin, I.M., and Shatsky, I.N. (2013). *Cap-independent translation initiation of Apaf-1 mRNA based on a scanning mechanism is determined by some features of the secondary structure of its 5′ untranslated region*. Biochem. Mosc. *78*, 157–165.

Applequist, S.E., Selg, M., Raman, C., and Jäck, H.M. (1997). *Cloning and characterization of HUPF1, a human homolog of the Saccharomyces cerevisiae nonsense mRNA-reducing UPF1 protein*. Nucleic Acids Res. *25*, 814–821.

Arita, M., Zhong, X., Min, Z., Hemmi, H., and Shimatake, H. (2003). *Multiple sites required for expression in 5′-flanking region of the hMLH1 gene*. Gene *306*, 57–65.

Atkin, A.L., Altamura, N., Leeds, P., and Culbertson, M.R. (1995). *The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm*. Mol. Biol. Cell *6*, 611–625.

Audigier, S., Guiramand, J., Prado-Lourenco, L., Conte, C., Gonzalez-Herrera, I.G., Cohen-Solal, C., Récasens, M., and Prats, A.-C. (2008). *Potent activation of FGF-2 IRES-dependent mechanism of translation during brain development*. RNA N. Y. N *14*, 1852–1864.

Avery, P., Vicente-Crespo, M., Francis, D., Nashchekina, O., Alonso, C.R., and Palacios, I.M. (2011). *Drosophila Upf1 and Upf2 loss of function inhibits cell growth and causes animal death in a Upf3-independent manner*. RNA N. Y. N *17*, 624–638.

Azzalin, C.M. (2012). *UPF1: A leader at the end of chromosomes*. Nucleus *3*, 16–21.

Azzalin, C.M., and Lingner, J. (2006a). *The double life of UPF1 in RNA and DNA stability pathways*. Cell Cycle *5*, 1496–1498.

Azzalin, C.M., and Lingner, J. (2006b). *The Human RNA Surveillance Factor UPF1 Is Required for S Phase Progression and Genome Stability*. Curr. Biol. *16*, 433–439.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). *Telomeric Repeat Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends*. Science *318*, 798–801.

Baird, S.D., Turcotte, M., Korneluk, R.G., and Holcik, M. (2006). *Searching for IRES*. RNA *12*, 1755–1785.

Balkwill, G.D., Derecka, K., Garner, T.P., Hodgman, C., Flint, A.P., and Searle, M.S. (2009). *Repression of translation of human estrogen receptor alpha by G-quadruplex formation.* Biochemistry (Mosc.) *48*, 11487–11495.

Balvay, L., Soto Rifo, R., Ricci, E.P., Decimo, D., and Ohlmann, T. (2009). *Structural and functional diversity of viral IRES*. Biochim. Biophys. Acta *1789*, 542–557.

Barbosa, C., and Romao, L. (2014). *Translation of the human erythropoietin transcript is regulated by an upstream open reading frame in response to hypoxia*. RNA *20*, 594–608.

Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., and Tampé, R. (2011). *Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1*. Proc. Natl. Acad. Sci. U. S. A. *108*, 3228–3233.

Bastide, A., Karaa, Z., Bornes, S., Hieblot, C., Lacazette, E., Prats, H., and Touriol, C. (2008). *An upstream open reading frame within an IRES controls expression of a specific VEGF-A isoform*. Nucleic Acids Res. *36*, 2434–2445.

Beaudoin, J.-D., and Perreault, J.-P. (2010). *5′-UTR G-quadruplex structures acting as translational repressors*. Nucleic Acids Res. *38*, 7022–7036.

Bellodi, C., Kopmar, N., and Ruggero, D. (2010). *Deregulation of oncogene-induced senescence and p53 translational control in X-linked dyskeratosis congenita*. EMBO J. *29*, 1865–1876.
Bentley, D.L. (2014). *Coupling mRNA processing with transcription in time and space*. Nat. Rev. Genet. *15*, 163–175.

Berk, Arnold. *Molecular cell biology*. Vol. 4. New York: WH Freeman, 2000.

Bernstein, C. (2015). *Epigenetic reduction of DNA repair in progression to gastrointestinal cancer*. World J. Gastrointest. Oncol. *7*, 30.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). *Role for a bidentate ribonuclease in the initiation step of RNA interference*. Nature *409*, 363–366.

Bert, A.G., Grépin, R., Vadas, M.A., and Goodall, G.J. (2006). *Assessing IRES activity in the HIF-1α and other cellular 5′ UTR*. RNA *12*, 1074–1083.

Bird TD. *Charcot-Marie-Tooth Neuropathy Type 1*. 1998 Aug 31 [Updated 2015 Mar 26]. *In*: Pagon RA, Adam MP, Ardinger HH, *et al.*, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2016. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1205/

Bisio (2015). *The 5'-untranslated region of p^{16INK4}a melanoma tumor suppressor acts as a cellular IRES, controlling mRNA translation under hypoxia through YBX1 binding*. Oncotarget.

Blais, J.D., Addison, C.L., Edge, R., Falls, T., Zhao, H., Wary, K., Koumenis, C., Harding, H.P., Ron, D., Holcik, M., *et al.* (2006). *Perk-dependent translational regulation promotes tumor cell adaptation and angiogenesis in response to hypoxic stress*. Mol. Cell. Biol. *26*, 9517–9532.

Blanco-Pérez, M., Pérez-Cañamás, M., Ruiz, L., and Hernández, C. (2016). *Efficient Translation of Pelargonium line pattern virus RNA Relies on a TED-Like 3´-Translational Enhancer that Communicates with the Corresponding 5´-Region through a Long-Distance RNA-RNA Interaction*. PLOS ONE *11*, e0152593.

Bochkov, Y., and Palmenberg, A. (2006). *Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location*. BioTechniques *41*, 283–292.

Bonnal, S. (2003). *A Single Internal Ribosome Entry Site Containing a G Quartet RNA Structure Drives Fibroblast Growth Factor 2 Gene Expression at Four Alternative Translation Initiation* Codons. J. Biol. Chem. *278*, 39330–39336.

Bornes, S., Prado-Lourenco, L., Bastide, A., Zanibellato, C., Iacovoni, J.S., Lacazette, E., Prats, A.-C., Touriol, C., and Prats, H. (2007). *Translational induction of VEGF internal ribosome entry site elements during the early response to ischemic stress*. Circ. Res. *100*, 305–308.

Braunstein, S., Karpisheva, K., Pola, C., Goldberg, J., Hochman, T., Yee, H., Cangiarella, J., Arju, R., Formenti, S.C., and Schneider, R.J. (2007). *A Hypoxia-Controlled Cap-Dependent to Cap-Independent Translation Switch in Breast Cancer*. Mol. Cell *28*, 501–512.

de Breyne, S., Yu, Y., Unbehaun, A., Pestova, T.V., and Hellen, C.U.T. (2009). *Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. Proc. Natl. Acad. Sci. U. S. A. 106, 9197–9202.*

Briscoe, J., and Thérond, P.P. (2013). *The mechanisms of Hedgehog signalling and its roles in development and disease*. Nat. Rev. Mol. Cell Biol. *14*, 416–429.

Brocato, J., Chervona, Y., and Costa, M. (2014). *Molecular Responses to Hypoxia-Inducible Factor 1 and Beyond*. Mol. Pharmacol. *85*, 651–657.

Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). *Structural basis for stop codon recognition in eukaryotes*. Nature *524*, 493–496.

Buecher, B., Cacheux, W., Rouleau, E., Dieumegard, B., Mitry, E., and Lièvre, A. (2013). *Role of microsatellite instability in the management of colorectal cancers*. Dig. Liver Dis. *45*, 441–449.

Bugaut, A., and Balasubramanian, S. (2012). *5′-UTR RNA G-quadruplexes: translation regulation and targeting*. Nucleic Acids Res. *40*, 4727–4741.

Bulygin, K.N., Khairulina, Y.S., Kolosov, P.M., Ven'yaminova, A.G., Graifer, D.M., Vorobjev, Y.N., Frolova, L.Y., Kisselev, L.L., and Karpova, G.G. (2010). *Three distinct peptides from the N domain of translation termination factor eRF1 surround stop codon in the ribosome*. RNA *16*, 1902–1914.

Buratowski, S. (2005). *Connections between mRNA 3' end processing and transcription termination*. Curr. Opin. Cell Biol. *17*, 257–261.

Buratti, E., Tisminetzky, S., Zotti, M., and Baralle, F.E. (1998). *Functional analysis of the interaction between HCV 5'UTR and putative subunits of eukaryotic translation initiation factor eIF3*. Nucleic Acids Res. *26*, 3179–3187.

Burroughs, A.M., Ando, Y., de Hoon, M.J.L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2011). *Deep-sequencing of human Argonaute-associated small RNA provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin*. RNA Biol. *8*, 158–177.

Cammas, A., Dubrac, A., Morel, B., Lamaa, A., Touriol, C., Teulade-Fichou, M.-P., Prats, H., and Millevoi, S. (2015). *Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation*. RNA Biol. *12*, 320–329.

Candeias, M.M., Powell, D.J., Roubalova, E., Apcher, S., Bourougaa, K., Vojtesek, B., Bruzzoni-Giovanelli, H., and Fåhraeus, R. (2006). *Expression of p53 and p53/47 are controlled by alternative mechanisms of messenger RNA translation initiation*. Oncogene *25*, 6936–6947.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). *The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumourigenesis*. Genes Dev. *16*, 2733–2742.

Carvalho, M.D., Carvalho, J.F., and Merrick, W.C. (1984). *Biological characterization of various forms of elongation factor 1 from rabbit reticulocytes*. Arch. Biochem. Biophys. *234*, 603–611.

Celik, A., Kervestin, S., and Jacobson, A. (2015). *NMD: At the crossroads between translation termination and ribosome recycling*. Biochimie *114*, 2–9.

Cerutti, L., Mian, N., and Bateman, A. (2000). *Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain*. Trends Biochem. Sci. *25*, 481–482.

Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). *Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2*. Mol. Cell *41*, 693–703.

Chang, Z., Zhang, W., Chang, Z., Song, M., Qin, Y., Chang, F., Guo, H., and Wei, Q. (2015). *Expression characteristics of FHIT, p53, BRCA2 and MLH1 in families with a history of oesophageal cancer in a region with a high incidence of oesophageal cancer*. Oncol. Lett. *9*, 430–436.

Chappell, S.A., Dresios, J., Edelman, G.M., and Mauro, V.P. (2006). *Ribosomal shunting mediated by a translational enhancer element that base pairs to 18S rRNA*. Proc. Natl. Acad. Sci. *103*, 9488–9493.
Chatel-Chaix, L., Clément, J.-F., Martel, C., Bériault, V., Gatignol, A., DesGroseillers, L., and Mouland, A.J. (2004). *Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles*. Mol. Cell. Biol. *24*, 2637–2648.

Chavatte, L., Seit-Nebi, A., Dubovaya, V., and Favre, A. (2002). *The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome*. EMBO J. *21*, 5302–5311.

Chawla, R., Redon, S., Raftopoulou, C., Wischnewski, H., Gagos, S., and Azzalin, C.M. (2011). *Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication*. EMBO J. *30*, 4047–4058.

Chen, T.-M., Shih, Y.-H., Tseng, J.T., Lai, M.-C., Wu, C.-H., Li, Y.-H., Tsai, S.-J., and Sun, H.S. (2014). *Overexpression of FGF9 in colon cancer cells is mediated by hypoxia-induced translational activation*. Nucleic Acids Res. *42*, 2932–2944.

Cheng, Y., Jia, C., Li, G., and Li, H. (2014). *Expression of Eukaryotic Initiation Factor 3f Is Associated with Prognosis in Gastric Carcinomas*. Oncol. Res. Treat. *37*, 198–202.

Cho, H., Kim, K.M., and Kim, Y.K. (2009). *Human Proline-Rich Nuclear Receptor Coregulatory Protein 2 Mediates an Interaction between mRNA Surveillance Machinery and Decapping Complex*. Mol. Cell *33*, 75–86.

Cobbold, L.C., Spriggs, K.A., Haines, S.J., Dobbyn, H.C., Hayes, C., de Moor, C.H., Lilley, K.S., Bushell, M., and Willis, A.E. (2008). *Identification of internal ribosome entry segment (IRES)-trans-acting factors for the Myc family of IRESs*. Mol. Cell. Biol. *28*, 40–49.

Colussi, T.M., Costantino, D.A., Zhu, J., Donohue, J.P., Korostelev, A.A., Jaafar, Z.A., Plank, T.-D.M., Noller, H.F., and Kieft, J.S. (2015). *Initiation of translation in bacteria by a structured eukaryotic IRES RNA*. Nature *519*, 110–113.

Connolly, E.P., Silvera, D., Badura, M.L., Venuto, T., and Schneider, R.J. (2011). *Over-expression of Translation Initiation Factor eIF4G Confers Robust Radioresistance to the Cancer Stem Cell Population in Inflammatory Breast Cancer*. Int. J. Radiat. Oncol. *81*, S85.

Conte, C., Riant, E., Toutain, C., Pujol, F., Arnal, J.-F., Lenfant, F., and Prats, A.-C. (2008). *FGF2 translationally induced by hypoxia is involved in negative and positive feedback loops with HIF-1alpha*. PloS One *3*, e3078.

Conte, C., Ainaoui, N., Delluc-Clavieres, A., Khoury, M.P., Azar, R., Pujol, F., Martineau, Y., Pyronnet, S., and Prats, A.-C. (2009). *Fibroblast growth factor 1 induced during myogenesis by a transcription-translation coupling mechanism*. Nucleic Acids Res. *37*, 5267–5278.

Cooper GM. *The Cell: A Molecular Approach*. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Available from: http://www.ncbi.nlm.nih.gov/books/NBK9839/

Cornelis, S., Bruynooghe, Y., Denecker, G., Van Huffel, S., Tinton, S., and Beyaert, R. (2000). *Identification and characterization of a novel cell cycle-regulated internal ribosome entry site.* Mol. Cell *5*, 597–605.

Cragun, D., Radford, C., Dolinsky, J.S., Caldwell, M., Chao, E., and Pal, T. (2014). *Panel-based testing for inherited colorectal cancer: a descriptive study of clinical testing performed by a US laboratory: Panel-based testing for inherited CRC*. Clin. Genet. *86*, 510–520.

Culbertson, M.R., and Leeds, P.F. (2003). *Looking at mRNA decay pathways through the window of molecular evolution*. Curr. Opin. Genet. Dev. *13*, 207–214.

Damiano, F., Alemanno, S., Gnoni, G.V., and Siculella, L. (2010). *Translational control of the sterol-regulatory transcription factor SREBP-1 mRNA in response to serum starvation or ER stress is mediated by an internal ribosome entry site*. Biochem. J. *429*, 603–612.

De La Cruz, J., Iost, I., Kressler, D., and Linder, P. (1997). *The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. *94*, 5201–5206.

De Pietri Tonelli, D. (2004). *Translational regulation of BACE-1 expression in neuronal and non-neuronal cells*. Nucleic Acids Res. *32*, 1808–1817.

Deaton, A.M., and Bird, A. (2011). *CpG islands and the regulation of transcription*. Genes Dev. *25*, 1010–1022.

Delatte, B., Wang, F., Ngoc, L.V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., *et al.* (2016). *Transcriptome-wide distribution and function of RNA hydroxymethylcytosine*. Science *351*, 282–285.

Deng, G., Chen, A., Hong, J., Chae, H.S., and Kim, Y.S. (1999). *Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression*. Cancer Res. *59*, 2029–2033.

Denning, G., Jamieson, L., Maquat, L.E., Thompson, E.A., and Fields, A.P. (2001). *Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein*. J. Biol. Chem. *276*, 22709–22714.

Descamps, G., Gomez-Bougie, P., Tamburini, J., Green, A., Bouscary, D., Maïga, S., Moreau, P., Le Gouill, S., Pellat-Deceunynck, C., and Amiot, M. (2012). *The cap-translation inhibitor 4EGI-1 induces apoptosis in multiple myeloma through Noxa induction*. Br. J. Cancer *106*, 1660–1667.

Dever, T.E., and Green, R. (2012a). *The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes*. Cold Spring Harb. Perspect. Biol. *4*, a013706–a013706.

Diederichs, S., Bartsch, L., Berkmann, J.C., Fröse, K., Heitmann, J., Hoppe, C., Iggena, D., Jazmati, D., Karschnia, P., Linsenmeier, M., *et al.* (2016). *The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations*. EMBO Mol. Med. *8*, 442–457.

Dikstein, R. (2012). *Transcription and translation in a package deal: The TISU paradigm*. Gene *491*, 1–4.

Dobbyn, H.C., Hill, K., Hamilton, T.L., Spriggs, K.A., Pickering, B.M., Coldwell, M.J., de Moor, C.H., Bushell, M., and Willis, A.E. (2008). *Regulation of BAG-1 IRES-mediated translation following chemotoxic stress*. Oncogene *27*, 1167–1174.

Dome, J.S., and Coppes, M.J. (2002). *Recent advances in Wilms tumor genetics*. Curr. Opin. Pediatr. *14*, 5–11.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., *et al.* (2012). *Topology of the human and mouse m6A RNA methylomes revealed by m^{6}A-seq*. Nature *485*, 201–206.

Dong, J., Lai, R., Nielsen, K., Fekete, C.A., Qiu, H., and Hinnebusch, A.G. (2004). *The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly*. J. Biol. Chem. *279*, 42157–42168.

Dueck, A., Ziegler, C., Eichner, A., Berezikov, E., and Meister, G. (2012). *microRNA associated with the different human Argonaute proteins*. Nucleic Acids Res. *40*, 9850–9862.

Elfakess, R., and Dikstein, R. (2008). *A Translation Initiation Element Specific to mRNA with Very Short 5'UTR that Also Regulates Transcription*. PLoS ONE *3*, e3094.

Elfakess, R., Sinvani, H., Haimov, O., Svitkin, Y., Sonenberg, N., and Dikstein, R. (2011). *Unique translation initiation of mRNA-containing TISU element*. Nucleic Acids Res. *39*, 7598–7609.

Elroy-Stein, O., and Merrick, W.C. (2007). *Translation Initiation Via Cellular Internal Ribosome Entry Sites*. Cold Spring Harb. Monogr. Arch. *48*, 155–172.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., *et al.* (2007). *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*. Nature *447*, 799–816.

Ender, C., and Meister, G. (2010). *Argonaute proteins at a glance*. J. Cell Sci. *123*, 1819–1823.

Erickson, F.L., and Hannig, E.M. (1996). *Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit*. EMBO J. *15*, 6311–6320.

Fabian, M.R., and White, K.A. (2004). *5'-3' RNA-RNA interaction facilitates cap- and poly(A) tail-independent translation of tomato bushy stunt virus mrna: a potential common mechanism for tombusviridae*. J. Biol. Chem. *279*, 28862–28872.

Faehnle, C.R., Elkayam, E., Haase, A.D., Hannon, G.J., and Joshua-Tor, L. (2013). *The Making of a Slicer: Activation of Human Argonaute-1*. Cell Rep. *3*, 1901–1909.

Fan, S., Li, Y., Yue, P., Khuri, F.R., and Sun, S.-Y. (2010). *The eIF4E/eIF4G interaction inhibitor 4EGI-1 augments TRAIL-mediated apoptosis through c-FLIP down-regulation and DR5 induction independent of inhibition of cap-dependent protein translation*. Neoplasia N. Y. NY *12*, 346.

Fatscher, T., Boehm, V., and Gehring, N.H. (2015). *Mechanism, factors, and physiological role of nonsense-mediated mRNA decay*. Cell. Mol. Life Sci. CMLS *72*, 4523–4544.

Faye, M.D., and Holcik, M. (2015). *The role of IRES trans-acting factors in carcinogenesis*. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1849*, 887–897.

Faye, M.D., Graber, T.E., Liu, P., Thakor, N., Baird, S.D., Durie, D., and Holcik, M. (2013). *Nucleotide Composition of Cellular Internal Ribosome Entry Sites Defines Dependence on NF45 and Predicts a Posttranscriptional Mitotic Regulon*. Mol. Cell. Biol. *33*, 307–318.

Faye, M.D., Beug, S.T., Graber, T.E., Earl, N., Xiang, X., Wild, B., Langlois, S., Michaud, J., Cowan, K.N., Korneluk, R.G., *et al.* (2014). *IGF2BP1 controls cell death and drug resistance in rhabdomyosarcomas by regulating translation of cIAP1*. Oncogene.

Feoktistova, K., Tuvshintogs, E., Do, A., and Fraser, C.S. (2013). *Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity*. Proc. Natl. Acad. Sci. *110*, 13339–13344.

Fernandez, J., Yaman, I., Mishra, R., Merrick, W.C., Snider, M.D., Lamers, W.H., and Hatzoglou, M. (2001). *Internal ribosome entry site-mediated translation of a mammalian mRNA is regulated by amino acid availability*. J. Biol. Chem. *276*, 12285–12291.

Fernandez, J., Bode, B., Koromilas, A., Diehl, J.A., Krukovets, I., Snider, M.D., and Hatzoglou, M. (2002). *Translation mediated by the internal ribosome entry site of the cat-1 mRNA is regulated by glucose availability in a PERK kinase-dependent manner*. J. Biol. Chem. *277*, 11780–11787.

Fernandez, J., Yaman, I., Huang, C., Liu, H., Lopez, A.B., Komar, A.A., Caprara, M.G., Merrick, W.C., Snider, M.D., Kaufman, R.J., *et al.* (2005). *Ribosome Stalling Regulates IRES-Mediated Translation in Eukaryotes, a Parallel to Prokaryotic Attenuation*. Mol. Cell *17*, 405–416.

Fiorini, F., Bagchi, D., Le Hir, H., and Croquette, V. (2015). *Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities*. Nat. Commun. *6*, 7581.

Fitzgerald, K.D., and Semler, B.L. (2009). *Bridging IRES elements in mRNA to the eukaryotic translation apparatus*. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1789*, 518–528.

Franks, T.M., Singh, G., and Lykke-Andersen, J. (2010). *Upf1 ATPase-Dependent mRNP Disassembly Is Required for Completion of Nonsense- Mediated mRNA Decay*. Cell *143*, 938–950.

Fraser, C.S., Berry, K.E., Hershey, J.W.B., and Doudna, J.A. (2007). *eIF3j Is Located in the Decoding Center of the Human 40S Ribosomal Subunit*. Mol. Cell *26*, 811–819.

Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., and Kisselev, L. (1996). *Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase*. RNA N. Y. N *2*, 334–341.

Frolova, L.Y., Tsivkovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov, V.M., Tatkov, S.I., and Kisselev, L.L. (1999). *Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis*. RNA N. Y. N *5*, 1014–1020.

Fu, Q., Chen, Z., Gong, X., Cai, Y., Chen, Y., Ma, X., Zhu, R., and Jin, J. (2015). *β-Catenin expression is regulated by an IRES-dependent mechanism and stimulated by paclitaxel in human ovarian cancer cells*. Biochem. Biophys. Res. Commun. *461*, 21–27.

Gaccioli, F., Huang, C.C., Wang, C., Bevilacqua, E., Franchi-Gazzola, R., Gazzola, G.C., Bussolati, O., Snider, M.D., and Hatzoglou, M. (2006). *Amino acid starvation induces the SNAT2 neutral amino acid transporter by a mechanism that involves eukaryotic initiation factor 2alpha phosphorylation and cap-independent translation*. J. Biol. Chem. *281*, 17929–17940.

Gallagher, J.W., Kubica, N., Kimball, S.R., and Jefferson, L.S. (2008). *Reduced Eukaryotic Initiation Factor 2B -Subunit Expression Suppresses the Transformed Phenotype of Cells Overexpressing the Protein*. Cancer Res. *68*, 8752–8760.

Gausachs, M., Mur, P., Corral, J., Pineda, M., González, S., Benito, L., Menéndez, M., Espinàs, J.A., Brunet, J., Iniesta, M.D., *et al.* (2012). *MLH1 promoter hypermethylation in the analytical algorithm of Lynch syndrome: a cost-effectiveness study*. Eur. J. Hum. Genet. EJHG *20*, 762–768.

Genuardi, M., Viel, A., Bonora, D., Capozzi, E., Bellacosa, A., Leonardi, F., Valle, R., Ventura, A., Pedroni, M., Boiocchi, M., *et al.* (1998). *Characterization of MLH1 and MSH2 alternative splicing and its relevance to molecular testing of colorectal cancer susceptibility*. Hum. Genet. *102*, 15–20.

Gerlitz, G., Jagus, R., and Elroy-Stein, O. (2002). *Phosphorylation of initiation factor-2 alpha is required for activation of internal translation initiation during cell differentiation*. Eur. J. Biochem. FEBS *269*, 2810–2819.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S., and Snyder, M. (2007). *What is a gene, post-ENCODE? History and updated definition*. Genome Res. *17*, 669–681.

Ghosh, U., and Adhya, S. (2016). *Non-Equivalent Roles of AGO1 and AGO2 in mRNA Turnover and Translation of Cyclin D1 mRNA*. J. Biol. Chem.

Giarnieri, E., Mancini, R., Pisani, T., Alderisio, M., and Vecchione, A. (2000). *Msh2, Mlh1, Fhit, p53, Bcl-2, and Bax expression in invasive and in situ squamous cell carcinoma of the uterine cervix*. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *6*, 3600–3606.

Goh, S.-H., Hong, S.-H., Hong, S.-H., Lee, B.-C., Ju, M.-H., Jeong, J.-S., Cho, Y.-R., Kim, I.-H., and Lee, Y.-S. (2011). *eIF3m expression influences the regulation of tumourigenesis-related genes in human colon cancer*. Oncogene *30*, 398–409.

Gomez, E., Mohammad, S.S., and Pavitt, G.D. (2002). *Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation*. EMBO J. *21*, 5292–5301.

Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y., and Maquat, L.E. (2009). *SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNA*. Genes Dev. *23*, 54–66.

Gonzalez-Herrera, I.G., Prado-Lourenco, L., Pileur, F., Conte, C., Morin, A., Cabon, F., Prats, H., Vagner, S., Bayard, F., Audigier, S., *et al.* (2006). *Testosterone regulates FGF-2 expression during testis maturation by an IRES-dependent translational mechanism*. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *20*, 476–478.

Graber, T.E., and Holcik, M. (2007). *Cap-independent regulation of gene expression in apoptosis*. Mol. Biosyst. *3*, 825.

Graber, T.E., Baird, S.D., Kao, P.N., Mathews, M.B., and Holcik, M. (2010). *NF45 functions as an IRES trans-acting factor that is required for translation of cIAP1 during the unfolded protein response*. Cell Death Differ. *17*, 719–729.

Grifo, J.A., Abramson, R.D., Satler, C.A., and Merrick, W.C. (1984). *RNA-stimulated ATPase activity of eukaryotic initiation factors*. J. Biol. Chem. *259*, 8648–8654.

Grillo, G. (2003). *PatSearch: a program for the detection of patterns and structural motifs in nucleotide sequences*. Nucleic Acids Res. *31*, 3608–3612.

Gromadski, K.B., Schummer, T., Stromgaard, A., Knudsen, C.R., Kinzy, T.G., and Rodnina, M.V. (2007). *Kinetics of the Interactions between Yeast Elongation Factors 1A and 1B , Guanine Nucleotides, and Aminoacyl-tRNA*. J. Biol. Chem. *282*, 35629–35637.

Grover, R., Ray, P.S., and Das, S. (2008). *Polypyrimidine tract binding protein regulates IRES-mediated translation of p53 isoforms*. Cell Cycle Georget. Tex *7*, 2189–2198.

Grover, R., Candeias, M.M., F\aahraeus, R., and Das, S. (2009). *p53 and little brother p53/47: linking IRES activities with protein functions*. Oncogene *28*, 2766–2772.

Grover, R., Sharathchandra, A., Ponnuswamy, A., Khan, D., and Das, S. (2011). *Effect of mutations on the p53 IRES RNA structure: implications for de-regulation of the synthesis of p53 isoforms*. RNA Biol. *8*, 132–142.

Gu, L., Zhu, N., Zhang, H., Durden, D.L., Feng, Y., and Zhou, M. (2009). *Regulation of XIAP Translation and Induction by MDM2 following Irradiation*. Cancer Cell *15*, 363–375.

Guan, D., Altan-Bonnet, N., Parrott, A.M., Arrigo, C.J., Li, Q., Khaleduzzaman, M., Li, H., Lee, C.-G., Pe'ery, T., and Mathews, M.B. (2008). *Nuclear Factor 45 (NF45) Is a Regulatory Subunit of Complexes with NF90/110 Involved in Mitotic Control*. Mol. Cell. Biol. *28*, 4629–4641.

Guan, X.-Y., Sham, J.S., Tang, T.C., Fang, Y., Huo, K.-K., and Yang, J.-M. (2001). *Isolation of a novel candidate oncogene within a frequently amplified region at 3q26 in ovarian cancer*. Cancer Res. *61*, 3806–3809.

Guerrette, S., Acharya, S., and Fishel, R. (1999). *The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer.* J. Biol. Chem. *274*, 6336–6341.

Haimov, O., Sinvani, H., and Dikstein, R. (2015). *Cap-dependent, scanning-free translation initiation mechanisms*. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1849*, 1313–1318.

Halaby, M.-J., Harris, B.R.E., Miskimins, W.K., Cleary, M.P., and Yang, D.-Q. (2015a). *Deregulation of IRES-mediated p53 translation in cancer cells with defective p53 response to DNA damage*. Mol. Cell. Biol. MCB.00365-15.

Halaby, M.-J., Li, Y., Harris, B.R., Jiang, S., Miskimins, W.K., Cleary, M.P., and Yang, D.-Q. (2015b). *Translational Control Protein 80 Stimulates IRES-Mediated Translation of p53 mRNA in Response to DNA Damage*. BioMed Res. Int. *2015*, 1–9.

Haraldsdottir, S., Hampel, H., Wu, C., Weng, D.Y., Shields, P.G., Frankel, W.L., Pan, X., de la Chapelle, A., Goldberg, R.M., and Bekaii-Saab, T. (2016). *Patients with colorectal cancer associated with Lynch syndrome and MLH1 promoter hypermethylation have similar prognoses*. Genet. Med.

Hayashi, T., Lamba, D.A., Slowik, A., Reh, T.A., and Bermingham-McDonogh, O. (2010). *A method for stabilising RNA for transfection that allows control of expression duration*. Dev. Dyn. *239*, 2034–2040.

Hellen, C.U.T. (2001). *Internal ribosome entry sites in eukaryotic mRNA molecules*. Genes Dev. *15*, 1593–1612.

Herczenik, E., and Gebbink, M.F.B.G. (2008). *Molecular and cellular aspects of protein misfolding and disease*. FASEB J. *22*, 2115–2133.

Hernández, G., Vázquez-Pianzola, P., Sierra, J.M., and Rivera-Pomar, R. (2004). *Internal ribosome entry site drives cap-independent translation of reaper and heat shock protein 70 mRNA in Drosophila embryos*. RNA N. Y. N *10*, 1783–1797.

Hinnebusch, A.G. (2011). *Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes*. Microbiol. Mol. Biol. Rev. *75*, 434–467.

Hinnebusch, A.G. (2014). *The scanning mechanism of eukaryotic translation initiation*. Annu. Rev. Biochem. *83*, 779–812.

Hocine, S., Singer, R.H., and Grunwald, D. (2010). *RNA Processing and Export*. Cold Spring Harb. Perspect. Biol. *2*, a000752–a000752.

Höck, J., and Meister, G. (2008). *The Argonaute protein family*. Genome Biol *9*, 210.

Holcik, M. (2015). *Could the eIF2α-Independent Translation Be the Achilles Heel of Cancer?* Front. Oncol. *5*.

Holcik, M., and Sonenberg, N. (2005). *Translational control in stress and apoptosis*. Nat Rev Mol Cell Biol *6*, 318–327.

Holcik, M., Lefebvre, C., Yeh, C., Chow, T., and Korneluk, R.G. (1999). *A new internal-ribosome-entry-site motif potentiates XIAP-mediated cytoprotection*. Nat. Cell Biol. *1*, 190–192.

Holcik, M., Sonenberg, N., and Korneluk, R.G. (2000). *Internal ribosome initiation of translation and the control of cell death*. Trends Genet. *16*, 469–473.

Holcik, M., Gordon, B.W., and Korneluk, R.G. (2003). *The Internal Ribosome Entry Site-Mediated Translation of Antiapoptotic Protein XIAP Is Modulated by the Heterogeneous Nuclear Ribonucleoproteins C1 and C2*. Mol. Cell. Biol. *23*, 280–288.

Hou, H.-Y., Kung, S.-H., Lin, C.-W., Wu, K.-Y., and Lu, W.-W. (2016). *Idarubicin is a broad-Spectrum enterovirus replication inhibitor that selectively targets the virus internal ribosomal entry site*. J. Gen. Virol.

Hsu, K.-S., Guan, B.-J., Cheng, X., Guan, D., Lam, M., Hatzoglou, M., and Kao, H.-Y. (2016). *Translational control of PML contributes to TNFα-induced apoptosis of MCF7 breast cancer cells and decreased angiogenesis in HUVECs*. Cell Death Differ. *23*, 469–483.

Huang, V., Zheng, J., Qi, Z., Wang, J., Place, R.F., Yu, J., Li, H., and Li, L.-C. (2013). *Ago1 Interacts with RNA Polymerase II and Binds to the Promoters of Actively Transcribed Genes in Human Cancer Cells*. PLoS Genet. *9*, e1003821.

Huang, Y., Jin, C., Hamana, T., Liu, J., Wang, C., An, L., McKeehan, W.L., and Wang, F. (2015). *Overexpression of FGF9 in Prostate Epithelial Cells Augments Reactive Stroma Formation and Promotes Prostate Cancer Progression*. Int. J. Biol. Sci. *11*, 948–960.

Hung, C.-Y., Yang, W.-B., Wang, S.-A., Hsu, T.-I., Chang, W.-C., and Hung, J.-J. (2014). *Nucleolin enhances internal ribosomal entry site (IRES)-mediated translation of Sp1 in tumourigenesis*. Biochim. Biophys. Acta BBA - Mol. Cell Res.

Hwang, J., and Kim, Y.K. (2013). *When a ribosome encounters a premature termination codon*. BMB Rep. *46*, 9–16.

Hwang, J., and Maquat, L.E. (2011). *Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question*. Curr. Opin. Genet. Dev. *21*, 422–430.

Iglesias, N., and Stutz, F. (2008). *Regulation of mRNP dynamics along the export pathway*. FEBS Lett. *582*, 1987–1996.

Imamachi (2012). *Up-frameshift protein 1 (UPF1): Multitalented entertainer in RNA decay*. Drug Discov. Ther.

Isidro, G., Matos, S., Gonçalves, V., Cavaleiro, C., Antunes, O., Marinho, C., Soares, J., and Boavida, M.G. (2003). *Novel MLH1 mutations and a novel MSH2 polymorphism identified by SSCP and DHPLC in Portuguese HNPCC families*. Hum. Mutat. *22*, 419–420.

Isken, O., Kim, Y.K., Hosoda, N., Mayeur, G.L., Hershey, J.W.B., and Maquat, L.E. (2008). *Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay*. Cell *133*, 314–327.

Ito, E., Yanagisawa, Y., Iwahashi, Y., Suzuki, Y., Nagasaki, H., Akiyama, Y., Sugano, S., Yuasa, Y., and Maruyama, K. (1999). *A Core Promoter and a Frequent Single-Nucleotide Polymorphism of the Mismatch Repair GenehMLH1*. Biochem. Biophys. Res. Commun. *256*, 488–494.

Jackson, R.J. (1991). *mRNA translation. Initiation without an end*. Nature *353*, 14–15.

Jackson, R.J. (2000). *A Comparative View of Initiation Site Selection Mechanisms*. Cold Spring Harb. Monogr. Arch. *39*, 127–183.

Jackson, R.J. (2013). *The Current Status of Vertebrate Cellular mRNA IRESs*. Cold Spring Harb. Perspect. Biol. *5*, a011569–a011569.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). *The mechanism of eukaryotic translation initiation and principles of its regulation*. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jang, S.K., Kräusslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988). *A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation*. J. Virol. *62*, 2636–2643.

Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R.E., Kim, Y.J., Dinh, T.T., Liu, J., Cui, X., Zheng, B., *et al.* (2011). *Argonaute10 and Argonaute1 Regulate the Termination of Floral Stem Cells through Two MicroRNA in Arabidopsis*. PLoS Genet *7*, e1001358.

Johnson, R.E., Klassen, R., Prakash, L., and Prakash, S. (2015). *A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands*. Mol. Cell *59*, 163–175.

Kapp, L.D., and Lorsch, J.R. (2004a). *GTP-dependent Recognition of the Methionine Moiety on Initiator tRNA by Translation Factor eIF2*. J. Mol. Biol. *335*, 923–936.

Kapp, L.D., and Lorsch, J.R. (2004b). *The molecular mechanics of eukaryotic translation*. Annu. Rev. Biochem. *73*, 657–704.

Karaskova, M., Gunisova, S., Herrmannova, A., Wagner, S., Munzarova, V., and Valasek, L.S. (2012). *Functional Characterization of the Role of the N-terminal Domain of the c/Nip1 Subunit of Eukaryotic Initiation Factor 3 (eIF3) in AUG Recognition*. J. Biol. Chem. *287*, 28420–28434.

Karetnikov, A., and Lehto, K. (2008). *Translation mechanisms involving long-distance base pairing interactions between the 5' and 3' non-translated regions and internal ribosomal entry are conserved for both genomic RNA of Blackcurrant reversion nepovirus*. Virology *371*, 292–308.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). *Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay*. Genes Dev. *20*, 355–367.

Katahira, J. (2015). *Nuclear Export of Messenger RNA*. Genes *6*, 163–184.

Kaygun, H., and Marzluff, W.F. (2005). *Regulated degradation of replication-dependent histone mRNA requires both ATR and Upf1*. Nat. Struct. 38 Mol. Biol. *12*, 794–800.

Kempf, E., Rousseau, B., Besse, B., and Paz-Ares, L. (2016). KRAS oncogene in lung cancer: focus on molecularly driven clinical trials. Eur. Respir. Rev. *25*, 71–76.

Khan, D., Sharathchandra, A., Ponnuswamy, A., Grover, R., and Das, S. (2013). *Effect of a natural mutation in the 5′ untranslated region on the translational control of p53 mRNA.* Oncogene *32*, 4148–4159.

Khan, D., Katoch, A., Das, A., Sharathchandra, A., Lal, R., Roy, P., Das, S., Chattopadhyay, S., and Das, S. (2015). *Reversible induction of translational isoforms of p53 in glucose deprivation*. Cell Death Differ. *22*, 1203–1218.

Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences*. Nucleic Acids Res. *34*, W676-682.

Kim, Y.K., Furic, L., Desgroseillers, L., and Maquat, L.E. (2005). *Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTR so as to elicit mRNA decay*. Cell *120*, 195–208.

Kim, C. Y., Jung, W. Y., Lee, H. J., Kim, H. K., Kim, A., & Shin, B. K. (2012). *Proteomic analysis reveals overexpression of moesin and cytokeratin 17 proteins in colorectal carcinoma*. *Oncology reports*, *27*(3), 608-620.

King, H.A., Cobbold, L.C., and Willis, A.E. (2010). *The role of IRES trans-acting factors in regulating translation initiation*. Biochem. Soc. Trans. *38*, 1581.

Kobayashi, H., and Tomari, Y. (2016). *RISC assembly: Coordination between small RNA and Argonaute proteins*. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1859*, 71–81.

Kochetov, A.V., Ahmad, S., Ivanisenko, V., Volkova, O.A., Kolchanov, N.A., and Sarai, A. (2008). *uORFs, reinitiation and alternative translation start sites in human mRNA*. FEBS Lett. *582*, 1293–1297.

Koesters, R., Adams, V., Betts, D., Moos, R., Schmid, M., Siermann, A., Hassam, S., Weitz, S., Lichter, P., Heitz, P.U., *et al.* (1999). *Human Eukaryotic Initiation Factor EIF2C1 Gene: cDNA Sequence, Genomic Organization, Localization to Chromosomal Bands 1p34–p35, and Expression*. Genomics *61*, 210–218.

Koh, D.C., and Mauro, V.P. (2009). *Reconciling contradictory reports regarding translation of BACE1 mRNA: initiation mechanism is altered by different expression systems*. RNA Biol. *6*, 54–58.

Kolosov, P., Frolova, L., Seit-Nebi, A., Dubovaya, V., Kononenko, A., Oparina, N., Justesen, J., Efimov, A., and Kisselev, L. (2005). *Invariant amino acids essential for decoding function of polypeptide release factor eRF1*. Nucleic Acids Res. *33*, 6418–6425.

Kolupaeva, V.G. (2005). *Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association*. RNA *11*, 470–486.

Kolupaeva, V.G., Pestova, T.V., Hellen, C.U., and Shatsky, I.N. (1998). *Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA*. J. Biol. Chem. *273*, 18599–18604.

Komar, A.A., and Hatzoglou, M. (2005). *Internal Ribosome Entry Sites in Cellular mRNA: Mystery of Their Existence*. J. Biol. Chem. *280*, 23425–23428.

Komar, A.A., and Hatzoglou, M. (2011). *Cellular IRES-mediated translation: The war of ITAFs in pathophysiological states*. Cell Cycle *10*, 229–240.

Kong, C., Ito, K., Walsh, M.A., Wada, M., Liu, Y., Kumar, S., Barford, D., Nakamura, Y., and Song, H. (2004). *Crystal structure and functional analysis of the eukaryotic class II release factor eRF3 from S. pombe*. Mol. Cell *14*, 233–245.

Kozak, M. (1979). *Migration of 40 S ribosomal subunits on messenger RNA when initiation is perturbed by lowering magnesium or adding drugs*. J. Biol. Chem. *254*, 4731–4738.

Kozak, M. (1989). *The scanning model for translation: an update*. J. Cell Biol. *108*, 229–241.

Kozak, M. (1991). *A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes*. Gene Expr. *1*, 111–115.

Kozak, M. (2002). *Pushing the limits of the scanning mechanism for initiation of translation*. Gene *299*, 1–34.

Kress, T.R., Sabò, A., and Amati, B. (2015). *MYC: connecting selective transcriptional control to global RNA production*. Nat. Rev. Cancer *15*, 593–607.

Kuhle, B., and Ficner, R. (2014). *eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining*. EMBO J. *33*, 1177–1191.

Kukuła, K., Chojnowska, L., Dąbrowski, M., Witkowski, A., Chmielak, Z., Skwarek, M., Kądziela, J., Teresińska, A., Małecki, M., Janik, P., *et al.* (2011). *Intramyocardial plasmid-encoding human vascular endothelial growth factor A165/basic fibroblast growth factor therapy using percutaneous transcatheter approach in patients with refractory coronary artery disease (VIF-CAD)*. Am. Heart J. *161*, 581–589.

Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). *An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation*. Nat. Chem. Biol. *3*, 218–221.

Kumari, S., Bugaut, A., and Balasubramanian, S. (2008). *Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5′ UTR of the NRAS proto-oncogene*. Biochemistry (Mosc.) *47*, 12664–12669.

Kupatt, C., Hinkel, R., Pfosser, A., El-Aouni, C., Wuchrer, A., Fritz, A., Globisch, F., Thormann, M., Horstkotte, J., Lebherz, C., *et al.* (2010). *Cotransfection of vascular endothelial growth factor-A and platelet-derived growth factor-B via recombinant adeno-associated virus resolves chronic ischemic malperfusion role of vessel maturation*. J. Am. Coll. Cardiol. *56*, 414–422.

Kupčinskaitė-Noreikienė, R., Skiecevičienė, J., Jonaitis, L., Ugenskienė, R., Kupčinskas, J., Markelis, R., Baltrėnas, V., Sakavičius, L., Semakina, I., Grižas, S., *et al.* (2013). *CpG island methylation of the MLH1, MGMT, DAPK, and CASP8 genes in cancerous and adjacent noncancerous stomach tissues*. Med. Kaunas Lith. *49*, 361–366.

Kuroha, K., Tatematsu, T., and Inada, T. (2009). *Upf1 stimulates degradation of the product derived from aberrant messenger RNA containing a specific nonsense mutation by the proteasome*. EMBO Rep. *10*, 1265–1271.

Kwabi-Addo, B., Ozen, M., and Ittmann, M. (2004). *The role of fibroblast growth factors and their receptors in prostate cancer*. Endocr. Relat. Cancer *11*, 709–724.

Lang, K.J., Kappel, A., and Goodall, G.J. (2002). *Hypoxia-inducible factor-1α mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia*. Mol. Biol. Cell *13*, 1792–1801.

Le, S.Y., and Maizel Jr, J.V. (1997). *A common RNA structural motif involved in the internal initiation of translation of cellular mRNA*. Nucleic Acids Res. *25*, 362–369.

Lee, J.H., Pestova, T.V., Shin, B.-S., Cao, C., Choi, S.K., and Dever, T.E. (2002). *Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation*. Proc. Natl. Acad. Sci. U. S. A. *99*, 16689–16694.

Lee, S.R., Pratt, G.A., Martinez, F.J., Yeo, G.W., and Lykke-Andersen, J. (2015). *Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity*. Mol. Cell *59*, 413–425.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). *MicroRNA genes are transcribed by RNA polymerase II*. EMBO J. *23*, 4051–4060.

Leprivier, G., Rotblat, B., Khan, D., Jan, E., and Sorensen, P.H. (2015). *Stress-mediated translational control in cancer cells*. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1849*, 845–860.

Leroy, C., Fialin, C., Sirvent, A., Simon, V., Urbach, S., Poncet, J., Robert, B., Jouin, P., and Roche, S. (2009). *Quantitative Phosphoproteomics Reveals a Cluster of Tyrosine Kinases That Mediates Src Invasive Activity in Advanced Colon Carcinoma Cells*. Cancer Res. *69*, 2279–2286.

Lewis, S.M., and Holcik, M. (2005). *IRES in distress: translational regulation of the inhibitor of apoptosis proteins XIAP and HIAP2 during cell stress.* Cell Death Differ. *12*, 547–553.

Lewis, S.M., and Holcik, M. (2008). *For IRES trans-acting factors, it is all about location*. Oncogene *27*, 1033–1035.

Lewis, S.M., Cerquozzi, S., Graber, T.E., Ungureanu, N.H., Andrews, M., and Holcik, M. (2007). *The eIF4G homolog DAP5/p97 supports the translation of select mRNA during endoplasmic reticulum stress*. Nucleic Acids Res. *36*, 168–178.

Li, G.-M. (2008). *Mechanisms and functions of DNA mismatch repair*. Cell Res. *18*, 85–98.

Li, G., Wang, N., Sun, C., and Li, B. (2014). *Decreased expression of eukaryotic initiation factor 3f is an adverse prognostic factor for stage I-III gastric cancer*. World J. Surg. Oncol. *12*, 72.

Li, L., Yu, C., Gao, H., and Li, Y. (2010a). *Argonaute proteins: potential biomarkers for human colon cancer*. BMC Cancer *10*, 38.

Li, W., Thakor, N., Xu, E.Y., Huang, Y., Chen, C., Yu, R., Holcik, M., and Kong, A.-N. (2010b). *An internal ribosomal entry site mediates redox-sensitive translation of Nrf2*. Nucleic Acids Res. *38*, 778–788.

Li, X., Yao, X., Wang, Y., Hu, F., Wang, F., Jiang, L., Liu, Y., Wang, D., Sun, G., and Zhao, Y. (2013). *MLH1 promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features*. PloS One *8*, e59064.

Li, X., Zhang, Q., Fan, K., Li, B., Li, H., Qi, H., Guo, J., Cao, Y., and Sun, H. (2016). *Overexpression of TRPV3 Correlates with Tumor Progression in Non-Small Cell Lung Cancer*. Int. J. Mol. Sci. *17*.

Liberman, N., Gandin, V., Svitkin, Y.V., David, M., Virgili, G., Jaramillo, M., Holcik, M., Nagar, B., Kimchi, A., and Sonenberg, N. (2015). *DAP5 associates with eIF2 and eIF4AI to promote Internal Ribosome Entry Site driven translation*. Nucleic Acids Res. *43*, 3764–3775.

Lien, I.-C., Horng, L.-Y., Hsu, P.-L., Wu, C.-L., Sung, H.-C., and Wu, R.-T. (2014). *Internal ribosome entry site of bFGF is the target of thalidomide for IMiDs development in multiple myeloma*. Genes Cancer *5*, 127.

Liu, B., and Qian, S.-B. (2014). *Translational reprogramming in cellular stress response: Translational reprogramming in stress*. Wiley Interdiscip. Rev. RNA *5*, 301–305.

Liu, T., Zhang, H., Xiong, J., Yi, S., Gu, L., and Zhou, M. (2015). *Inhibition of MDM2 homodimerization by XIAP IRES stabilises MDM2, influencing cancer cell survival*. Mol. Cancer *14*.

Liu, Y., Hidayat, S., Su, W., Deng, X., Yu, D., and Yu, B. (2007). *Expression and activity of mTOR and its substrates in different cell cycle phases and in oral squamous cell carcinomas of different malignant grade*. Cell Biochem. Funct. *25*, 45–53.

Liu, Y., Du, F., Chen, W., Yao, M., Lv, K., and Fu, P. (2014). *EIF5A2 is a novel chemoresistance gene in breast cancer*. Breast Cancer.

Lomakin, I.B., and Steitz, T.A. (2013). *The initiation of mammalian protein synthesis and mRNA scanning mechanism*. Nature *500*, 307–311.

López-Lastra, M., Rivas, A., and Barría, M.I. (2005). *Protein synthesis in Eukaryotes; the growing biological relevance of cap-independent translation initiation*. Biol. Res. *38*, 121.

Lozano, G., and Martínez-Salas, E. (2015). *Structural insights into viral IRES-dependent translation mechanisms.* Curr. Opin. Virol. *12*, 113–120.

Luna, R., Gaillard, H., González-Aguilera, C., and Aguilera, A. (2008). *Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus*. Chromosoma *117*, 319–331.

Luo, A., Kong, J., Hu, G., Liew, C.-C., Xiong, M., Wang, X., Ji, J., Wang, T., Zhi, H., Wu, M., *et al.* (2004). *Discovery of Ca^{2+}-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray*. Oncogene *23*, 1291–1299.

Macejak, D.G., and Sarnow, P. (1991). *Internal initiation of translation mediated by the 5[prime] leader of a cellular mRNA*. Nature *353*, 90–94.

Majumdar, R. (2003). *Mammalian Translation Initiation Factor eIF1 Functions with eIF1A and eIF3 in the Formation of a Stable 40 S Preinitiation Complex*. J. Biol. Chem. *278*, 6580–6587.

Malbert-Colas, L., Ponnuswamy, A., Olivares-Illana, V., Tournillon, A.-S., Naski, N., and Fåhraeus, R. (2014). *HDMX Folds the Nascent p53 mRNA following Activation by the ATM Kinase*. Mol. Cell *54*, 500–511.

Maquat, L.E. (2005). *Nonsense-mediated mRNA decay in mammals*. J. Cell Sci. *118*, 1773–1776.

Maquat, L.E., and Gong, C. (2009). *Gene expression networks: competing mRNA decay pathways in mammalian cells*. Biochem. Soc. Trans. *37*, 1287–1292.

Marcel, V., Ghayad, S.E., Belin, S., Therizols, G., Morel, A.-P., Solano-Gonzàlez, E., Vendrell, J.A., Hacot, S., Mertani, H.C., Albaret, M.A., *et al.* (2013). *p53 Acts as a Safeguard of Translational Control by Regulating Fibrillarin and rRNA Methylation in Cancer*. Cancer Cell *24*, 318–330.

Marcotrigiano, J., Gingras, A.C., Sonenberg, N., and Burley, S.K. (1999). *Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G*. Mol. Cell *3*, 707–716.

Marques-Ramos, A. (2013). *Translational regulation mediated by internal ribosome entry sites of the mTOR and d133P53 human transcripts*. Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Martineau, Y., Le Bec, C., Monbrun, L., Allo, V., Chiu, I.-M., Danos, O., Moine, H., Prats, H., and Prats, A.-C. (2004). *Internal ribosome entry site structural motifs conserved among mammalian fibroblast growth factor 1 alternatively spliced mRNA*. Mol. Cell. Biol. *24*, 7622–7635.

Martínez-Salas, E., Lozano, G., Fernandez-Chamorro, J., Francisco-Velilla, R., Galan, A., and Diaz, R. (2013). *RNA-Binding Proteins Impacting on Internal Initiation of Translation*. Int. J. Mol. Sci. *14*, 21705–21726.

Marzluff, W.F. (2005). *Metazoan replication-dependent histone mRNA: a distinct set of RNA polymerase II transcripts*. Curr. Opin. Cell Biol. *17*, 274–280.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). *Metabolism and regulation of canonical histone mRNA: life without a poly(A) tail*. Nat. Rev. Genet. *9*, 843–854.

Masoud, G.N., and Li, W. (2015). *HIF-1α pathway: role, regulation and intervention for cancer therapy*. Acta Pharm. Sin. B *5*, 378–389.

Mathews, M. B., Sonenberg, N., Hershey, J. W. B. Editors. (2007) Traslational control in Biology and Medicine. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Mei, M., Liu, D., Dong, S., Ingvarsson, S., Goodfellow, P.J., and Chen, H. (2010). *The MLH1 −93 promoter variant influences gene expression*. Cancer Epidemiol. *34*, 93–95.

Mendell, J.T., ap Rhys, C.M.J., and Dietz, H.C. (2002). *Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts*. Science *298*, 419–422.

Merkulova, T.I., Frolova, L.Y., Lazar, M., Camonis, J., and Kisselev, L.L. (1999). *C-terminal domains of human translation termination factors eRF1 and eRF3 mediate their in vivo interaction*. FEBS Lett. *443*, 41–47.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). *Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTR and near Stop Codons*. Cell *149*, 1635–1646.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., and Jaffrey, S.R. (2015). *5' UTR m6A Promotes Cap-Independent Translation. Cell 163*, 999–1010.

Mihailovich, M., Thermann, R., Grohovaz, F., Hentze, M.W., and Zacchetti, D. (2007). *Complex translational regulation of BACE1 involves upstream AUG and stimulatory elements within the 5' untranslated region*. Nucleic Acids Res. *35*, 2975–2985.

Milev, M.P., Ravichandran, M., Khan, M.F., Schriemer, D.C., and Mouland, A.J. (2012). *Characterization of Staufen1 Ribonucleoproteins by Mass Spectrometry and Biochemical Analyses Reveal the Presence of Diverse Host Proteins Associated with Human Immunodeficiency Virus Type 1*. Front. Microbiol. *3*.

Moerke, N.J., Aktas, H., Chen, H., Cantel, S., Reibarkh, M.Y., Fahmy, A., Gross, J.D., Degterev, A., Yuan, J., Chorev, M., *et al.* (2007). *Small-Molecule Inhibition of the Interaction between the Translation Initiation Factors eIF4E and eIF4G*. Cell *128*, 257–267.

Moffat, J.G., and Tate, W.P. (1994). *A single proteolytic cleavage in release factor 2 stabilises ribosome binding and abolishes peptidyl-tRNA hydrolysis activity*. J. Biol. Chem. *269*, 18899–18903.

Mokrejš, M., Mašek, T., Vopálenskỳ, V., Hlubuček, P., Delbos, P., and Pospíšek, M. (2010). *IRESite–a tool for the examination of viral and cellular internal ribosome entry sites*. Nucleic Acids Res. *38*, D131.

Montanaro, L., Calienni, M., Bertoni, S., Rocchi, L., Sansone, P., Storci, G., Santini, D., Ceccarelli, C., Taffurelli, M., Carnicelli, D., *et al.* (2010). *Novel dyskerin-mediated mechanism of p53 inactivation through defective mRNA translatio*n. Cancer Res. *70*, 4767–4777.

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). *Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells*. Nat. Cell Biol. *12*, 543–552.

Mora, L., Heurgué-Hamard, V., Champ, S., Ehrenberg, M., Kisselev, L.L., and Buckingham, R.H. (2003). *The essential role of the invariant GGQ motif in the function and stability in vivo of bacterial release factors RF1 and RF2*. Mol. Microbiol. *47*, 267–275.

Morfoisse, F., Kuchnio, A., Frainay, C., Gomez-Brouchet, A., Delisle, M.-B., Marzi, S., Helfer, A.-C., Hantelys, F., Pujol, F., Guillermet-Guibert, J., *et al.* (2014). *Hypoxia Induces VEGF-C Expression in Metastatic Tumor Cells via a HIF-1α-Independent Translation-Mediated Mechanism*. Cell Rep. *6*, 155–167.

Morris, M., Negishi, Y., Pazsint, C., Schonhoft, J., and Basu, S. (2010). *An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES*. J Am Chem Soc *132*, 17831–17839.

Mühlemann, O., and Jensen, T.H. (2012). *mRNP quality control goes regulatory*. Trends Genet. *28*, 70–77.

Mühlemann, O., and Lykke-Andersen, J. (2010). *How and where are nonsense mRNA degraded in mammalian cells?* RNA Biol. *7*, 28–32.

Müller, B., Blackburn, J., Feijoo, C., Zhao, X., and Smythe, C. (2007). *DNA-activated protein kinase functions in a newly observed S phase checkpoint that links histone mRNA abundance with DNA replication*. J. Cell Biol. *179*, 1385–1398.

Müller-McNicoll, M., and Neugebauer, K.M. (2013). *How cells get the message: dynamic assembly and function of mRNA–protein complexes*. Nat. Rev. Genet. *14*, 275–287.

Nakayama, K. (2009). *Cellular Signal Transduction of the Hypoxia Response*. J. Biochem. (Tokyo) *146*, 757–765.

NCBI, R. Entrez Gene: *MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2* (E. coli).

Neugebauer, K.M. (2002). *On the importance of being co-transcriptional*. J. Cell Sci. *115*, 3865–3871.

Nguyen, K.T., Holloway, M.P., and Altura, R.A. (2012). *The CRM1 nuclear export protein in normal development and disease*. Int. J. Biochem. Mol. Biol. *3*, 137.

Oberer, M., Marintchev, A., and Wagner, G. (2005). *Structural basis for the enhancement of eIF4A helicase activity by eIF4G*. Genes Dev. *19*, 2212–2223.

Ogawa, A. (2013). *Ligand-Dependent Upregulation of Ribosomal Shunting*. ChemBioChem *14*, 1539–1543.

Ohlfest, J.R., Demorest, Z.L., Motooka, Y., Vengco, I., Oh, S., Chen, E., Scappaticci, F.A., Saplis, R.J., Ekker, S.C., Low, W.C., *et al.* (2005). *Combinatorial antiangiogenic gene therapy by nonviral gene transfer using the sleeping beauty transposon causes tumor regression and improves survival in mice bearing intracranial human glioblastoma*. Mol. Ther. J. Am. Soc. Gene Ther. *12*, 778–788.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2012). *N-and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5: SMG-7 during NMD*. Nucleic Acids Res. *40*, 1251–1266.

Olsen, D.S., Savner, E.M., Mathew, A., Zhang, F., Krishnamoorthy, T., Phan, L., and Hinnebusch, A.G. (2003). *Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo*. EMBO J. *22*, 193–204.

Onofre, C., Tomé, F., Barbosa, C., Silva, A.L., and Romão, L. (2015). *Expression of human hemojuvelin is tightly regulated by two upstream open reading frames that respond to iron overload in hepatic cells*. Mol. Cell. Biol. MCB.01462-14.

Orphanides, G., and Reinberg, D. (2002). *A unified theory of gene expression*. Cell *108*, 439–451.

Ott, G., Rosenwald, A., and Campo, E. (2013). *Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification*. Hematol. Educ. Program Am. Soc. Hematol. Am. Soc. Hematol. Educ. Program *2013*, 575–583.

Özeş, A.R., Feoktistova, K., Avanzino, B.C., and Fraser, C.S. (2011). *Duplex Unwinding and ATPase Activities of the DEAD-Box Helicase eIF4A Are Coupled by eIF4G and eIF4B*. J. Mol. Biol. *412*, 674–687.

Ozretić, P., Bisio, A., Musani, V., Trnski, D., Sabol, M., Levanat, S., and Inga, A. (2015). *Regulation of human PTCH1b expression by different 5' untranslated region cis -regulatory elements*. RNA Biol. *12*, 290–304.

Paek, K.Y., Hong, K.Y., Ryu, I., Park, S.M., Keum, S.J., Kwon, O.S., and Jang, S.K. (2015). *Translation initiation mediated by RNA looping*. Proc. Natl. Acad. Sci. *112*, 1041–1046.

Pal, T., Permuth-Wey, J., and Sellers, T.A. (2008). *A review of the clinical relevance of mismatch-repair deficiency in ovarian cancer*. Cancer *113*, 733–742.

Palumbo, A., Da Costa, N.M., Esposito, F., De Martino, M., D'Angelo, D., de Sousa, V.P.L., Martins, I., Nasciutti, L.E., Fusco, A., and Pinto, L.F.R. (2016). *HMGA2 overexpression plays a critical role in the progression of esophageal squamous carcinoma*. Oncotarget. *7*(18), 25872-25884

Parisi, C., Giorgi, C., Batassa, E.M., Braccini, L., Maresca, G., D'agnano, I., Caputo, V., Salvatore, A., Pietrolati, F., Cogoni, C., *et al.* (2011). *Ago1 and Ago2 differentially affect cell proliferation, motility and apoptosis when overexpressed in SH-SY5Y neuroblastoma cells*. FEBS Lett. *585*, 2965–2971.

Pause, A., and Sonenberg, N. (1992). *Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A*. EMBO J. *11*, 2643–2654.

Pelletier, J., and Sonenberg, N. (1988). *Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA*. Nature *334*, 320–325.

Perlick, H.A., Medghalchi, S.M., Spencer, F.A., Kendzior, R.J., and Dietz, H.C. (1996). *Mammalian orthologues of a yeast regulator of nonsense transcript stability*. Proc. Natl. Acad. Sci. *93*, 10928–10932.

Pestova, T.V., and Kolupaeva, V.G. (2002). *The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection*. Genes Dev. *16*, 2906–2922.

Pestova, T.V., Borukhov, S.I., and Hellen, C.U.T. (1998). *Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons*. Nature *394*, 854–859.

Pestova, T.V., Lomakin, I.B., Lee, J.H., Choi, S.K., Dever, T.E., and Hellen, C.U. (2000). *The joining of ribosomal subunits in eukaryotes requires eIF5B*. Nature *403*, 332–335.

Pestova, T.V., Lorsch, J.R., and Hellen, C.U.T. (2007). *The mechanisms of translation initiation in eukaryotes.* In: M.Mathews, N. Sonenberg, J.W.B. Hershey (Eds.), Translational control in biology and medicine (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Phatnani, H.P., and Greenleaf, A.L. (2006). *Phosphorylation and functions of the RNA polymerase II CTD*. Genes Dev. *20*, 2922–2936.

Piccirillo, C.A., Bjur, E., Topisirovic, I., Sonenberg, N., and Larsson, O. (2014). *Translational control of immune responses: from transcripts to translatomes*. Nat. Immunol. *15*, 503–511.

Pickering, B.M., Mitchell, S.A., Spriggs, K.A., Stoneley, M., and Willis, A.E. (2004). *Bag-1 internal ribosome entry segment activity is promoted by structural changes mediated by poly(rC) binding protein 1 and recruitment of polypyrimidine tract binding protein 1*. Mol. Cell. Biol. *24*, 5595–5605.

Pisarev, A.V., Hellen, C.U.T., and Pestova, T.V. (2007). *Recycling of eukaryotic posttermination ribosomal complexes*. Cell *131*, 286–299.

Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U.T., and Pestova, T.V. (2010). *The role of ABCE1 in eukaryotic posttermination ribosomal recycling*. Mol. Cell *37*, 196–210.

Pisareva, V.P., and Pisarev, A.V. (2014). *eIF5 and eIF5B together stimulate 48S initiation complex formation during ribosomal scanning*. Nucleic Acids Res.

Pisareva, V.P., and Pisarev, A.V. (2016). *DHX29 reduces leaky scanning through an upstream AUG codon regardless of its nucleotide context*. Nucleic Acids Res.

Pogue-Geile, K., Geiser, J.R., Shu, M., Miller, C., Wool, I.G., Meisler, A.I., and Pipas, J.M. (1991). *Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein*. Mol. Cell. Biol. *11*, 3842–3849.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). *Integrating mRNA processing with transcription*. Cell *108*, 501–512.

Pyronnet, S., Pradayrol, L., and Sonenberg, N. (2000). *A cell cycle-dependent internal ribosome entry site*. Mol. Cell *5*, 607–616.

Qi, J., Dong, Z., Liu, J., and Zhang, J.-T. (2014). *EIF3i promotes colon oncogenesis by regulating COX-2 protein synthesis and β-catenin activation*. Oncogene *33*, 4156–4163.

Ray, P.S., Grover, R., and Das, S. (2006). *Two internal ribosome entry sites mediate the translation of p53 isoforms*. EMBO Rep. *7*, 404–410.

Renaud-Gabardos, E. (2015). *Internal ribosome entry site-based vectors for combined gene therapy*. World J. Exp. Med. *5*, 11.

Richter-Cook, N.J. (1998). *Purification and Characterization of a New Eukaryotic Protein Translation Factor: eukaryotic initiation factor 4H*. J. Biol. Chem. *273*, 7579–7587.

Riley, A., Jordan, L.E., and Holcik, M. (2010). *Distinct 5' UTR regulate XIAP expression under normal growth conditions and during cellular stress*. Nucleic Acids Res. *38*, 4665–4674.

Roberts, R., Zhang, J., Mayberry, L.K., Tatineni, S., Browning, K.S., and Rakotondrafara, A. (2015). *A unique 5' translation element discovered in Triticum mosaic virus* . J. Virol. JVI.02099-15.

Rogers, G.W., Edelman, G.M., and Mauro, V.P. (2004). *Differential utilization of upstream AUG in the -secretase mRNA suggests that a shunting mechanism regulates translation*. Proc. Natl. Acad. Sci. *101*, 2794–2799.

Rohban, S., and Campaner, S. (2015). *Myc induced replicative stress response: How to cope with it and exploit it*. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1849*, 517–524.

Rougemaille, M., Villa, T., Gudipati, R.K., and Libri, D. (2008). *mRNA journey to the cytoplasm: attire required*. Biol. Cell *100*, 327–342.

Rozovsky, N., Butterworth, A.C., and Moore, M.J. (2008). *Interactions between eIF4AI and its accessory factors eIF4B and eIF4H*. RNA *14*, 2136–2148.

Rubsamen, D., Blees, J.S., Schulz, K., Doring, C., Hansmann, M.-L., Heide, H., Weigert, A., Schmid, T., and Brune, B. (2012). *IRES-dependent translation of egr2 is induced under inflammatory conditions*. RNA *18*, 1910–1920.

Rubtsova, M.P., Sizova, D.V., Dmitriev, S.E., Ivanov, D.S., Prassolov, V.S., and Shatsky, I.N. (2003). *Distinctive properties of the 5'-untranslated region of human hsp70 mRNA*. J. Biol. Chem. *278*, 22350–22356.

Safar, A.M., Spencer, H., Su, X., Coffey, M., Cooney, C.A., Ratnasinghe, L.D., Hutchins, L.F., and Fan, C.-Y. (2005). *Methylation profiling of archived non-small cell lung cancer: a promising prognostic system*. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *11*, 4400–4405.

Salas-Marco, J., and Bedwell, D.M. (2004). *GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination*. Mol. Cell. Biol. *24*, 7769–7778.

Salvati, E., Zizza, P., Rizzo, A., Iachettini, S., Cingolani, C., D'Angelo, C., Porru, M., Randazzo, A., Pagano, B., Novellino, E., *et al.* (2014). *Evidence for G-quadruplex in the promoter of vegfr-2 and its targeting to inhibit tumor angiogenesis*. Nucleic Acids Res. *42*, 2945–2957.

Sammons, M.A., Antons, A.K., Bendjennat, M., Udd, B., Krahe, R., and Link, A.J. (2010). *ZNF9 Activation of IRES-Mediated Translation of the Human ODC mRNA Is Decreased in Myotonic Dystrophy Type 2*. PLoS ONE *5*, e9301.

Scappaticci, F.A., Smith, R., Pathak, A., Schloss, D., Lum, B., Cao, Y., Johnson, F., Engleman, E.G., and Nolan, G.P. (2001). *Combination angiostatin and endostatin gene transfer induces synergistic antiangiogenic activity in vitro and antitumor efficacy in leukemia and solid tumors in mice*. Mol. Ther. J. Am. Soc. Gene Ther. *3*, 186–196.

Scarpa, M., Scarpa, M., Castagliuolo, I., Erroi, F., Kotsafti, A., Basato, S., Brun, P., D'Incà, R., Rugge, M., Angriman, I., *et al.* (2016). *Aberrant gene methylation in non-neoplastic mucosa as a predictive marker of ulcerative colitis-associated CRC*. Oncotarget.

Schepens, B., Tinton, S.A., Bruynooghe, Y., Parthoens, E., Haegman, M., Beyaert, R., and Cornelis, S. (2007). *A role for hnRNP C1/C2 and Unr in internal initiation of translation during mitosis*. EMBO J. *26*, 158–169.

Schimmer, A.D. (2004). *Inhibitor of apoptosis proteins: translating basic knowledge into clinical practice*. Cancer Res. *64*, 7183–7190.

Schoenberg, D.R., and Maquat, L.E. (2012). *Regulation of cytoplasmic mRNA decay*. Nat. Rev. Genet.

Sekiyama, N., Arthanari, H., Papadopoulos, E., Rodriguez-Mias, R.A., Wagner, G., and Léger-Abraham, M. (2015). *Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilising the binding of unphosphorylated 4E-BP1*. Proc. Natl. Acad. Sci. *112*, E4036–E4045.

Semler, B.L., and Waterman, M.L. (2008). *IRES-mediated pathways to polysomes: nuclear versus cytoplasmic routes*. Trends Microbiol. *16*, 1–5.

Serquina, A.K.P., Das, S.R., Popova, E., Ojelabi, O.A., Roy, C.K., and Gottlinger, H.G. (2013). *UPF1 Is Crucial for the Infectivity of Human Immunodeficiency Virus Type 1 Progeny Virions*. J. Virol. *87*, 8853–8861.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). *Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p^{16INK4}a*. Cell *88*, 593–602.

Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L., and de Lange, T. (2009). M*ammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication*. Cell *138*, 90–103.

Shahbazian, D., Parsyan, A., Petroulakis, E., Topisirovic, I., Martineau, Y., Gibbs, B.F., Svitkin, Y., and Sonenberg, N. (2010). *Control of Cell Survival and Proliferation by Mammalian Eukaryotic Initiation Factor 4B*. Mol. Cell. Biol. *30*, 1478–1485.

Sharathchandra, A., Lal, R., Khan, D., and Das, S. (2012). *Annexin A2 and PSF proteins interact with p53 IRES and regulate translation of p53 mRNA*. RNA Biol. *9*, 1429–1439.

Shatsky, I. N., Dmitriev, S. E., Terenin, I. M., & Andreev, D. E. (2010). Cap-and IRES-independent scanning mechanism of translation initiation as an alternative to the concept of cellular IRESs. *Molecules and cells*, *30*(4), 285-293.

Shay, K.P., Michels, A.J., Li, W., Kong, A.-N.T., and Hagen, T.M. (2012). *Cap-independent Nrf2 translation is part of a lipoic acid-stimulated detoxification stress response*. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1823*, 1102–1109.

Shen, J., Yin, J.-Y., Li, X.-P., Liu, Z.-Q., Wang, Y., Chen, J., Qu, J., Xu, X.-J., McLeod, H.L., He, Y.-J., *et al.* (2014). *The Prognostic Value of Altered eIF3a and Its Association with p27 in Non-Small Cell Lung Cancers*. PLoS ONE *9*, e96008.

Sherrill, K.W., and Lloyd, R.E. (2008). *Translation of cIAP2 mRNA Is Mediated Exclusively by a Stress-Modulated Ribosome Shunt*. Mol. Cell. Biol. *28*, 2011–2022.

Sherrill, K.W., Byrd, M.P., Van Eden, M.E., and Lloyd, R.E. (2004). *BCL-2 translation is mediated via internal ribosome entry during cell stress*. J. Biol. Chem. *279*, 29066–29074.

Shi, Y., and Manley, J.L. (2015). *The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site*. Genes Dev. *29*, 889–897.

Shi, Y., Yang, Y., Hoang, B., Bardeleben, C., Holmes, B., Gera, J., and Lichtenstein, A. (2016). *Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress*. Oncogene *35*, 1015–1024.

Shin, B.-S., Maag, D., Roll-Mecak, A., Arefin, M.S., Burley, S.K., Lorsch, J.R., and Dever, T.E. (2002). *Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity*. Cell *111*, 1015–1025.

Shoemaker, C.J., and Green, R. (2011). *Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast*. Proc. Natl. Acad. Sci. U. S. A. *108*, E1392-1398.

Showkat, M., Beigh, M.A., and Andrabi, K.I. (2014). mTOR Signaling in Protein Translation Regulation: Implications in Cancer Genesis and Therapeutic Interventions. Mol. Biol. Int. *2014*, 1–14.

Shuda, M., Kondoh, N., Tanaka, K., Ryo, A., Wakatsuki, T., Hada, A., Goseki, N., Igari, T., Hatsuse, K., Aihara, T., *et al.* (2000). *Enhanced expression of translation factor mRNA in hepatocellular carcinoma*. Anticancer Res. *20*, 2489–2494.

Silvera, D., and Schneider, R.J. (2009). *Inflammatory breast cancer cells are constitutively adapted to hypoxia*. Cell Cycle Georget. Tex *8*, 3091.

Silvera, D., Formenti, S.C., and Schneider, R.J. (2010). *Translational control in cancer*. Nat Rev Cancer *10*, 254–66.

Simon, A.E. (2015). *3'UTR of carmoviruses*. Virus Res. *206*, 27–36.

Simon, A.E., and Miller, W.A. (2013). *3′ Cap-Independent Translation Enhancers of Plant Viruses*. Annu. Rev. Microbiol. *67*, 21–42.

Sinvani, H., Haimov, O., Svitkin, Y., Sonenberg, N., Tamarkin-Ben-Harush, A., Viollet, B., and Dikstein, R. (2015). *Translational Tolerance of Mitochondrial Genes to Metabolic Energy Stress Involves TISU and eIF1-eIF4GI Cooperation in Start Codon Selection*. Cell Metab. *21*, 479–492.

Skabkin, M.A., Skabkina, O.V., Dhote, V., Komar, A.A., Hellen, C.U.T., and Pestova, T.V. (2010). *Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling*. Genes Dev. *24*, 1787–1801.

Soengas, M.S., Capodieci, P., Polsky, D., Mora, J., Esteller, M., Opitz-Araya, X., McCombie, R., Herman, J.G., Gerald, W.L., Lazebnik, Y.A., *et al.* (2001). *Inactivation of the apoptosis effector Apaf-1 in malignant melanoma*. Nature *409*, 207–211.

Sokabe, M., and Fraser, C.S. (2014). *Human Eukaryotic Initiation Factor 2 (eIF2)-GTP-Met-tRNAi Ternary Complex and eIF3 Stabilise the 43 S Preinitiation Complex*. J. Biol. Chem. *289*, 31827–31836.

Sokabe, M., Fraser, C.S., and Hershey, J.W.B. (2012). *The human translation initiation multi-factor complex promotes methionyl-tRNAi binding to the 40S ribosomal subunit*. Nucleic Acids Res. *40*, 905–913.

Sonenberg, N., and Hinnebusch, A.G. (2007). *New Modes of Translational Control in Development, Behavior, and Disease*. Mol. Cell *28*, 721–729.

Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A., and Barford, D. (2000). *The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis*. Cell *100*, 311–321.

Spriggs, K., Stoneley, M., Bushell, M., and Willis, A. (2008). *Re-programming of translation following cell stress allows IRES-mediated translation to predominate*. Biol. Cell *100*, 27–38.

Spriggs, K.A., Bushell, M., Mitchell, S.A., and Willis, A.E. (2005). *Internal ribosome entry segment-mediated translation during apoptosis: the role of IRES-trans-acting factors*. Cell Death Differ. *12*, 585–591.

Spriggs, K.A., Cobbold, L.C., Jopling, C.L., Cooper, R.E., Wilson, L.A., Stoneley, M., Coldwell, M.J., Poncet, D., Shen, Y.-C., Morley, S.J., *et al.* (2009a). *Canonical Initiation Factor Requirements of the Myc Family of Internal Ribosome Entry Segments*. Mol. Cell. Biol. *29*, 1565–1574.

Spriggs, K.A., Cobbold, L.C., Ridley, S.H., Coldwell, M., Bottley, A., Bushell, M., Willis, A.E., and Siddle, K. (2009b). *The human insulin receptor mRNA contains a functional internal ribosome entry segment*. Nucleic Acids Res. *37*, 5881–5893.

Stein, I., Itin, A., Einat, P., Skaliter, R., Grossman, Z., and Keshet, E. (1998). *Translation of vascular endothelial growth factor mRNA by internal ribosome entry: implications for translation under hypoxia*. Mol. Cell. Biol. *18*, 3112.

Stoneley, M., Chappell, S.A., Jopling, C.L., Dickens, M., MacFarlane, M., and Willis, A.E. (2000a). *c-Myc protein synthesis is initiated from the internal ribosome entry segment during apoptosis*. Mol. Cell. Biol. *20*, 1162–1169.

Stoneley, M., Subkhankulova, T., Le Quesne, J.P.C., Coldwell, M.J., Jopling, C.L., Belsham, G.J., and Willis, A.E. (2000b). *Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment*. Nucleic Acids Res. *28*, 687–694.

Stoneley, M., Paulin, F. E., Quesne, J. P. L., Chappell, S. A., & Willis, A. E. (1998). C-Myc 5'untranslated region contains an internal ribosome entry segment. *Oncogene*, *16*(3), 423-428.

Subkhankulova, T., Mitchell, S.A., and Willis, A.E. (2001). *Internal ribosome entry segment-mediated initiation of c-Myc protein synthesis following genotoxic stress*. Biochem. J. *359*, 183.

Sun, C., Todorovic, A., Querol-Audí, J., Bai, Y., Villa, N., Snyder, M., Ashchyan, J., Lewis, C.S., Hartland, A., Gradia, S., *et al.* (2011a). *Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3).* Proc. Natl. Acad. Sci. *108*, 20473–20478.

Sun, D., Liu, W.-J., Guo, K., Rusche, J.J., Ebbinghaus, S., Gokhale, V., and Hurley, L.H. (2008). *The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents.* Mol. Cancer Ther. *7*, 880–889.

Sun, J., Conn, C.S., Han, Y., Yeung, V., and Qian, S.-B. (2011b). *PI3K-mTORC1 Attenuates Stress Response by Inhibiting Cap-independent Hsp70 Translation*. J. Biol. Chem. *286*, 6791–6800.

Takagi, M., Absalon, M.J., McLure, K.G., and Kastan, M.B. (2005). *Regulation of p53 translation and induction after DNA damage by ribosomal protein L26 and nucleolin*. Cell *123*, 49–63.

Takeda, M. (2004). *A Unique Role of an Amino Terminal 16-Residue Region of Long-Type GATA-6*. J. Biochem. (Tokyo) *135*, 639–650.

Tange, T.Ø., Nott, A., and Moore, M.J. (2004). *The ever-increasing complexities of the exon junction complex*. Curr. Opin. Cell Biol. *16*, 279–284.

Tarbé, N.G., Rio, M.C., Hummel, S., Weidle, U.H., and Zöller, M. (2005). *Overexpression of the small transmembrane and glycosylated protein SMAGP supports metastasis formation of a rat pancreatic adenocarcinoma line*. Int. J. Cancer *117*, 913–922.

Tawfik, H.M., El-Maqsoud, N.M.R.A., Hak, B.H.A.A., and El-Sherbiny, Y.M. (2011). *Head and neck squamous cell carcinoma: mismatch repair immunohistochemistry and promoter hypermethylation of hMLH1 gene*. Am. J. Otolaryngol. *32*, 528–536.

Terenin, I.M., Andreev, D.E., Dmitriev, S.E., and Shatsky, I.N. (2013). *A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent*. Nucleic Acids Res. *41*, 1807–1816.

Thakor, N., and Holcik, M. (2012). *IRES-mediated translation of cellular messenger RNA operates in eIF2 - independent manner during stress*. Nucleic Acids Res. *40*, 541–552.

Tomonaga, T., Matsushita, K., Yamaguchi, S., Oohashi, T., Shimada, H., Ochiai, T., Yoda, K., and Nomura, F. (2003). *Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer*. Cancer Res. *63*, 3511.

Topisirovic, I., and Sonenberg, N. (2011). *mRNA Translation and Energy Metabolism in Cancer: The Role of the MAPK and mTORC1 Pathways*. Cold Spring Harb. Symp. Quant. Biol. *76*, 355–367.

Tsai, B.P., Jimenez, J., Lim, S., Fitzgerald, K.D., Zhang, M., Chuah, C.T.H., Axelrod, H., Wilson, L., Ong, S.T., Semler, B.L., *et al.* (2014). *A novel Bcr-Abl-mTOR-eIF4A axis regulates IRES-mediated translation of LEF-1.* Open Biol. *4*, 140180–140180.

Turnpenny, P. D., & Ellard, S. (2011). *Emery's elements of medical genetics*. Elsevier Health Sciences.

Ungureanu, N.H., Cloutier, M., Lewis, S.M., de Silva, N., Blais, J.D., Bell, J.C., and Holcik, M. (2006). *Internal ribosome entry site-mediated translation of Apaf-1, but not XIAP, is regulated during UV-induced cell death*. J. Biol. Chem. *281*, 15155–15163.

Vagner, S., Gensac, M.C., Maret, A., Bayard, F., Amalric, F., Prats, H., and Prats, A.C. (1995). *Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes*. Mol. Cell. Biol. *15*, 35–44.

Vaklavas, C., Meng, Z., Choi, H., Grizzle, W.E., Zinn, K.R., and Blume, S.W. (2015). *Small molecule inhibitors of IRES-mediated translation*. Cancer Biol. Ther. *16*, 1471–1485.

Valásek, L., Nielsen, K.H., and Hinnebusch, A.G. (2002). *Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation in vivo*. EMBO J. *21*, 5886–5898.

Valeri, N., Gasparini, P., Fabbri, M., Braconi, C., Veronese, A., Lovat, F., Adair, B., Vannini, I., Fanini, F., Bottoni, A., *et al.* (2010). *Modulation of mismatch repair and genomic stability by miR-155.* Proc. Natl. Acad. Sci. U. S. A. *107*, 6982–6987.

Van Eden, M.E. (2004). *Demonstrating internal ribosome entry sites in eukaryotic mRNA using stringent RNA test procedures*. RNA *10*, 720–730.

Van Stry, M., Oguin, T.H., Cheloufi, S., Vogel, P., Watanabe, M., Pillai, M.R., Dash, P., Thomas, P.G., Hannon, G.J., and Bix, M. (2012). *Enhanced Susceptibility of Ago1/3 Double-Null Mice to Influenza A Virus Infection*. J. Virol. *86*, 4151–4157.

Vilar, E., and Gruber, S.B. (2010). *Microsatellite instability in colorectal cancer–the stable evidence*. Nat. Rev. Clin. Oncol. *7*, 153–162.

Villa, N., Do, A., Hershey, J.W.B., and Fraser, C.S. (2013). *Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome*. J. Biol. Chem. *288*, 32932–32940.

Villaflores, O.B., Chen, Y.-J., Chen, C.-P., Yeh, J.-M., and Wu, T.-Y. (2012). *Effects of curcumin and demethoxycurcumin on amyloid-β precursor and tau proteins through the internal ribosome entry sites: A potential therapeutic for Alzheimer's disease*. Taiwan. J. Obstet. Gynecol. *51*, 554–564.

van Vliet, C., Thomas, E.C., Merino-Trigo, A., Teasdale, R.D., and Gleeson, P.A. (2003). *Intracellular sorting and transport of proteins*. Prog. Biophys. Mol. Biol. *83*, 1–45.

Vumbaca, F., Phoenix, K.N., Rodriguez-Pinto, D., Han, D.K., and Claffey, K.P. (2008). *Double-Stranded RNA-Binding Protein Regulates Vascular Endothelial Growth Factor mRNA Stability, Translation, and Breast Cancer Angiogenesis*. Mol. Cell. Biol. *28*, 772–783.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). *The spliceosome: design principles of a dynamic RNP machine*. Cell *136*, 701–718.

Waki, T., Tamura, G., Tsuchiya, T., Sato, K., Nishizuka, S., and Motoyama, T. (2002). *Promoter methylation status of E-cadherin, hMLH1, and p16 genes in nonneoplastic gastric epithelia*. Am. J. Pathol. *161*, 399–403.

Wang, X.-L. (2012). *Detection of eukaryotic translation initiation factor 4E and its clinical significance in hepatocellular carcinoma*. World J. Gastroenterol. *18*, 2540.

Wang, S., Rosenwald, I.B., Hutzler, M.J., Pihan, G.A., Savas, L., Chen, J.-J., and Woda, B.A. (1999). *Expression of the eukaryotic translation initiation factors 4E and 2α in non-Hodgkin's lymphomas*. Am. J. Pathol. *155*, 247–255.

Wang, X., Zhao, Y., Xiao, Z., Chen, B., Wei, Z., Wang, B., Zhang, J., Han, J., Gao, Y., Li, L., *et al.* (2009). *Alternative Translation of OCT4 by an Internal Ribosome Entry Site and its Novel Function in Stress Response*. Stem Cells *27*, 1265–1275.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., *et al.* (2013). *N^6-methyladenosine-dependent regulation of messenger RNA stability*. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). *N^6-methyladenosine Modulates Messenger RNA Translation Efficiency.* Cell *161*, 1388–1399.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014). *N^6-methyladenosine modification destabilises developmental regulators in embryonic stem cells*. Nat. Cell Biol. *16*, 191–198.

Wani, M., Afroze, D., Makhdoomi, M., Hamid, I., Wani, B., Bhat, G., Wani, R., and Wani, K. (2012). *Promoter methylation status of DNA repair gene (hMLH1) in gastric carcinoma patients of the Kashmir valley*. Asian Pac. J. Cancer Prev. APJCP *13*, 4177–4181.

Weingarten-Gabbay, S., Khan, D., Liberman, N., Yoffe, Y., Bialik, S., Das, S., Oren, M., and Kimchi, A. (2014). *The translation initiation factor DAP5 promotes IRES-driven translation of p53 mRNA*. Oncogene *33*, 611–618.

Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A.A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., and Segal, E. (2016). *Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes*. Science *351*.

Weipoltshammer, K., and Schöfer, C. (2016). *Morphology of nuclear transcription*. Histochem. Cell Biol. *145*, 343–358.

Welch, E.M., Barton, E.R., Zhuo, J., Tomizawa, Y., Friesen, W.J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C.R., Hwang, S., *et al.* (2007). *PTC124 targets genetic disorders caused by nonsense mutations*. Nature *447*, 87–91.

Welsh, S.J., Dale, A.G., Lombardo, C.M., Valentine, H., de la Fuente, M., Schatzlein, A., and Neidle, S. (2013). *Inhibition of the hypoxia-inducible factor pathway by a G-quadruplex binding small molecule*. Sci. Rep. *3*, 2799.

Will, C.L., and Luhrmann, R. (2011). *Spliceosome Structure and Function*. Cold Spring Harb. Perspect. Biol. *3*, a003707–a003707.

Wu, H., Zeng, H., Lam, R., Tempel, W., Kerr, I.D., and Min, J. (2015). *Structure of the human MLH1 N-terminus: implications for predisposition to Lynch syndrome*. Acta Crystallogr. Sect. F Struct. Biol. Commun. *71*, 981–985.

Xiao, Z.-S., Simpson, L.G., and Quarles, L.D. (2003). *IRES-dependent translational control of Cbfa1/Runx2 expressio*n. J. Cell. Biochem. *88*, 493–505.

Xing, L., Liang, C., and Kleiman, L. (2011). *Coordinate Roles of Gag and RNA Helicase A in Promoting the Annealing of Formula to HIV-1 RNA*. J. Virol. *85*, 1847–1860.

Yamakawa, N., Okuyama, K., Ogata, J., Kanai, A., Helwak, A., Takamatsu, M., Imadome, K. -i., Takakura, K., Chanda, B., Kurosaki, N., *et al.* (2014). *Novel functional small RNA are selectively loaded onto mammalian Ago1*. Nucleic Acids Res. *42*, 5289–5301.

Yamamoto, Y., Singh, C.R., Marintchev, A., Hall, N.S., Hannig, E.M., Wagner, G., and Asano, K. (2005). *The eukaryotic initiation factor (eIF) 5 HEAT domain mediates multifactor assembly and scanning with distinct interfaces to eIF1, eIF2, eIF3, and eIF4G*. Proc. Natl. Acad. Sci. U. S. A. *102*, 16164–16169.

Yaman, I., Fernandez, J., Liu, H., Caprara, M., Komar, A.A., Koromilas, A.E., Zhou, L., Snider, M.D., Scheuner, D., and Kaufman, R.J. (2003). *The zipper model of translational control: a small upstream ORF is the switch that controls structural remodeling of an mRNA leader*. Cell *113*, 519–531.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). *Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay*. Genes Dev. *15*, 2215–2228.

Yang, J.-S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010). *Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis*. Proc. Natl. Acad. Sci. *107*, 15163–15168.

Yang, L., Chen, D., Duan, R., Xia, L., Wang, J., Qurashi, A., Jin, P., and Chen, D. (2007). *Argonaute 1 regulates the fate of germline stem cells in Drosophila*. Dev. Camb. Engl. *134*, 4265–4272.

Yang, X., Hao, Y., Ferenczy, A., Tang, S.C., and Pater, A. (1999). *Overexpression of anti-apoptotic gene BAG-1 in human cervical cancer*. Exp. Cell Res. *247*, 200–207.

Yeh, S.H., Yang, W.B., Gean, P.W., Hsu, C.Y., Tseng, J.T., Su, T.P., Chang, W.C., and Hung, J.J. (2011). *Translational and transcriptional control of Sp1 against ischaemia through a hydrogen peroxide-activated internal ribosomal entry site pathway*. Nucleic Acids Res. *39*, 5412–5423.

Yin, X., Kim, R.H., Sun, G., Miller, J.K., and Li, B.D. (2014). *Overexpression of Eukaryotic Initiation Factor 4E Is Correlated with Increased Risk for Systemic Dissemination in Node-Positive Breast Cancer Patients*. J. Am. Coll. Surg. *218*, 663–671.

Yin, Y., Stephen, C.W., Luciani, M.G., and Fåhraeus, R. (2002). *p53 Stability and activity is regulated by Mdm2-mediated induction of alternative p53 translation products*. Nat. Cell Biol. *4*, 462–467.

Young, R.M., Wang, S.-J., Gordan, J.D., Ji, X., Liebhaber, S.A., and Simon, M.C. (2008). *Hypoxia-mediated Selective mRNA Translation by an Internal Ribosome Entry Site-independent Mechanism*. J. Biol. Chem. *283*, 16309–16319.

Youssef, S.S., Fahmy, A.M., Omran, M.H., Mohamed, A.S., El Desouki, M.A., and El-Awady, M.K. (2014). *In vitro Inhibition of Hepatitis C Virus by Antisense Oligonucleotides in PBMC Compared to Hepatoma Cells*. BioMed Res. Int. *2014*, 1–7.

Zhang, B.B., Zhou, G., and Li, C. (2009). *AMPK: an emerging drug target for diabetes and the metabolic syndrome*. Cell Metab. *9*, 407–416.

Zhang, J., Cho, S.-J., Shu, L., Yan, W., Guerrero, T., Kent, M., Skorupski, K., Chen, H., and Chen, X. (2011). *Translational repression of p53 by RNPC1, a p53 target overexpressed in lymphomas*. Genes Dev. *25*, 1528–1543.

Zhang, J., Dinh, T.N., Kappeler, K., Tsaprailis, G., and Chen, Q.M. (2012). *La Autoantigen Mediates Oxidant Induced De Novo Nrf2 Protein Translation*. Mol. Cell. Proteomics *11*, M111.015032-M111.015032.

Zhong, X., Arita, M., Yamada, K., Sugiyama, H., Tan, K., Kanazawa, S., Koike, J., Teramoto, T., and Hemmi, H. (2007). *A Single Nucleotide Substitution (−107C → G) in the hMLH1 Promoter Found in Colorectal Cancer Population Reduces Transcriptional Activity*. Biochem. Genet. *45*, 671–681.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). *Dynamic m^6A mRNA methylation directs translational control of heat shock response*. Nature *526*, 591–594.

Zhou, J., Rode, K.A., and Qian, S.-B. (2016). *m^6A: A novel hallmark of translation.* Cell Cycle *15*, 309–310.

Zitvogel, L., Tahara, H., Cai, Q., Storkus, W.J., Muller, G., Wolf, S.F., Gately, M., Robbins, P.D., and Lotze, M.T. (1994). *Construction and characterization of retroviral vectors expressing biologically active human interleukin-12*. Hum. Gene Ther. *5*, 1493–1506.

Zou, J., Luo, H., Zeng, Q., Dong, Z., Wu, D., and Liu, L. (2011). *Protein kinase CK2alpha is overexpressed in colorectal cancer and modulates cell proliferation and invasion via regulating EMT-related genes*. J. Transl. Med. *9*, 97.

Zuo, C., Zhang, H., Spencer, H.J., Vural, E., Suen, J.Y., Schichman, S.A., Smoller, B.R., Kokoska, M.S., and Fan, C.-Y. (2009). *Increased microsatellite instability and epigenetic inactivation of the hMLH1 gene in head and neck squamous cell carcinoma*. Otolaryngol.--Head Neck Surg. Off. J. Am. Acad. Otolaryngol.-Head Neck Surg. *141*, 484–490.