Aus der Klinik für Kinder- und Jugendmedizin
(Prof. Dr. med. J. Gärtner)
der Medizinischen Fakultät der Universität Göttingen

LDHBx and MDH1x are controlled by physiological translational readthrough in Homo sapiens

INAUGURAL - DISSERTATION

zur Erlangung des Doktorgrades
der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von
Fabian Schüren
aus
Bremen

Göttingen 2015

Dekan:	Prof. Dr. rer. nat. H. K. Kroemer
I. Berichterstatter/in:	PD Dr. rer. nat. S. Thoms
II. Berichterstatter/in:	Prof. Dr. rer. nat. B. Schwappach-Pignataro
III. Berichterstatter/in:	Prof. Dr. mult. T. Meyer
Tag der mündlichen Prüfung:	07.04.2016

Hiermit erkläre ich, die Dissertation mit dem Titel „LDHBx and MDH1x are controlled by physiological translational readthrough in *Homo sapiens*“ eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den ....

(Unterschrift)

Contents

List of Figures

List of Tables

Nomenclature

Nucleobases

Character	Nucleobases	Group of Nucleobases
A	adenine	purine
G	guanine	purine
U	uracil	pyrimidine
T	thymine	pyrimidine
C	cytosine	pyrimidine

Abbreviations

Abbreviation	Meaning
A-site	ribosomal site, binds aminoacyl-tRNA
Aa	amino acid(s)
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CoA	coenzyme A
CSF	codon substitution frequencies
DAPI	4′,6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DRVL	dual reporter using Venus and luciferase
E-site	ribosomal site, binds free tRNA and releases it
ER	endoplasmatic reticulum
eRF1	eukaryotic release factor 1
FAD/$FADH_2$	flavin adenine dinucleotide
fs	feature selection
FTR	functional translational readthrough
GFP	green fluorescent protein
LDH	lactate dehydrogenase
Leu	leucine, hydrophobic amino acid
LIN	basic model
LINfs3	model, refined using feature selection
LINfs5	model, refined using feature selection
LINiter	iterative model
Lys	lysine, polar amino acid

Abbreviation	Meaning
MCS	multi-cloning site
MCT	monocarboxylate transporter
MDH	malate dehydrogenase
mRNA	messenger RNA
NAD^+/$NADH+H^+$	nicotinamide adenine dinucleotide
$NADP^+$/NADPH	nicotinamide adenine dinucleotide phosphate
NLS	nuclear localisation signal
ORF	open reading frame
P-site	ribosomal site, binds peptidyl-tRNA
PBD	peroxisome biogenesis disorder
PDE2	phosphodiesterase 2
PEX5/7	peroxisomal import receptor recognising PTS1/2
PTS1/2	peroxisomal targeting signal 1/2
RCDP	rhizomelic chondrodysplasia punctata
RF	release factor
RNA	ribonucleic acid
ROI	region of interest
RTP	readthrough propensity
SCC	stop codon context
SD	standard deviation
Ser	serine, polar amino acid
TMD	transmembrane domain
TR	translational readthrough
tRNA	transfer RNA
UTR	untranslated region
WB	Western blot
ZSD	Zellweger spectrum disease

Organisms

Viruses

Abbreviation	Virus
AKV	Aujeszkyvirus, a murine leukemia virus
BWYV	Beet western yellow virus
FeLV	Feline leukemia virus
Mo-MuLV	Moloney-murine leukemia virus

Abbreviation	Virus
Qβ	Bacteriophage Qβ
SINV	Sindbis virus
TMV	Tobacco mosaic virus

Prokaryotes & Fungi

Species
Bacillus subtilis
Escherichia coli
Salmonella typhimurium
Saccharomyces cerevisiae
Ustilago maydis
Yarrowia lipolytica

Metazoa

Species
Drosophila melanogaster
Caenorhabditis elegans
Mus musculus
Felis catus
Homo sapiens

1 Introduction

This introduction will provide the reader with sufficient information regarding the topic of functional translational readthrough (FTR) so as to allow the reader an understanding of the findings of this study. Translation and termination are explained, followed by examples of translational readthrough (TR) in lower organisms, which might suggest a first hint that the phenomenon appears in abundance in different organisms. Exhaustive systems biology approaches that show surprisingly little redundancy in their results allowed systematic analysis of TR in recent years. One of these systems biology approaches was designed and applied by our group. These methods led to the detection and verification of the first FTR genes found in *Homo sapiens*. These genes and their target organelle in the cell are elucidated. The likely functions of the FTR genes in their target organelle - the peroxisome - are discussed.

1.1 Translation and termination

Information, which is stored in our genome in form of DNA, is processed and copied to messenger RNA (mRNA). The mRNA translocates from the nucleus to the cell's cytosol where it is translated into enzymes or structural proteins. These function as the biological machinery, infrastructure or skeleton of the cell. Translation occurs at ribosomes that build peptide chains (Figure 1 on the following page), which in turn fold into e.g. functional enzymes. Ribosomes are assembled from two subunits for each translation and present with three slots for transfer RNAs (tRNAs), namely the E-, P-, and A-site. The process of translation can be divided into four steps, initiation at the mRNA's initiation site, elongation, translocation of mRNA through the ribosome and termination of translation at the termination site. It is the step of termination that is of particular interest for this study.
When one of the three stop codons (UAG, UAA, or UGA) enters the ribosomal A-site it is generally bound by a release factor (RF) instead of a tRNA carrying an amino acid. The eukaryotic release factor 1 (eRF1) recognises each of the three stop codons and bridges the distance from a stop codon to the peptidyl transferase center in the ribosome. A loop of eRF1 assisted by the peptidyl transferase utilises a water molecule to cleave the peptidyl-tRNA bond of the tRNA at the ribosomal P-site (Berg et al. 2011). Subsequently the new peptide is released from the ribosome and the latter dissembles into its subunits.

1.2 (Functional) translational readthrough in lower organisms

As nature is rarely absolute or binary, also termination of translation is not free of errors. Instead of the release factor a competing near-cognate tRNA carrying an amino acid might bind

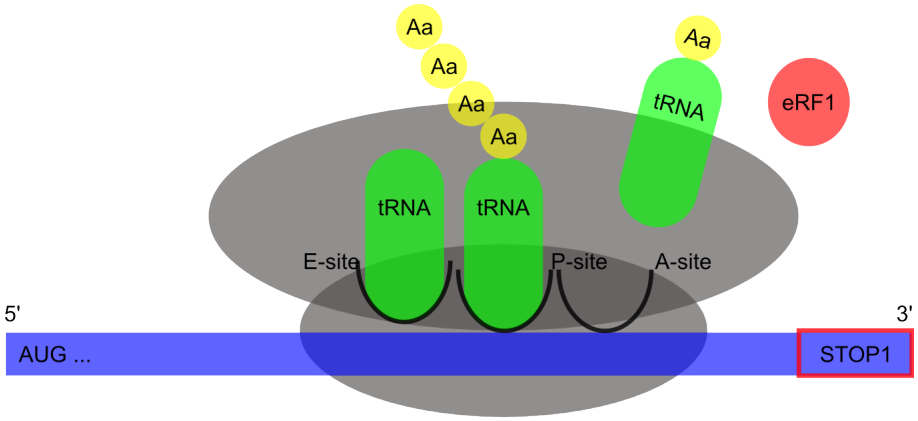


Figure 1: A ribosome that translates a mRNA starts at the 5′ end at the start codon (AUG) and stops generally at the 3′ end at one of the three stop codons (UAG, UAA, or UGA). tRNAs bind to codons with their respective, specific anticodons and carry amino acids with them. The amino acids form a peptide. Release factor eRF1 will compete with near-cognate tRNAs when the stop codon enters the ribosomal A-site in an eukaryotic cell.

the stop codon in the ribosomal A-site and translation continues until the second in-frame stop codon. Such errors occur with frequencies up to 0.01 and 0.1% (Namy et al. 2001; Harrell et al. 2002). In some cases termination efficiency is reduced by one or two orders of magnitude, which consequently results in stop suppression between 1 and 10 %. Stop suppression leads to insertion of amino acids for stop codons (Gesteland and Atkins 1996) and results in ribosomes that read through the stop codon into the downstream region. Thus they continue translation and extend the protein C-terminally beyond its normal length. Elevated levels of ribosomal readthrough above rates of 0.1% are defined here as translational readthrough (TR) (Figure 2). The extend to which TR occurs varies between the different stop codons in most

Figure 2: Ribosomes that translate mRNAs generally terminate translation at the first stop codon they encounter and dislocate from the mRNA. In rare cases the error rate that leads to stop suppression is elevated and ribosomes exhibit translational readthrough (TR). The first stop codon is decoded as a standard amino acid and the ribosome continues translation into the 3′ region downstream of the first stop codon. Translation is continued until a second in-frame stop codon enters the ribosomal A-site and eRF1 terminates translation.

organisms. TR is often influenced either by nucleotides in the vicinity of the stop codon, by distal mRNA structures downstream the stop codon, or both kinds of regulatory elements, for further information please refer to Paragraph 1.5 on page 9. TR and *global redefinition* of stop codons are distinct phenomena, as TR leads to insertion of a standard amino acid (Namy et al.

2001), while the term *redefinition* of a stop codon is used to describe globally regulated insertion of non-standard amino acids (e.g. selenocysteine), the latter is not known to be used by viruses (Firth and Brierley 2012). The important effect of TR is probably that it allows continued translation into the downstream region, while the specific amino acid that is added to the peptide at the termination site is assumed to be less crucial (Firth et al. 2011). Mutant tRNAs can be stop codon suppressors and elevate readthrough strongly, but TR occurs with 10^{-1} to 10^{1} also in normal physiology. Interestingly, TR can be induced and elevated by drugs such as antibiotics of the group of aminoglycosides, this is used and investigated as a therapeutic option for genetic disorders caused by premature stop codons (Bidou et al. 2012; Keeling et al. 2014). Aminoglycosides can also be utilised to confirm TR candidates and differentiate this mechanism from other recoding events.

Detection of TR ventured from viruses, namely bacteriophage *Qβ* and *tobacco mosaic virus* (TMV) (Weiner and Weber 1971; Hofstetter et al. 1974; Pelham 1978), to prokaryotes (reviewed in Engelberg-Kulka (1981)), to eukaryotes (Bonetti et al. 1995). TR was also detected by chance in mammals (Geller and Rich 1980; Yamaguchi et al. 2012).

Figure 3: Ribosomes that translate mRNAs generally terminate translation at the first stop codon they encounter and dislocate from the mRNA. In rare cases the error rate that leads to stop suppression is elevated and ribosomes exhibit translational readthrough (TR). The first stop codon is decoded as a standard amino acid and the ribosome continues translation into the 3′ region downstream of the first stop codon. This might give rise to new functional domains or signals in the extension, which were hidden in the untranslated region before. It allows expression of the parental proteins and a small amount of extended proteins with new functions. Translation is continued until a second in-frame stop codon enters the ribosomal A-site and eRF1 terminates translation.

Already with the detection of TR in viruses it became apparent that the phenomenon can fulfil a function as extensions with new domains are appended to the C-termini of proteins (Figure 3). Functional translational readthrough (FTR) is a term used for TR that gives rise to new isoforms, which possess new functions and thus differ not only structurally but also functionally from their parental proteins. FTR provides organisms with a curious ability to regulate protein expression as only a small percentage of the original protein is diverted from its task without disturbing its original function.

Viruses use TR and FTR among other recoding mechanisms (e.g. ribosomal hopping, or frameshifting) to expand the coding capacity of their small genome (Atkins et al. 1990; Firth and Brierley 2012). For example TR that ranges from 6 to 24 % was found in three plant viruses and one animal virus (Urban et al. 1996). Several examples of FTR in viruses are known.

The *Sindbis virus* (SINV) utilises TR to synthesise a viral RNA dependent RNA polymerase (Li and Rice 1993), TMV exploits TR to express a RNA replicase (Pelham 1978), and *feline leukemia virus* (FeLV) requires TR to express a protease (Yoshinaka et al. 1985b). Propagation of the virus depends on FTR for Qβ (Weiner and Weber 1971; Hofstetter et al. 1974), and *beet western yellow virus* (BWYV) (Bruyere et al. 1997). The importance and possible regulatory capacity of FTR can be seen in the example of *Moloney-murine leukemia virus* (Mo-MuLV), which expresses a *gag* gene and its TR product a gag-pol fusion protein (Yoshinaka et al. 1985a), the propagation of the virus depends on the parental protein, the fusion protein and possibly even on a fixed ratio between the two proteins (Felsenstein and Goff 1988).
Also in eukaryotes examples of TR and FTR exist. TR of UAG stop codons in yeast (*Saccharomyces cerevisiae*) was analysed, the standard amino acids tyrosine, lysine, and tryptophan were incorporated for the stop codons, which led to TR of 4 % (Fearon et al. 1994). Furthermore the genes *CST6* and *RCK2* are undergoing TR of 0.6 %, and 0.8 %, respectively (Namy et al. 2002). FTR was observed in other fungi species, e.g. enzymes of the glycolysis pathway (3-phosphoglycerate kinase), and the pentose phosphate pathway (D-ribulose-5-phosphate-3-epimerase) exhibit dual localisation to peroxisomes due to FTR in *Ustilago maydis* (Freitag et al. 2012; Stiebler et al. 2014). FTR of 2.2 % was proven for phosphodiesterase 2 (*PDE2*) in yeast. The process appends a 21 amino acid extension to the protein and thereby alters its localisation, and reduces its activity. Although the protein is non-essential its FTR leads to higher cAMP levels in yeast cells and influences the cell's stress response (Namy et al. 2002). In yeast systems biology approaches were applied, these led to the detection of two TR candidates (*IMP3*, and *BSC4*), and of underrepresented stop codon environments, which are thought to influence TR (Namy et al. 2003; Williams et al. 2004). Systems biology approaches allow systematic screens for TR candidates in a genome or transcriptome and are discussed in the following paragraph.

1.3 Systems biology approaches to translational readthrough

While the first examples of TR and FTR concern the analysis of individual genes and proteins and often have been found by serendipity, recently, systems biology approaches (Table 1 on the following page) have been employed to identify TR and FTR in metazoa. I will shortly introduce characteristics of metrics such as PhyloCSF, and ribosomal profiling. The method of our *in silico* regression model will be laid out in more detail. The systems biology approaches have been and will be applied in different species to investigate TR. I participated in the development of the *in silico* regression model and the results of this systems biology approach will be described in detail in Section 3 on page 32.
PhyloCSF is a comparative metric, which belongs to the group of phylogenetic approaches and utilises multiple sequence alignments to evaluate coding potential in gene sequences (Lin et al.

PhyloCSF	**Ribosome Profiling**
Relies on multiple species alignment to distinguish potential protein coding regions from non-coding regions by comparison of patterns of codon substitution frequencies in a genome-wide scan throughout related organisms, it identifies translational readthrough candidates by an evaluation of coding potential up- and downstream of annotated stop codons.	Ribosome profiling utilises ribosomal footprints in combination with deep sequencing. It identifies new coding regions, previously annotated as untranslated regions by their ribosomal footprint density. It can be applied to the whole transcriptome of an organism.
Mass spectrometry	***In silico* regression model**
Protein digest by trypsin or pepsin, followed by de novo sequencing allows identification of proteins or extensions. Extensions found in proteins domains that follow the annotated/canonical termination region hint readthrough or recoding events.	Calculates linear regression between stop codon context and experimental TR values to predict TR genes in a genome-wide scan. It was used to identify the high translational readthrough consensus in mammals. It can be combined with search algorithms that pinpoint functional protein domains.

Table 1: Recently employed systems biology approaches that identified numerous TR candidates and events in metazoa including *Homo sapiens*.

2011) to detect recoding events that lead to translation in presumingly untranslated regions (UTRs). As a systematic approach applicable to whole genomes and using several related species, the metric PhyloCSF combines phylogenetic information with a model that measures codon substitution frequencies (CSF). It utilises prior information of the genome that is to be analysed e.g. codon frequencies and branch lengths between species. The application of PhyloCSF requires a target sequence, one or more informant genomes that it can be aligned to, and known coding gene annotations in one of the data sets (Lin et al. 2011), which are required to train the model. The detection and verification of TR candidates found by PhyloCSF involves additional analyses and experiments, as PhyloCSF provides information about putative protein coding regions and thereby can only predict TR candidates. Sequencing of 12 related Drosophila genomes (Clark et al. 2007) allowed the application of comparative genomic analysis such as CSF on the genome of *Drosophila melanogaster*. This analysis revealed protein coding signatures in the extension between the designated stop codon and the next downstream, in-frame stop codon of 49 TR candidates (Lin et al. 2007). New transcriptional evidence (modENCODE) and PhyloCSF (Lin et al. 2011) were used to enlarge the number of TR candidates in *D. melanogaster* to 283 genes excluding non-TR events (Jungreis et al. 2011). Few candidates were proven to undergo TR using green fluorescent protein (GFP) transgenic flies (4 candidates), and mass spectrometry (9 candidates) including the known *syn* and *kel* genes (Jungreis et al. 2011).

Ribosome profiling recognises translating ribosomes in the 3′ UTRs and thus identifies recod-

ing events including TR (Ingolia et al. 2009, 2011). A ribosome protects the mRNA fragment (ribosome footprint), which is being translated by the ribosome, from digestion by nucleases. This in combination with the technique of deep sequencing is used by ribosome profiling. The approach allows localisation of translating ribosomes and thereby measurement of translation with sub-codon resolution (Ingolia et al. 2009). The libraries of mRNA fragments are created by digestion of mRNA, recovery of ribosome footprints, and their quantitative conversion into complementary DNA (cDNA) (Ingolia et al. 2011). The translation efficiency is defined as the ratio of ribosome footprints to mRNA abundance (Dunn et al. 2013; Ingolia et al. 2009). Ribosome profiling identified 307 TR events with readthrough rates above background levels (0.1 %) in *Drosophila melanogaster*, these events were not predicted by the PhyloCSF or CSF approaches (Dunn et al. 2013). Fortythree out of the 283 readthrough candidates found by application of PhyloCSF (Jungreis et al. 2011) were confirmed by ribosomal profiling (Dunn et al. 2013). The analysed lysates were prepared from *D. melanogaster* embryos and S2-cells, a cell line derived from *D. melanogaster* embryos. This restricted the results of this study (Dunn et al. 2013) to one developmental state. Given the premise that readthrough's regulation depends on the developmental state of the organism, confirmation of more than just 43 of the 283 phylogenetically predicted candidates might have been possible in further grown flies. Eight of 15 tested constructs showed TR in a dual reporter assay (Dunn et al. 2013), an experiment to prove TR biochemically via endogenous readthrough products was not carried out. In yeast, 30 examples of readthrough were found using ribosome profiling (Dunn et al. 2013). Fortytwo TR events that do not result from selenocysteine insertion were identified in *Homo sapiens* (Dunn et al. 2013), their functional relevance was not tested (Loughran et al. 2014).

***In silico* regression model (Schueren et al. 2014):** A genome wide *in silico* analysis of TR scanned a database of 200,000 human transcripts (Ensembl) for stop codon contexts (SCCs). The SCCs comprise 15 nucleotides (positions −6 to +9) surrounding the stop codon (positions +1 to +3). Our *in silico* regression model computed a readthrough propensity (RTP) score for 42,000 unique human transcripts. Therefore a binary vector represents the SCCs using 51 dimensions (12 positions × 4 nucleotides +3 stop codons), the binary vector was combined with experimentally assessed readthrough values. The *in silico* regression model assigned linear regression coefficients for all possible nucleotides in each position of the SCC, see Figure 4 on the following page for a rough sketch of the model. The design and numerical implementation of the model was done by Dr. T. Lingner, I participated in the development of one subunit, the RTP scanner. The model consists of two parts, a RTP scanner and a PTS1 scanner, which searches for putative peroxisomal targeting signal 1 (PTS1) sequences. Both scanner produce their own score, one RTP score and one PTS1 posterior probability. The RTP for a certain context was calculated as a sum over the regression coefficients of the context's positions. A first model (LIN) (Schueren et al. 2014) was trained with 66 sequences and their respective readthrough values (Floquet et al. 2012). A second model (LINiter) was trained with 15 ad-

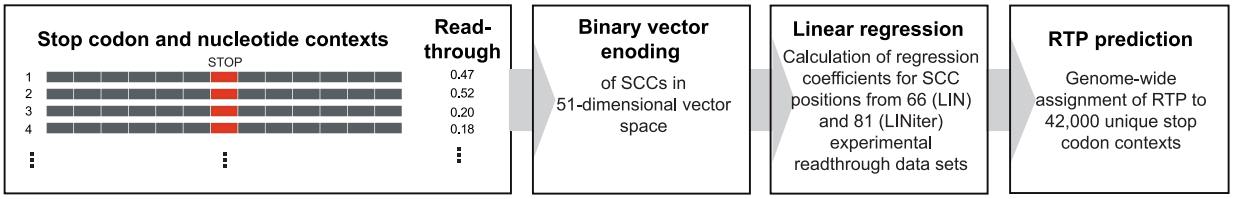


Figure 4: A schematic representation of the workflow of the RTP scanner. The stop codon contexts (SCCs) are represented as binary vectors, these and their experimental readthrough values are used to determine the coefficients employed by a linear regression model (LIN or LINiter). To predict the RTP score for a given SCC, the position-specific regression coefficients are added up. The sketch is taken from Schueren et al. (2014) and was originally designed by PD Dr. rer. nat. S. Thoms.

ditional experimentally tested sequences (Schueren et al. 2014). Feature selection (fs) led to models (LINfs5, and LINfs3) with yet higher correlation between readthrough and RTP and to a consensus for high RTP in human genes comprising only the stop codon and the positions −6, +4 to +7, or the stop codon and the positions +4 to +6 for LINfs5, and LINfs3, respectively. The PTS1 posterior probability predictor is an adaptation to human PTS1 based on a previously designed algorithm for plant PTS1 (Lingner et al. 2011). The importance of the PTS1 signal, that directs enzymes to the peroxisomal matrix is highlighted in Paragraph 1.6 on page 11.

The *in silico* regression model allows a prediction of readthrough propensity based on linear regression between an experimentally prepared training set of readthrough values and their respective sequences. Therefore the approach requires an annotated genome to search for SCCs and experimental evaluation of some of those; subsequently the model is able to predict RTP for the whole genome. The *in silico* regression model detected 38 readthrough candidates in the human genome that contain the high TR motif UGA CUA (stop codon underlined) and 19 candidates that comprehend the high TR consensus UGA CUA G (stop codon underlined), 5 of which were confirmed as TR events recently (Schueren et al. 2014; Loughran et al. 2014; Stiebler et al. 2014; Dunn et al. 2013). The confirmed TR events include the genes *LDHB* and *MDH1*, which are the protagonists of this study and thus play the main part in the result part (Section 3 on page 32), background information on the enzymes (LDH, MDH) are given in Paragraph 1.6 on page 11. As the influence of SCCs and more distal elements on TR is very interesting and important, these and also the revealed consensus for high TR in mammals including *Homo sapiens* will be explained and supported with additional information (Paragraph 1.5 on page 9).

1.4 (Functional) translational readthrough in higher organisms

Before systems biology approaches were applied, three genes of *Drosophila melanogaster* (*syn*, *kel*, and *hdc*) were proven as TR events (Klagges et al. 1996; Robinson and Cooley

1997; Steneberg and Samakovlis 2001) and two TR candidates (*Sxl*, and *oaf*) were identified (Samuels et al. 1991; Bergstrom et al. 1995). The expression rates of full length *kel* and open reading frame 1 (ORF1) *kel* vary with respect to the tissue and the fly's developmental state (Robinson and Cooley 1997).

The systems biology approach PhyloCSF predicted 283 TR candidates in *D. melanogaster* (Jungreis et al. 2011). Four of the 283 candidates (*Abd-B*, *cnc*, *Sp1*, and *z*) were confirmed as TR events using GFP transgenic flies. Interestingly not all TR events were found in the same developmental state of the organism (either in embryo, larvae, or both), kindly refer to Jungreis et al. (2011). Nine other candidates including known readthrough genes (*kel*, and *sync*) were confirmed to undergo TR by analysis of mass spectrometric data.

Ribosome profiling detected 350 TR events and thereby confirmed 43 of the phylogenetically predicted candidates in *D. melanogaster* (Dunn et al. 2013). From the 307 newly identified TR events only 14 of the extensions reached positive scores by application of PhyloCSF. This might indicate recent occurrence of most of these TR events in evolutionary terms. A comparison of the footprint density in different cell types led to the hypothesis that readthrough is differentially regulated between Drosophila cell types (Dunn et al. 2013). Eight out of 15 tested TR events predicted in *D. melanogaster* showed readthrough in a reporter assay using over-expression of constructs (Dunn et al. 2013).

FTR is defined as translational readthrough that leads to an isoform of the parental protein, which fulfils a different function than the latter one. This is assumed to be used in normal physiology of organisms. TR candidates with predicted hidden PTS1 were found in *D. melanogaster* (NADP-dependent isocitrate dehydrogenase), and *Caenorhabditis elegans* (inorganic pyrophosphatase). These candidates are particularly interesting, because the predicted targeting to peroxisomes was found to be realised in different ways (including alternative splicing, and PTS2) in different organisms (including *Ustilago maydis*, and Drosophila) (Stiebler et al. 2014). A number of potentially functional signals in the readthrough extensions, including a PTS1, nuclear localisation signals (NLS), eight transmembrane domains (TMDs), and one prenylation signal were found in TR candidates of *D. melanogaster*. Three NLS were shown to be functional using a GFP reporter, thus FTR was proven for these TR events in *D. melanogaster* (Dunn et al. 2013).

For a long time, the rabbit β-globin protein has been the only protein known to undergo stop codon suppression, without involvement of mutant tRNAs, in mammals (Geller and Rich 1980; Hatfield et al. 1988; Chittum et al. 1998). TR appends a 22 Aa extension to the β-globin protein, but it is probably not conserved in mammals (Jungreis et al. 2011). The application of PhyloCSF to a genome alignment of 29 mammalian species yielded the genes *SACM1L*, *OPRK1*, *OPRL1*, and *BRI3BP* as human readthrough candidates (Lindblad-Toh et al. 2011; Jungreis et al. 2011). Of these, two were experimentally confirmed (Loughran et al. 2014). Recently, the *MPZ* gene (myelin protein zero, P0) has been found to give rise to *L-MPZ* by readthrough (Yamaguchi et al. 2012). A functional significance of *L-MPZ* is likely

but not proven. However, the extension is known to contain antigenic sites for neuropathy-associated antibodies and mutations in *MPZ* can cause Charcot Marie–Tooth disease and Dejerine–Sottas disease (Hayasaka et al. 1993). A similar screen of 3′UTRs for conserved peptide sequences and an in-frame second stop codon in five mammalian transcriptomes revealed three readthrough events, *VEGFA*, *MTCH2*, and *AGO1* with 11, 13 and 24 % readthrough in a reporter assay respectively (Eswarappa et al. 2014). FTR of VEGFA will be discussed in the context of my results (Section 4 on page 62).
Ribosome profiling was applied to human foreskin fibroblasts and detected 42 TR events, of these 95 % exhibit TR > 1.2 % (Dunn et al. 2013). The four TR candidates (*SACM1L*, *OPRK1*, *OPRL1*, and *BRI3BP*) found by PhyloCSF (Jungreis et al. 2011) and three additional candidates (*ACP2*, *AQP4*, *MAPK10*) were tested in HEK-293T cells (Loughran et al. 2014). The candidates *AQP4*, *OPRL1*, *OPRK1*, and *MAPK10* exhibit readthrough rates between 7 and 31 % in mammalian translation systems, thus they were confirmed as TR events. Endogenous expression levels of these proteins, however, were too low to allow estimation of TR by ribosome profiling. Only very little TR was found for candidates with a predicted RNA secondary structure (*SACM1L*, *ACP2*) (Loughran et al. 2014). *AQP4* showed ∼ 3.5 % context driven readthrough in a dual reporter assay (Schueren et al. 2014) and was detected endogenously using Western blotting (Loughran et al. 2014). The influence that context and distal elements (e.g. RNA secondary structures) might exert on TR will be elucidated in Paragraph 1.5. FTR in *Homo sapiens* will be discussed in context of my results in the discussion (Section 4 on page 62).

1.5 Stop codon contexts and distal elements influence translational readthrough

As indicated throughout the preceding paragraphs, TR and FTR can be mediated and influenced by nucleotides in the immediate environment of a stop codon (SCC), or by nucleotide sequences that are located more distal of the stop codon in the 3'UTR and sometimes form elaborate secondary RNA structures. Translational readthrough is generally also influenced by the stop codon itself. The mechanisms are most probably distinct and can occur independently of each other. However, in possibly all organisms TR exists in both forms, and often both mechanisms are combined to regulate or ensure TR.
In viruses examples for TR mediated by context/distal elements alone and for TR influenced by a combination of context and distal elements exist. Distal elements are not generally required for TR in all viruses but might be necessary for stop suppression in some viruses (ten Dam et al. 1990; Honigman et al. 1991). A consensus STOP CAR YYA [R=A/G, Y=C/T] (stop codon underlined) for TR, which depends on context but is independent of stop codons, was found in TMV (Skuzeski et al. 1991). Of this consensus nucleotide positions +4, and +5

are most influential, but also position +6 exerts influence (Zerfass and Beier 1992; Urban et al. 1996). Distal elements in form of a pseudoknot structure in the 3′ vicinity of the *gag* gene's stop codon exist in MuLV, *aujeszkyvirus* (AKV), M7 Baboon endogenous virus, and FeLV (ten Dam et al. 1990; Wills et al. 1991). Beside the pseudoknot structure also a purine rich sequence (positions +4 to +8) immediately downstream the stop codon influences TR of the *gag* gene in MuLV (Feng et al. 1992; Honigman et al. 1991). In BWYV positions +8, and +9 of a SCC, but also two distal regions 3′ the stop codon influence TR (Bruyere et al. 1997; Brown et al. 1996). In SINV expression of the viral RNA polymerase nsP4 is regulated by context driven FTR. The SCC UGA CUA (stop codon underlined) regulates TR in plant and animal viruses (Beier and Grimm 2001), but introduction of distal stem loop structures increases TR 3 to 4 fold, the stems of these distal elements are more important than the loops (Firth et al. 2011).

In prokaryotes the SCC downstream and characteristics of the amino acids upstream the stop codon might influence TR. In general stop codon dependent TR in *E. coli* varies with SCCs as follows UAA/UGA C~A<G<U and UAG C<A~U<G (stop codon underlined) (Poole et al. 1995). Context effects were found for positions +4 to +10 in *E. coli*, with generally TR being most influenced by purines (Bossi 1983; Miller and Albertini 1983), or specifically an A residue at position +4 (Engelberg-Kulka 1981). Interestingly, highly expressed genes show a strong bias in position +4 against A or C and against usage of leaky stop codons UGA or UAG (Brown et al. 1990b). In *E. coli* tRNAs in the A-site of ribosomes might reduce the termination efficiency of stop codons residing in P-site via tRNA-tRNA interactions (Smith and Yarus 1989). High/low TR was also correlated with the amino acid and its charge (acidic/basic) at the E-site of the ribosome in *E. coli* (Mottagui-Tabar et al. 1994) and *Bacillus subtilis* and *Salmonella typhimurium* (Mottagui-Tabar and Isaksson 1998). It was shown that structural elements of release factor 2 (RF2) in *E. coli* are close enough to positions +4 to +6 (Poole et al. 1997, 1998) to allow site-directed cross-linking with the mRNA. Changes in positions +4 to +6 influence the efficiency of cross-linking between RF2 and position +1 (first base of stop codon), which might reflect changes of mRNA orientation with respect to RF2 (Poole et al. 1998).

Also eukaryotes exhibit TR that is influenced by SCC, distal elements, or both mechanisms. In general eukaryotes favour the consensus U (A/G) (G/A) (C/G) (stop codon underlined) for TR (Brown et al. 1990a). The SCC UAG CAA (stop codon underlined) in yeast allows high readthrough and resembles the TMV readthrough context (Kopczynski et al. 1992). The effect exerted by upstream contexts on TR depends on the downstream contexts (Williams et al. 2004), also synergistic effects of up- and downstream contexts were observed (Bonetti et al. 1995). The consensus UAG CAR NBA [R=A/G, N=any base, B=U/C/G] (stop codon underlined), that is very similar to the one found in TMV infected tobacco plant cells (Kopczynski et al. 1992), promotes high readthrough rates > 5 % in yeast (Namy et al. 2001), the result is supported by a genome-wide *in silico* study (Williams et al. 2004). Readthrough in yeast that

used the mentioned consensus is not promoted by a mRNA-tRNA interaction of positions +7 to +9 thereby the mechanism might involve the tRNA pairing with nucleotides at positions +4 to +6 or more direct effects due to the primary nucleotide sequence as no secondary structure is predictable for found consensus (Namy et al. 2001).
In *D. melanogaster* the tetranucleotide termination signal UGA C (stop codon underlined) is the least common signal among non-TR transcripts (3.1 %), while it is the most common signal used by genes that were classified as TR candidates (Jungreis et al. 2011). Termination efficiency is reduced by use of UGA>UAG>UAA and the base C>U>G>A at position +4 in the given order. This indicates context dependent TR, however, in *D. melanogaster* FTR that depends on distal structures and is stop codon independent was also observed for the *hdc* gene (Steneberg and Samakovlis 2001).
Like the other organisms also mammals utilise both ways to regulate or promote TR. A strong bias of position +4 of SCCs found in mammalian translation systems indicates the influence of this position on TR. A purine/pyridimine split was observed as pyrimidines (C/T) at position +4 promote translational readthrough and selenocysteyl-tRNA selection (McCaughan et al. 1995). High suppression rates were observed for UAG stop codons followed by a C residue at position +4 (Martin et al. 1993; Howard et al. 2000). Upstream the stop codon a correlation of a purine at position −1 and high readthrough was found (Cassan and Rousset 2001; Loughran et al. 2014).
In mammalian translation systems UGA, and UAG are most often decoded as tryptophan and glutamine, respectively. The stop codons differ in their termination efficiency (UAA>UAG>UGA) (Harrell et al. 2002), which is equivalent to the order found in *D. melanogaster*. In human a consensus for high TR was systematically derived by our *in silico* regression model (Schueren et al. 2014), also other studies found TR events and candidates with this consensus (Loughran et al. 2014; Stiebler et al. 2014). In a study of mammalian genomes that revealed four TR events (*AQP4*, *OPRL1*, *OPRK1*, and *MAPK10*) evidence was found, that supports a larger influential context encompassing positions +4 to +15 (Loughran et al. 2014). The high TR consensus in mammals supports TR of up to 3 % and constructs with larger contexts, which exhibit a combination of SCC and distal elements, displayed readthrough of up to 31 % (Loughran et al. 2014; Schueren et al. 2014). The findings so far support the hypothesis that TR can be regulated via the context in the stop codons vicinity (SCC) and distal elements in the 3′ region of the stop codon, however, each mechanism functions also on its own.

1.6 Peroxisomes and their functional significance

Our *in silico* regression model scanned for SCCs with high RTP values and additionally it evaluated the PTS1 posterior probability of the C-terminal end of the extensions. We used this scanner for putative functional PTS1 domains in the TR extensions to search for FTR in *Homo*

sapiens. As we concentrated our systematic search for FTR on possible peroxisomal targeting domains, an overview of the most important features of peroxisomes is given here.
Peroxisomes - formerly also known as microbodies - are cell organelles (de Duve 1969) that are found in every eukaryotic cell. The organelle builds a cell compartment with a matrix rich of proteins, which is enclosed by a single membrane (Wanders and Waterham 2006b). Though the organelle was regarded as relatively isolated, the evidence exists that peroxisomes are linked to other cell organelles, such as the endoplasmatic reticulum (ER) and mitochondria (Thoms et al. 2009). Peroxisomes play an essential role in many metabolic pathways (Wanders and Waterham 2006b) and react to environmental prompts by modification of biogenesis, morphology, number, and protein composition (Schrader et al. 2013). Monocarboxylate transporters (MCTs) like those found in cell membranes (Halestrap and Price 1999), and mitochondrial inner membranes (Brooks et al. 1999) were recently also detected in peroxisomal membranes (McClelland et al. 2003). MCTs support the metabolic activity of peroxisomes. It is not yet clear whether peroxisomes form by a budding process and originate from the ER or require the spatial closeness to the ER only for a transfer of peroxisomal membrane proteins, though evidence grows for an important role of the ER (Thoms et al. 2012). In one matter peroxisomes appear relatively independent of the ER as they are able to import enzymes from the cytosol into their matrix. These enzymes are directed to the peroxisomal matrix by peroxisomal targeting signal 1, and 2 (PTS1, and 2), which are found C- and N-terminally, respectively. The targeting signal PTS1 is recognised by import receptor PEX5, which directs peroxisomal matrix proteins into the organelle and subsequently returns to the cytosol (Liu et al. 2012; Smith and Aitchison 2013). The import receptor PEX7 recognises PTS2 (Marzioch et al. 1994). Over 90 % of the targeting signals found for peroxisomal matrix proteins are C-terminal and thus belong to the PTS1 group. The PTS1 does not lead to a binary decision, in contrast it directs the cytosolic protein partially into peroxisomes as the affinity of PEX5 to the PTS1 strongly depends on the amino acid sequence of the latter (Maynard et al. 2004). The typical PTS1, Ser-Lys-Leu (SKL) does not suffice for protein import into peroxisomes, up to nine amino acids further upstream of this tripeptide at the very C-terminus contribute, and the tripeptide itself exists in functional variants as well (Schueren et al. 2014; Brocard and Hartig 2006). Interestingly, peroxisomes have the rare ability to import folded enzymes and even oligomeric enzymes built from several subunits (Walton et al. 1995; McNew and Goodman 1996; Lanyon-Hogg et al. 2010).
Peroxisomes have a multitude of metabolic functions that extend beyond the textbook example of hydrogen peroxide metabolism. Fatty acids are metabolised in peroxisomes via β-oxidation (Lazarow and De Duve 1976), and human peroxisomes are also capable of α-oxidation. Currently it is assumed that 90 % of short/medium chain fatty acids are metabolised in mitochondria, the remaining 10 % in peroxisomes (Gladden 2004). Very long chain fatty acids are shortened in peroxisomes (Lodhi and Semenkovich 2014) and then shuttled to mitochondria, where they are processed in β-oxidation. The peroxisomal β-oxidation supplies

acetyl-CoA for the biosynthesis of functional lipids such as bile acids (Hayashi and Miwa 1989), and phospholipids (Hayashi and Takahata 1991). The organelle is also involved in glyoxylate detoxification (Wanders and Waterham 2006b) and peroxisomes might be related to leucine metabolism (McGroarty et al. 1974). Peroxisomal β-oxidation fuels anabolic processes in the peroxisomes, $FADH_2$ and NADH are produced during the oxidation and need to be re-oxidised to allow the β-oxidation to continue. In peroxisomes the $FADH_2$ transfers electrons directly to O_2 and consequently the pool of FAD used for β-oxidation is refilled (Gladden 2004). The mechanism for NADH re-oxidation in peroxisomes was long enigmatic. Then it was shown, that NAD^+/NADH and acetyl-CoA couldn't translocate over the peroxisomal membrane in *Saccharomyces cerevisiae* (van Roermund et al. 1995), and that peroxisomal β-oxidation could be stimulated by addition of oxaloacetate, and pyruvate (Osmundsen 1982; McClelland et al. 2003), while the addition of cytosolic lactate dehydrogenase (LDH) had no enhancing effect (Osmundsen 1982; Gladden 2004). As NAD^+/NADH can't cross the peroxisomal membrane, the findings indicate that monocarboxylates like lactate and pyruvate can translocate over the peroxisomal membrane. These monocarboxylates might be used to ensure redox-balance by use of dehydrogenases such as LDH and malate dehydrogenase (MDH). This hypothesis requires lactate to either translocate directly over the peroxisomal membrane or to use transporters as part of a shuttle system. Lactate shuttles exist in many forms as cell-to-cell shuttles, as intracellular lactate shuttles, as astrocyte-neurone lactate shuttle and possibly as peroxisomal shuttles (Gladden 2004). The stimulating effect of pyruvate on peroxisomal metabolism could be overcome by addition of a MCT blocker (McClelland et al. 2003). This supports the hypothesis that lactate/pyruvate use MCTs to translocate over the peroxisomal membrane (McClelland et al. 2003) instead of a direct shuttle system (Visser et al. 2007). This indicates that the re-oxidation of NADH takes place in peroxisomal matrix, as cytosolic LDH doesn't contribute to reaction inside the peroxisome and pyruvate can serve as a substrate for LDH inside the peroxisome. LDH would catalyse the reaction of pyruvate to lactate and thereby re-oxidise NADH.

The presence of LDH in peroxisomes and participation in the re-oxidation of NADH, produced by β-oxidation of palmitoyl-CoA, was confirmed (Baumgart et al. 1996). LDH was found to be to 0.33 to 1.2 % in peroxisomes (McGroarty et al. 1974; Volkl and Fahimi 1985; Baumgart et al. 1996). LDHA and also MDH1 were detected in peroxisomes by a mass spectrometry based proteomics survey (Gronemeyer et al. 2013). The way how LDH or MDH enter the peroxisome and their specific function in the peroxisome remained unclear. The redox-balance of the peroxisome might depend on the presence of LDH in peroxisomes, which highlights the enzymes importance in this particular organelle.

Lactate dehydrogenase (LDH) catalyses the conversion between lactate and pyruvate in the glycolysis pathway in an ordered sequential reaction (Figure 5 on the following page). A ternary complex is formed out of the enzyme and its substrates and products. LDH first binds the co-factor NADH and subsequently pyruvate, it releases the product lactate prior to NAD^+

(Berg et al. 2011). Beside lactate and pyruvate also glyoxylate is a substrate of LDH (Salido et al. 2012).

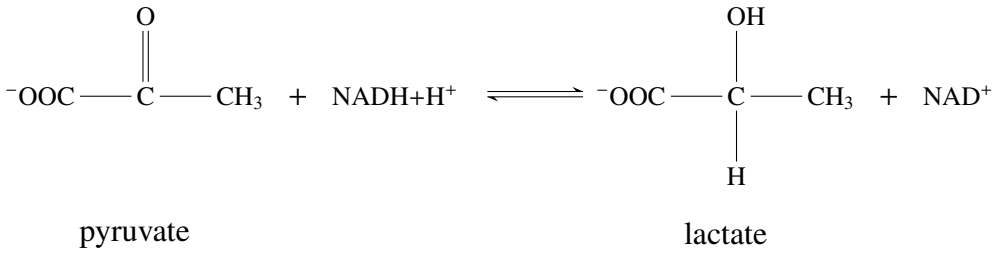


Figure 5: Reaction catalysed by lactate dehydrogenase (LDH)

LDH, the textbook example of the concept of isozymes, was resolved electrophoretically into five isozymes, which present with a molecular weight of 135 kDa and different net electrical charges (Markert 1963). The isozymes LDH 1-5 consist of four subunits (35 kDa) recruited from two different subunits M (muscle) and H (heart). These subunits M and H are encoded in the genes *LDHA* (OMIM® 2014a), located on chromosome 11p15.4 and *LDHB* (OMIM® 2014b), located on chromosome 12p12.2-p12.1, respectively. The different tetrameric isozymes are shown with their composition of subunits (Figure 6). As can be

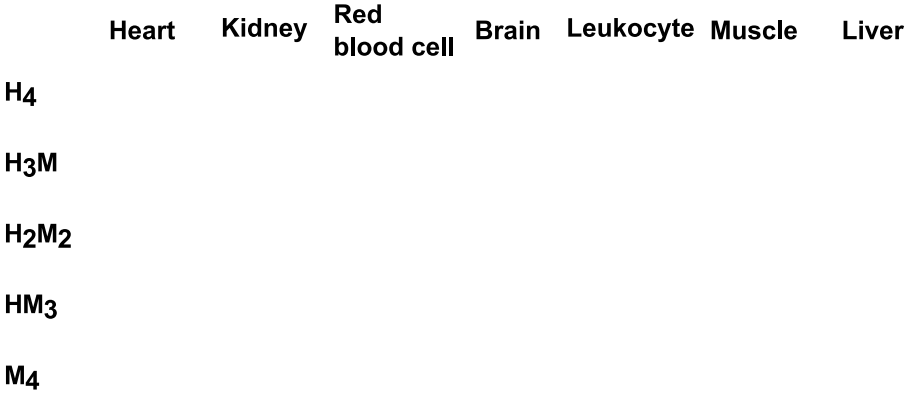


Figure 6: The LDH isozymes and their expression pattern in adult rat tissues, after (Urich 2013). The isozymes are LDH-1 (H_4), LDH-2 (H_3M_1), LDH-3 (H_2M_2), LDH-4 (H_1M_3), and LDH-5 (M_4).

expected by the composition of isozymes from their subunits, LDH 1-5 generate from an equal mixture of the two subunits in binomial distribution, in a proportion of 1 : 4 : 6 : 4 : 1 (Markert 1963). The homotetramer LDH-1 has the highest affinity to its substrates and works best in aerobic environments while LDH-5 has the lowest substrate affinity of all tetramers and works best in an anaerobic environment. LDH-1 unlike LDH-5 is allosterically inhibited by high levels of pyruvate (Berg et al. 2011). LDH is mostly found in a cell's cytosol and contributes to the anaerobic glycolysis with the shown reaction (Figure 5). This reaction allows a continued anaerobic glycolysis as it refills the pool of cytosolic NAD^+. The reverse reaction allows a refill of the pyruvate pool used for gluconeogenesis in e.g. liver cells.

Mammalian cells express two isoforms of malate dehydrogenase (MDH), which are the cytoplasmic MDH1, and the mitochondrial MDH2. MDH exists as a homodimeric enzyme

and is an ancestor of LDH in evolutionary terms. These enzymes participate in citrate cycle, gluconeogenesis (MDH2) and urea cycle (MDH1) and both isoforms contribute to the malate-aspartate-shuttle mainly in heart muscle and liver cells (Dettmer et al. 2005). Malate dehydrogenase catalyses the oxidation of malate (+NAD^+) to oxaloacetat acid (+$NADH+H^+$) (Figure 7).

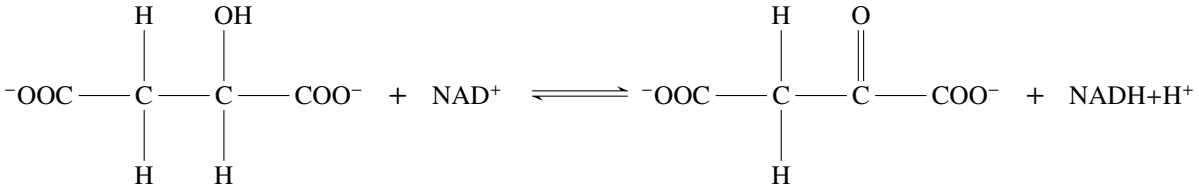


Figure 7: Malate dehydrogenase catalyses the interconversion of malate and oxaloacetic acid.

In mitochondria the reaction is driven by the use of oxaloacetate and NADH by citrate synthase, and the electrontransport chain, respectively (Berg et al. 2011).
In the yeast *Yarrowia lipolytica* the cytosolic and peroxisomal isoforms of MDH are derived from the same gene by alternative splicing (Kabran et al. 2012). In *Saccharomyces cerevisiae* the isoform MDH3 was found in peroxisomes (McAlister-Henn et al. 1995). The peroxisomal MDH is involved in the glyoxylate pathway, a pathway that takes place in cytosol and peroxisomal matrix (Kabran et al. 2012). Recently MDH1 was found in peroxisomes of human liver cells by a proteomics survey using mass spectrometry (Gronemeyer et al. 2013). It seems MDH is diverted to peroxisomes throughout different organisms. Until 2014 it remained unclear how MDH entered the peroxisomal matrix in human cells.
The importance of functioning peroxisomes becomes apparent in severe diseases related to peroxisomes. These diseases are grouped into two classes, namely the peroxisome biogenesis disorders (PBD), which are autosomal recessive disorders and genetically heterogeneous (Weller et al. 2003), and single peroxisomal protein defects. The single peroxisomal protein defects can be grouped by the metabolic pathway they disturb: the ether phospholipid synthesis, the peroxisomal β-oxidation, the peroxisomal α-oxidation, the glyoxylate detoxification, and the H_2O_2-metabolism (Wanders and Waterham 2006a). The PDB are divided into Zellweger spectrum diseases (ZSD) and rhizomelic chondrodysplasia punctata (RCDP), which exhibits a different phenotype (Weller et al. 2003) and includes a defect in PEX7. Patients that suffer from severe form of ZSD have commonly a life expectancy of one year or less and show defects in several peroxisomal metabolic functions. These patients exhibit neurodevelopmental disorders. RCDP patients present with skeletal abnormalities and neurological symptoms, and on the molecular level both ether lipid synthesis and α-oxidation is affected (Thoms et al. 2009).

1.7 The study's goal

The goal of this study was to find physiological functional translational readthrough (FTR) in *Homo sapiens* by identification of functional domains or signals in extensions that are appended by TR and lead to differences between the parental and extended proteins. I planned to analyse the most promising putative FTR candidates (genes: *LDHB*, *MDH1*) in detail regarding their respective TR and new function. I contributed to the design of the RTP scanner that led to the detection of the consensus for high TR in humans and characterised the TR of LDHB and MDH1 as well as quantified their import into peroxisomes.

2 Material and Methods

The chemicals used are listed in Table 2.

Table 2: The chemicals used for this study listed with their manufacturing company.

Product	**Company**	**Product**	**Company**
Agar-Agar, Kobe I	Roth GmbH	LB-Medium	Roth GmbH
Agarose, molecular grade	Bioline, UK	Na_2CO_3	Merck KGaA
BigDye	Life technologies GmbH, Germany	Na_2HPO_4*H_2O	Roth GmbH
Bromophenol blue	Merck KGaA	$Na_2S_2O_3$	Sigma-Aldrich, USA
BSA	New England Biolabs	$Na_2S_2O_4$	Sigma-Aldrich, USA
Complete EDTA-free - Protease inhibitor tbl.	Roche	NaAc (pH 4.6, 85 mM)	Sigma-Aldrich, USA
Cutsmart	New England Biolabs	NaH_2PO_4	Merck KGaA
Dithiothreitol (DTT)	SERVA Electrophoresis GmbH	NEBuffer 3.1	New England Biolabs
DMEM (Dulbecco's minimal essential medium)	Biochrom GmbH	Nonidet P 40	Boehringer Mannheim, Germany
DNAse I	Thermo Scientific	Orange G	Merck KGaA
dNTPs	PEQLAB Biotechnologie GmbH	PageRuler Prestained Protein Ladder	Thermo Scientific
EDTA	Roth GmbH	PBS Dulbecco	Biochrom GmbH, Germany

continued on next page...

Table 2: *continued*

Product	**Company**	**Product**	**Company**
FBS	Biochrom GmbH	Phenylmethanesul-fonylfluoride (PMSF)	Roth GmbH
Formamide	Sigma-Aldrich, USA	Roti-Load1 (4×)	Roth GmbH
GC-Buffer (5×)	Thermo Scientific	SDS stock (20 %)	Roth GmbH
Gel Red	Biotium, Inc.	Skimmed milk powder	Roth GmbH
Imidazol	Roth GmbH	T4 DNA Ligase	Thermo Scientific
IPTG	Sigma-Aldrich, USA	T4 DNA Ligase Buffer (10×)	Thermo Scientific
KH_2PO_4	Roth GmbH	TAE	Clontech
LB-Agar	Roth GmbH	Tris-HCl	Roth GmbH
		Triton ×100	Roth GmbH

The devices and software used in this study are listed in Table 3 on the next page and Table 4 on the following page, respectively.
The reaction kits used are listed in Table 5 on page 20. The organisms used are listed in Table 6 on page 20. The plasmids and oligonucleotides used are listed in the appendix (Table 17 on page 80, Table 18 on page 82). The restriction enzymes used for cloning and plasmid preparation were ordered from New England Biolabs, the enzymes are: *Xho*I, *Eco*RI-HF, *Eco*RV, *Bsp*EI, *Bst*EII, *Cla*I, *Msc*I, *Sac*II, and *Nde*I. Buffers and solutions used in this study are listed in Table 7 on page 21.

Device	Company
Äkta Purifier	Amersham Biosciences
Amicon Ultra-4 10K MWCO	Merck Millipore, USA
BioDocAnalyze	Biometra
Centrifuge 5417 R	Eppendorf
Centrifuge GPR	Beckman
Genetic Analyzer 3130 (Sequencer)	Hitachi Applied Biosystems
His Trap HP 5 ml	Amersham Biosciences
Homogenizer/Mixer	Qiagen
Large volume centrifuge CPR	Beckmann
LAS-4000	Fujifilm
NanoDrop 1000 - spectrophotometer	Thermo Scientific
Special accuracy wheighing machine - 2001 MP2	Sartorius
Standard Power Pack P25	Biometra
Synergy Mx plate reader	Biotek
T3000 Thermocycler	Biometra
Thermostar Shaker/Incubator	DMG
Vortex Genie 2	Bender & Hobein AG
Weighing machine - excellence	Sartorius

Table 3: The devices used for this study listed with their label and their manufacturing company.

Software	Company/Source
ND1000 for NanoDrop Spectrometer	Thermo Scientific
SLR ferm 1	PD Dr. rer. nat. S. Thoms
Gen5 v1.08	Biotek
OligoCalc	(Kibbe 2007)
Sequence conversion	(Joosse and Hannemann 2013)
Prenylation site predictor	(Maurer-Stroh and Eisenhaber 2005)
TMD predictor - ΔG prediction server v1.0	(Hessa et al. 2007)
TMD predictor - TopPred 1.10	(Claros and von Heijne 1994)
TMD predictor - TMPred	(Hofmann and Stoffel 1993)
TMD predictor - TMHMM Server, v 2.0	(CBS 2013)
TMD predictor - Split 4.0 server	(Jureti et al. 2002)

Table 4: The software used for this study listed with label and providing company.

Reaction kit	**Company**
High Pure PCR Product Purification Kit	Roche, Switzerland
Nucleospin Gel and PCR Clean-up	Macherey-Nagel
NucleoSpin Plasmid	Macherey-Nagel
NucleoBond Xtra Midi/Maxi	Macherey-Nagel
Lumi-light WB substrate	Roche, Switzerland
Transfection Kit	Qiagen, Netherlands
BCA-Assay	Interchim, France

Table 5: The reaction kits used for this study listed with their name and manufacturing company.

Product	**Company**
BL21 Codon Plus Competent Cells	AGILENT TECHNOLOGIES, USA
TOP10 Competent Cells	Thermo Fisher Scientific
BIO*Blue* 10^9 Chemically Competent Cells	BIOLINE

Table 6: The organisms used for this study listed with their name and distributor.

Table 7: Lists buffers and solutions used in this study with their trivial name, purpose and ingredients.

Buffer name	**Purpose**	**Ingredients**
Annealing buffer (10×)	Annealing of oligonucleotides	DPEC-treated water, NaCl (1 M), Tris-HCl (pH 7.5, 100 mM), EDTA (10 mM)
Annealing solution	Annealing of oligonucleotides	H_2O, oligonucleotides (each 100 pM), annealing buffer (1×)
Digestion solution	Digest of plasmids	H_2O, Cutsmart, plasmid (300 ng), restriction enzymes, BSA (20 μg/ml)
Ligation solution	Ligation of insert into plasmid	H_2O, vector (ca. 50 ng), insert (ca. 10 ng), T4 DNA ligase, buffer T4 DNA (1×)
PCR reaction solution	Constructs FL-LDHB/ FL-LDHB-[UGG]	H_2O, GC-buffer (1×), dNTPs (0.3 mM), primer$_{for}$ (0.3 μM), primer$_{rev}$ (0.3 μM), polymerase (Thermo Scientific)
Sequencing solution	For PCR amplification of target sequence	Sequencing buffer (1×), Big Dye (1×), primer$_{for}$ (0.5 μl), template DNA (200 ng), DEPC-treated water
Cleaning solution	Clean-Up for sequencing analysis	DEPC-treated water, ethanol (60 %), NaAc (pH 4.6, 85 mM)
Loading buffer (6×)	Gel electrophoresis	TAE (1×), saccharose (50 %), SDS (1 %), Orange G (0.25 %)
Buffer A	Protein purification	PBS (pH 7.5), 200 mM NaCl, 25 mM imidazole, 1 mM DTT, protease inhibitor PI 1 tbl., 0.1 mM phenylmethanesulfonylfluoride (PMSF), 0.1 mg/ml lysozyme, 22.5 μg/l DNAse I, 1 mM β-mercaptoethanol (ME)
Lysis buffer	Tissue lysis	H_2O, 20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 % Triton ×100, fresh: 5 mM Dithiothreitol (DTT), 0.1 mM PMSF, protease inhibitor PI 1× (Complete, Roche)
SDS-running buffer 10×	WB running buffer	H_2O, 30.3 g Tris-base, 144 g glycin, 50 ml 10 % SDS

continued on next page...

Table 7: *continued*

Buffer name	Purpose	Ingredients
Lämmli buffer	WB loading buffer	H_2O, 160 mM Tris-HCl (pH 6.8), 8 % SDS, 0.01 g bromophenol blue, 2 ml glycerol
Transfer buffer	WB transfer buffer	H_2O, 24.8 mM Tris-base, 160 mM glycin, 200 ml methanol
PBS 10×	PBS buffer	H_2O, 10 % (w/v) NaCl, 0.25 % (w/v) KCl, 1.8 % (w/v) $Na_2HPO_4*H_2O$, 0.3 % (w/v) KH_2PO_4, adjusted with NaOH to pH 7.4
PBS-Tween with Casein	blocking solution	PBS, 1 % Tween, 5 % (w/v) skimmed milk powder
IP-buffer	IP of tissue LDHB	H_2O, 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA (pH 8), 150 mM NaCl, 0.1 % (w/v) Non-ident P-40, 1 mM DTT, 1:25 Protease inhibitor
PBS-NaCl	PBS with NaCl	PBS, 150 mM NaCl
Sodium phosphate buffer (0.1 M pH 7.4)	Buffer for proteomics	H_2O, 0.3 % (w/v) NaH_2PO_4, 1 % (w/v) Na_2HPO_4

2.1 DNA cloning

Plasmids used in this study are listed in Table 17 on page 80. Oligonucleotides used in this study are listed in Table 18 on page 82. The backbone of the dual reporter assay, vector pDRVL (PST1360, Figure 8 on the following page) encoding an N-terminal Venus tag and a C-terminal hRluc tag, was derived from pEXP-Venus-hRluc (PST1327, a gift from A. Muntau and S. Gersting) by introduction of a multi-cloning site (MCS, 33 bp) containing *Bst*EII, *Cla*I, *Bsp*EI, and *Bsi*WI restriction sites between position 1483 and 1517 of PST1327.

pDRVL The backbone vector for the dual reporter constructs was created by ligating pre-annealed oligonucleotides OST963 and OST964 into the *Xho*I site of pEXP-Venus-hRluc. The oligonucleotides OST963 and OST964 were delivered dry and dissolved in DEPC-treated water to a concentration of 100 µM. The oligonucleotides were then subjected to primer hybridisation. The annealing buffer is listed in Table 7 on the previous page. The annealing reaction was carried out in a thermocycler (T3000, Biometra) using three steps, first denaturation at

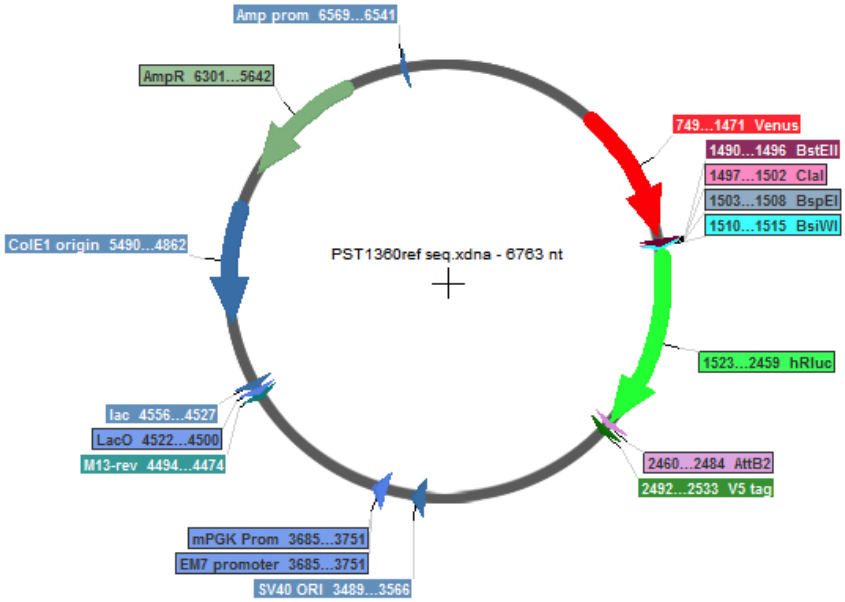


Figure 8: The Map of pDRVL (6763 bp) shows the ampicillin resistance gene, the genes encoding Venus and hRluc as well as the restriction sites of the multi-cloning site (*Bst*EII, *Cla*I, *Bsp*EI, and *Bsi*WI).

98 °C for 5 s, then gradual reduction of temperature to allow primer annealing at optimal temperature of 40 °C for 5 s. After further gradual reduction of temperature the reaction solution is kept at 10 °C.

The concentration of the annealed oligonucleotides OST963 and OST964 was measured with a spectrometer (NanoDrop 1000, Thermo Scientific). The vector pEXP-Venus-hRluc was digested to open the *Xho*I restriction site using a reaction solution (Table 7 on page 21). The reaction solution was incubated at 37 °C for 2 h. Subsequently the restriction enzyme *Xho*I was inactivated (incubation of reaction solution for 20 min at 65 °C). The digested vector was cleaned according to manufacturers manual (High Pure PCR Product Purification Kit, Roche) and the concentration of the digested and cleaned vector pEXP-Venus-hRluc was measured using a spectrometer (NanoDrop 1000, Thermo Scientific). The ligation solution (Table 7 on page 21) was prepared with a vector/insert ratio of 1 : 10 and the solution was incubated at room temperature over night.

The transformation of the ligated vector into competent *E. coli* cells (TOP10, Table 6 on page 20) was done with 20 µl cells and 2 µl ligation solution. The mixture was kept for 30 min on ice, then on 42 °C for 30 s, for 2 min on ice again, and subsequently incubated in 0.5 ml medium (LB-Medium, Table 2 on page 17) at 37 °C for 1 h on a shaker. The cells were plated on LB-Medium plates (ampicillin resistance) and incubated over night at 37 °C. Picked colonies were incubated in 3 ml LB-Medium (1 % (v/v) ampicillin) over night at 37 °C. A sample of 2 ml underwent a DNA preparation procedure (NucleoSpin Plasmid, Machery-Nagel) according to manufacturers manual, except for the use of only 40 instead of 50 µl elution buffer to yield higher concentrations of DNA. 1 µl of the elution was taken for a restriction test di-

gest (restriction sites *Eco*RV, *Bst*EII, Buffer: Cutsmart, Table 7 on page 21) to check if the cells contained the correct plasmid. This was done by electrophoresis (gel: 1 % (w/v) agarose, 1 %(v/v) gel red in 1× TAE buffer, 6× loading buffer: 50 % (w/v) saccharose, 1 % (w/v) SDS, 0.25 % (w/v) Orange G in 1× TAE buffer) with 80 – 100 V at room temperature, the gels were analysed at 366 nm excitation wavelength, the gels were used with O'GeneRuler™ DNA Ladder (Thermo Scientific).

Promising clones were sequenced (Paragraph 2.2 on page 26). Mini-prep solutions of chosen plasmids were incubated in medium (100 ml LB-Medium with respective antibiotic 0.1 % (v/v), over night at 37 °C) and subjected to a DNA preparation reaction kit (NucleoBond Xtra Midi/Maxi, Machery-Nagel) according to manufacturers manual. The concentration of the resulting elution was measured using the spectrometer (NanoDrop 1000, Thermo Scientific). pDRVL was cut open for insertion of inserts using a digest that contained the restriction enzyme *Bsp*EI, BSA, NEBuffer 3.1, pDRVL, DEPC-treated water. As *Bst*EII has a different temperature optimum as BspEI it was added 2 h later and at 60 °C instead of 37 °C. The cut vector was then cleaned from enzymes and plasmid debris by gel purification according to manufacturers manual (Table 5 on page 20).

All listed dual reporter reporter constructs PST1418, 1419-1423, 1424-1426, 1430, 1435, 1437, 1444, 1466-1470, 1472-1473, 1475-1477, 1479-1481, 1445, 1446-1454, and 1455 were derived from pDRVL by insertion of pre-annealed oligonucleotides OST1144-1145, 1148-1157, 1160-1165, 1158-1159, 1190-1191, 1198-1199, 1229-1230, 1449-1458, 1461-1464, 1467-1472, 1475-1480, 1207-1208, 1211-1228, and 1235-1236 into *Bsp*EI and *Bst*EII sites (Table 18 on page 82). The oligonucleotides were ordered from Metabion (Steinkirchen, Germany) and annealed, ligated and transformed, plated and processed as described for pDRVL. Colonies were picked, prepared with NucleoSpin Plasmid (Macherey-Nagel) and digested, run on an agarose gel (1 %) to check for the insert. The concentration of positive clones was measured, subsequently they were subjected to sequencing. Clones with correct sequences were prepared using NucleoBond Xtra Midi/Maxi (Macherey-Nagel), the concentration was obtained using the spectrometer (ND1000).

To construct plasmids PST1431, 1432, the full length (FL-)LDHB and FL-LDHB-[UGG] constructs for expression in *E. coli* and subsequent purification, the primers OST1166-1168 were used with the templates PST1365 (LDHB) and PST1389 (LDHB-[UGG]) in a PCR reaction to amplify FL-LDHB and FL-LDHB-[UGG], refer to Table 7 on page 21 for solution and Table 8 on the next page for the programme of thermocycler (T3000, Biometra).

The amplified inserts (Figure 9 on the following page) were retrieved using gel purification according to manufacturers manual (Table 5 on page 20). The inserts concentration was measured using NanoDrop spectrometer (ND1000). The insert were digested overnight at 37 °C, to prepare their ends for ligation (H_2O, Cutsmart buffer, *Nde*I, *Eco*RI-HF). The vector was prepared for ligation (H_2O, Cutsmart buffer, *Nde*I, *Eco*RI-HF, *Sac*II) overnight at 37 °C. The digested inserts and vector were run on 1% agarose gel and cut out for gel purification according

Purpose	T [°C]	Duration	Repetitions
Initial denaturation	95	3 min	
Denaturatoin	98	20 s	25 cycles
Primer annealing	60	30 s	
Primer extension	72	1 min 2 s (60 s/kb)	
Final extension	72	4 min	
Cooling	4	∞	

Table 8: The programme as it was used for the PCR to amplify FL-LDHB and FL-LDHB-[UGG].

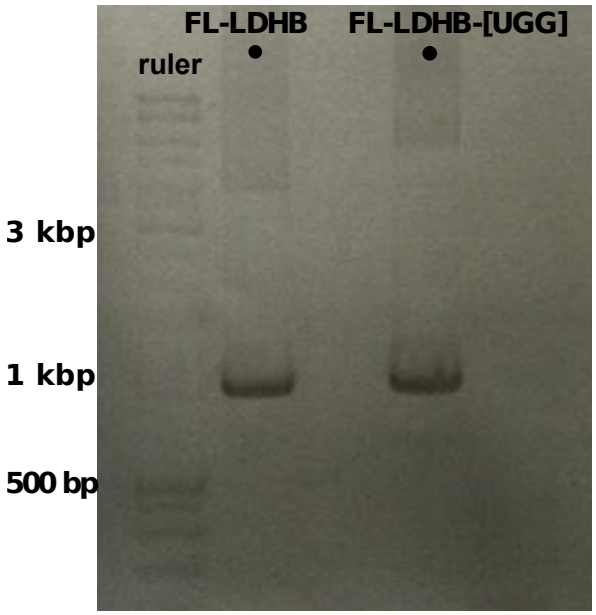


Figure 9: PCR product of inserts FL-LDHB, FL-LDHB-[UGG] for pET-41(+) vector, the insert runs with the 1 kbp band. Gel: 1 % agarose. Acquisition settings: brightness 50, contrast 50, time 440, gamma 50. Ruler, O'GeneRulerTM DNA Ladder (Thermo Scientific). (k)bp, (kilo) base-pair

to manufacturers manual (Table 5 on page 20). The elution's concentration was measured with a spectrometer (ND1000). Then the inserts were ligated into PST884 (pET41a(+), 5933 bp, kanamycin resistance, Merck Millipore, USA) at ratio (vector:insert) of 1 : 10 over night at 6 °C. The resulting constructs pET41a-FL-LDHB and pET41a-FL-LDHB-[UGG] were transformed into competent cells (BIO*Blue*, Bioline), the cells were plated (LB-medium plates, kanamycin 1:1000). Colonies were picked, incubated in 3 ml LB-media (with kanamycin) over night at 37 °C. Preparation of grown colonies using NucleoSpin Plasmid (Macherey-Nagel). Test digestion (H_2O, Cutsmart, *Xma*I, *Nde*I) for 1 h at 37 °C of colonies, control of digest using 1% agarose gel. Positive clones were sequenced. Clones (PST1431, 1432) with correct sequences were prepared using NucleoBond Xtra Midi/Maxi (Macherey-Nagel), the concentration was obtained using a spectrometer (ND1000).

2.2 Sequencing reaction to confirm plasmids

The region of interest of each constructed plasmid was sequenced, therefore a sequencing solution (Table 7 on page 21) with 200 ng of plasmid DNA was subjected to a thermocycler (T3000, Biometra; 25 cycles of 96 °C for 10 s, and 60 °C for 4 min). The sequencing solution was added to 350 µl of cleaning solution (Table 7 on page 21), the solution was mixed (Vortex Genie 2, Table 3 on page 19), and incubated at room temperature for 15 min. The solution was spun at 14, 000 rpm for 15 min, supernatant was removed, 300 µl ethanol (70 %) was added, the solution was mixed, spun for 7 min and supernatant was removed. The remains were incubated for 10 min at 37 °C with open lid and the dried pellets were resuspended in 10 µl formamide. Subsequently the plasmid solution was analysed in a sequencer (Table 3 on page 19). For DRVL-constructs the primer p24 (pEC/YFP∗61) was used. The sequencing results were analysed and evaluated using SerialCloner (version 2.6, Serial Basics) and 4Peaks (version 1.7.2, Nucleobytes Inc.).

2.3 Cell culture and transfection

HeLa cells and human skin fibroblasts were maintained in low glucose Dulbecco's minimal essential medium (DMEM), HEK, HT1080, U118, U373 and COS-7 cells in high glucose DMEM. Culture media were supplemented with 1 % (w/v) glutamine, 5˜10 % (v/v) heat inactivated fetal calf serum (FCS), 100 units/ml penicillin, and 100 µg/ml streptomycin. For U118 cells, 1 % non-essential amino acids and 1 % pyruvate were added to the media.
Cells were split when confluent reached 90 %, the new media (see above) was pre-warmed, the cells media was aspirated, cells were washed with PBS (10 ml), PBS was removed, cells were incubated at 37 °C in a humidified 5 % CO_2 incubator for 2 min with trypsin (1 ml) to loosen the cells from the surface. Trypsin was neutralised with 9 ml medium, the cells were spun at 800 rpm for 5 min, the pellet was retrieved and suspended in 5 ml of media. One ml of the suspension was added to 11 ml of fresh media, the solution was incubated at 37 °C in a humidified 5 % CO_2 incubator for further use.
Cells were transfected using Effectene transfection reagent (Transfection kit, Qiagen, Germany) as described by the manufacturer. Plasmids were diluted in Buffer EC (Transfection kit, Qiagen, Germany), and Enhancer (Transfection kit, Qiagen, Germany) and incubated for 5 min at room temperature. Effectene was added and incubated for 10 min at room temperature. Pre-warmed medium was added to the HeLa cells and to the transfection mixture which was then added to cells and incubated at 37 °C in a humidified 5 % CO_2 incubator for 24 h. Then, 6 h after transfection, transfection reagent was removed, and, where indicated, geneticin (G418) was added at a concentration of 100 µg/ml.

After I learned how to proceed in cell culture and after the DRVL assay was optimised, transfections in cell culture were carried out by C. Dickel.

2.4 Dual reporter assay and readthrough calculation

The dual reporter assay was carried out in two versions, the 12-well and the 96-well protocol. The 96-well protocol allows a higher throughput of constructs and measurements. The dual reporter constructs used Venus and hRluc, Venus is characterised as follows: $\lambda_{exc} = 515\,\text{nm}$, $\lambda_{em} = 528\,\text{nm}$, quantum yield 0.57, photostability: 50 % brightness after 15 s.

12-well protocol Transfection: 1×10^5 cells per well were seeded 24 h before transfection. Media was removed using aspiration, cells were washed with PBS (1×), transfected with 300 ng of plasmid DNA. Media was changed again 6 h after transfection. Cells were washed with PBS (1×), 1 ml warm media was added per well, which contained 100 µg of G418 if indicated. Cells were incubated for 24 h at 37 °C in a humidified 5 % CO_2 incubator.
Harvest: Media was removed using aspiration, cells were washed with PBS (1×), PBS was removed and cells were treated with trypsin (100 µl for 2 min at 37 °C. After addition of 1 ml media the cells were spun at 1000 rpm for 5 min, the media was removed. The wet pellet was resuspended and lysed in Renilla Luciferase Assay Lysis Buffer (Promega, Madison, Wisconsin) according to the manufacturer's manual, plate was shaken for 15 min at room temperature. Cells were spun down (14, 000 rpm, 2 min, 4 °C) and supernatants were stored at −80 °C.
For Venus fluorescence measurement, cell lysates were diluted 1 : 25 in PBS and analysed at 485 nm excitation, 530 nm emission (sensitivity: 130) using a Synergy Mx plate reader (Biotek, Winooski, Vermont). Settings of measurement in *well mode*: shake (medium, 2 s), delay (1 s), read with detection method *fluorescence*, read type *end point*, read speed *normal*, optics position *top*, top probe vertical offset 4 mm. PBS was used as a blank control for fluorescence measurements. Undiluted lysates (20 µl) were used to measure hRluc luminescence by the Renilla Luciferase Assay System (Promega) and the Synergy Mx plate reader (Biotek). An automated injector was used to add 100 µl Renilla Luciferase Assay Reagent. Luminescence measurement in well mode settings: shake (medium, 2 s), dispensor (100 µl Renilla Luciferase Assay Reagent, speed $v = 225\,\mu\text{l/s}$), shake (medium, 1 s), delay 1 s, read type *endpoint*, integration time 10 s, emission *hole*, optics position *top*, sensitivity 150, top probe vertical offset 15 mm. Renilla Luciferase Assay Reagent was used as a blank control for hRluc luminescence measurements. Each construct was analysed in n biological replicates with n being $3 \leq n \leq 10$ and each biological sample was measured in triplets.
The mean and standard deviation of these replicate's triplets were computed. To obtain TR rates, the ratio of hRluc/Venus signal was calculated, and the readthrough of control construct pDRVL was set to 100 %. The quotient q of the fluorescence f and luminescence l signal and

its standard deviation σ_q were calculated for each replicate using uncertainty propagation for computation of standard deviation (Equation 1). The standard deviations of the fluorescence and luminescence signals being σ_f and σ_l, respectively.

$$\sigma_q = \sqrt{\sigma_f^2 \times \left(\frac{\partial q}{\partial f}\right)^2 + \sigma_l^2 \times \left(\frac{\partial q}{\partial l}\right)^2} \tag{1}$$

Let $w_i = 1/\sigma_{qi}^2$ be the weight of a readthrough value from replicate i with σ_{qi} being the error of the ratios q_i. Then the weighted mean $\bar{x}$ of the replicates and its error $\sigma_{\bar{x}}$ were calculated according to Equation 2 and 3, respectively.

$$\bar{x} = \frac{\sum_i^n (x_i w_i)}{\sum_i^n w_i} \tag{2}$$

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{\sum_i^n w_i}} \tag{3}$$

96-well protocol Cells were washed with PBS and lysed by Renilla Luciferase Assay Lysis Buffer (Promega, Madison, Wisconsin) with 50 µl per well. The 96-well plate was incubated at 20 °C on a shaker for 15 min. Lysates were used for analysis subsequently. For Venus fluorescence measurement, cell lysates were diluted 1 : 6 in PBS (final volume 60 µl) and analysed at 485 nm excitation, 530 nm emission (sensitivity: 130) using a Synergy Mx plate reader (Biotek, Winooski, Vermont). PBS was used as a blank control for fluorescence measurements. Undiluted lysates (40 µl) were used to measure hRluc luminescence by the Renilla Luciferase Assay System (Promega) and the Synergy Mx plate reader (Biotek). An automated injector was used to add 100 µl Renilla Luciferase Assay Reagent. Luminescence was read 2 s after injection and integrated over 10 s (sensitivity: 150). Renilla Luciferase Assay Reagent was used as a blank control for hRluc luminescence measurements. Each construct was analysed in n biological replicates with n being $4 \leq n \leq 6$.
To obtain readthrough rates, the ratio of hRluc/Venus signal was calculated, and the readthrough of control construct pDRVL was set to 100 %. Subsequently the arithmetic mean and standard deviation were computed for each construct from its biological replica.

2.5 Immunofluorescence, microscopy and quantification

Transfected LDHB and LDHA fusion constructs were detected in HeLa cells by combined direct fluorescence and immunofluorescence experiments (antibodies, Table 9 on the following page). Endogenous LDHB was analysed in HeLa, U118, human skin fibroblasts and COS-7 cells by immunofluorescence.
Fluorescence microscopy was done using a 100× oil objective (1.3 NA) with a Zeiss Im-

Construct	primary Ab	secondary Ab
endogenous LDHB	anti-LDHB mouse monoclonal antibodies, dilution 1:500 (Abnova, Taiwan)	Alexa488 labeled Ab, dilution 1:200 (MoBiTech, Germany)
PEX14	anti-PEX14 rabbit polyclonal antibodies, dilution 1:200 (ProteinTech, Chicago, Illinois)	Cy3 labeled Ab, dilution 1:200 (Jackson Immuno Research, West Grove, Pennsylvania)
PEX14 (when eCFP-LDHA present)	anti-PEX14 rabbit polyclonal antibodies, dilution 1:200 (ProteinTech, Chicago, Illinois)	Alexa647 labeled Ab, dilution 1:200 (Jackson Immuno Research, West Grove, Pennsylvania)

Table 9: List of antibodies and constructs used for immunofluorescence. Primary and secondary antibodies are listed with their respective dilutions and manufacturers.

ager M1 fluorescence wide field scope equipped with the Zeiss Axiocam HRm Camera and Zeiss Axiovision 4.8 acquisition software. Acquisition settings: DAPI (λ_{exc} = 350 nm, λ_{em} = 470 nm), eYFP (λ_{exc} = 514 nm, λ_{em} = 527 nm), Alexa647 (λ_{exc} = 652 nm, λ_{em} = 668 nm), Alexa488 (λ_{exc} = 499 nm, λ_{em} = 519 nm), Cy3 (λ_{exc} = 512 nm, λ_{em} = 570 nm). z-Stacks with 30 images and 0.25 µm spacing were recorded and subjected to deconvolution. Where necessary, linear contrast enhancements were applied (Axiovision).
To estimate TR of endogenous LDHB and to quantify induction of endogenous LDHB by geneticin, fluorescence images from samples prepared with anti-LDHB and anti-PEX14 antibodies were recorded under identical conditions and subjected to deconvolution. To quantify TR induction the LDHB/PEX14 intensities were measured, and the same threshold ratios were applied to all channel pairs (ImageJ, 1.47v). Induction is expressed as the ratio of LDHB/PEX14 ratios with and without geneticin treatment, respectively. For TR estimation no treatment with geneticin was carried out, and cytosol removal was only required for one group of samples. To quantify translational readthrough the signal of the anti-LDHB antibody was measured in cells with cytosol, and set to 100 %, the anti-LDHB antibody's signal was measured after cytosol removal and normed to the 100 % control. Thus TR i.e. the remaining LDHB in peroxisomes could be quantified.

2.6 Image analysis using ImageJ

The cells cores were selected in the DAPI channel of the analysed images using the threshold function, this region of interest (ROI) was then used in the LDHB and PEX14 channels to set the pixel values in that particular region to zero. The plug-in JACoP (just another colocalisation plug-in, (Bolte and Cordelieres 2006)) for ImageJ (Schneider et al. 2012) was used to compute Pearson and Mander coefficients. The Pearson coefficient gives the correlation of

fluorophore signal between Alexa488 (LDHB) and Cy3 (PEX14) channels. Mander coefficient M1 acts as a score for pixels with LDHB signal that overlap with pixels showing PEX14 signal with respect to the overall intensity of LDHB signal, M2 is defined conversely for the PEX14 signal. Thus the Mander coefficients are not disturbed by intensity differences between the channels. As a multi-channel image might exhibit a shift between its channel the function Van Steensel's CCF of JACoP is a practical tool. The tools shifts images against each other in the x-axis and calculates a Pearson coefficient r each time, the maximum of $r(dx)$ is the Pearson coefficient of a picture without a shift. Only a x-axis can be corrected, not a z-/y-shift. This tool was used to check against $r_{max}(dx \neq 0)$.

2.7 Purification of full length LDHB and full length LDHB-[UGG]

For expression of FL-LDHB and FL-LDHB-[UGG] *E. coli* cells (BL21 Codon Plus RIL, Agilent Technologies, USA) were used. The transformed cells were incubated in pre-culture over night (shaker 300 rpm, 37 °C, 100 ml, 2.5 mg kanamycin, 3 mg chloramphenicilin). The main culture (750 ml, 9.375 mg kanamycin, 11.25 mg chloramphenicilin) was inoculated to an optical density (OD) of $OD = 0.08$. A sample of 1 ml was taken every hour (3×) for analysis of the growth speed on western blot stained with coomassie (Bio-Rad, USA) over night. After 1.5 h induction with 0.5 IPTG (Sigma-Aldrich, USA) and temperature was lowered to 20 °C. After another 3.5 h the cells were harvested and spun at high speed for 20 min. The wet pellet was lysed in buffer A (Table 7 on page 21) (Buffer with PBS pH 7.5, 200 mM NaCl, 25 mM imidazole, 1 mM DTT, protease inhibitor PI 1 tbl., 0.1 mM phenylmethanesulfonylfluoride (PMSF), 0.1 mg/ ml lysozyme, 22.5 µg/ l DNAse I) and the lysate rotated for 30 min at 4 °C. Lysates were sonicated (20 × 20 s, amplitude 40 %), spun down to remove debris (27, 000 g, 30 min, 4 °C) and filtered (filter 0.45 µm). Purification was carried out with imidazole gradient (Purifier Äkta, Amersham Biosciences; column: His_Trap_HP 5 ml, Amersham Biosciences), that was created using buffer B (buffer with PBS pH 7.5, 200 mM NaCl, 500 mM imidazole), for further parameters see Figure 35 on page 73. The fractions of purified protein were checked on commassie stained (over night at 4 °C) SDS-gel (12 %, 20 mA, sample loaded with 1 : 1 Roti-buffer (Carl Roth GmbH + Co. KG, Germany Karlsruhe) (Figure 36 on page 74). The fractions containing most protein were pooled (FL-LDHB fractions 17, 18, 19, and FL-LDHB-[UGG] fractions 15, 16, 17) and enriched using a centrifugal filter device (Amicon-Ultra-4, 10 K MWCO, Table 3 on page 19). The enriched protein solution's concentrations were measured with BCA assay (Interchim, France) and adjusted to a concentration of approximately 0.004 µg/µl (dilution with PBS pH 7.5 and 200 mM NaCl) for enzymatic activity measurement in central laboratory of UMG. The measured activity was recorded in the units U/l. The mean and standard deviation of 20 samples was taken and the equivalent mean in the units U/mg was calculated, the respective standard deviation was calculated with

uncertainty propagation (Equation 1 on page 28).
For analysis of purified FL-LDHB and FL-LDHB-[UGG] by mass spectroscopy the solution's buffer was exchanged for a sodium phosphate buffer (0.1 M, pH 7.4) to reduce the background of imidazole.

2.8 Immunoprecipitation of endogenous protein from tissue lysates

tissue acquisition and lysis Tissue of different organs from wild-type rats were shock frozen in fluid nitrogen (boiling temperatur −195.8 °C) less than 10 min after the animals death (CO_2 induced unconsciousness followed by decapitation). Heart muscle, skeletal muscle, liver, kidney, and fat tissue were acquired and stored at −80 °C. Lysis buffer was prepared fresh (Table 7 on page 21), 0.15 g tissue was mixed with 1 ml lysis buffer and homogenised 5× for 5 s (TissueRuptor, Qiagen). The lysate was spun at 14,000 rpm at 4 °C for 20 min, supernatant was divided for use in BCA-assay (25 µl), WB analysis (25 µl) and storage at −20 °C.
For the western blot lysates were sonificated (amplitude 40 %, 20 s) and incubated with 1× Lämmli-buffer at 95 °C for 10 min. Western blot was carried out with following settings, current in stacking gel $I = 20\,\text{mA}$, current in separation gel $I > 20\,\text{mA}$, current while blotting $I = 1.2\,\text{mA/cm}^2$, incubation over night at 4 °C with primary antibody (monoclonal anti-LDHB mouse antibody, dilution 1 : 5,000), after three washing steps, incubation for 1 h at room temperature with secondary antibody (goat anti-mouse antibody, dilution 1 : 5,000). Images were acquired with Lumi light kit (Roche) and a reader (LAS-4000, Fujifilm).

IP Two aliquots for each tissue were prepared with 1 mg of protein in lysate solution, which was added to lysis buffer to a final volume of 300 µl. For each tissue sample both aliquots were incubated for 1 h at 4 °C, one with, one without anti-LDHB antibody. Protein G, agarose beads (20 µl) were added to aliquots and rotated for 3 h at 4 °C. Samples were spun for 3 min at 7,000 rpm at 4 °C. Supernatant was discarded, volume (300 µl) refilled with IP-buffer (Table 7 on page 21). The washing step was repeated 5× with IP-buffer, then with PBS-NaCl (150 mM). Bead pellets were dried (aspiration) and resuspended in 40 µl Lämmli-buffer (with 600 mM DTT). The samples and controls (taken prior to addition of antibody, and after pull down of LDHB by beads from supernatant) were incubated (with Lämmli-buffer) at 70 °C for 10 min. These were used for western blot, Coomassie-stained gels and a silver-gel (not shown).

3 Results

3.1 Genome wide database scan in *Homo sapiens* for readthrough prediction

In order to find functional translational readthrough (FTR) genes that are expressed in cells of *Homo sapiens*, it was necessary to predict high translational readthrough (TR) candidates in a first step. Therefore we employed the *in silico* regression model (Schueren et al. 2014), that is explained in the introduction (Paragraph 1.3 on page 4). For a first round of whole-genome readthrough propensity (RTP) prediction, we extracted 42,000 unique stop codon contexts (SCCs) out of the 200,000 transcript stored in the Ensembl database and calculated RTP by adding up the regression coefficients of all relevant positions. A sortable list of RTP values for all human transcripts computed by our first model LIN is contained in Dataset 1 (Schueren et al. 2014).

3.2 Selection of candidates for experimental assessment of translational readthrough

To expand the data basis used to train the model and to obtain evidence that the algorithm indeed predicts TR values I selected SCCs from the list of RTP values for all human transcripts computed by the LIN *in silico* regression model. I scanned the transcripts that seemed to be good test objects for transmembrane domains (TMDs) (one of 5 employed predictors found a TMD for *THG1L*), putative farnesylation sites (CaaX-box, found in extensions of *PRDM10*, *EDN1*, *IRAK3*, and prior to annotated stop codon in *FBXL20*), and possible ER-retention signals (KDEL-sequence, found in *EDEM3* and *LEPRE1*), and most importantly PTS1 signals, see Table 4 on page 19 for employed predictors. Finally I selected 15 candidates for experimental assessment of TR values. I focused on PTS1 signals to identify peroxisomal matrix proteins and thus TMDs had to be excluded. The readthrough candidates I wanted to test using a dual reporter assay comprised genes with high (RTP>0.2), intermediate (0.2>RTP>0.05), and low (RTP<0.05) predicted RTP values (Column LIN, Table 15 on page 78). The results of my experimental assessment of TR rates for these 15 candidates are described in Paragraph 3.3 on the next page (Table 15 on page 78).

3.3 Dual reporter assay using Venus/luciferase to characterise readthrough candidates

For experimental analysis, SCCs spanning 10 nucleotides upstream and downstream of the stop codon were expressed with a 5′/N-terminal yellow fluorescent protein (Venus) and a 3′/C-terminal humanized Renilla luciferase (hRluc) tag. Normal termination of translation results in expression of only Venus protein. Stop suppression leads to the expression of hRluc. Thus Venus serves as an internal expression control. TR is expressed as luciferase activity per Venus fluorescence. This dual reporter assay (DRVL) excludes introns and exon junction complexes and, due to the relatively short stretch of variable nucleotides between the reporters, also does not allow for extensive RNA structures that could modulate readthrough (Paragraph 1.5 on page 9). Consequently, this form of the dual reporter assay focuses on the assessment of TR not influenced by specific distal elements. The candidates (pDRVL-X) selected to expand the training data used for the *in silico* regression model were tested and showed TR between 0.10 % (±0.006 %) and 2.91 % (±0.15 %) relative to the 100 % readthrough control (pDRVL) expressing the Venus-hRluc fusion protein without an intervening stop codon region (Figure 10 on the following page). The TR rates of the candidates are also listed (Table 15 on page 78) with their respective RTP values.
The construct pDRVL-VASN-[UAA UAA] serves as a negative control, as it contains a tandem termination site (2 times the tightest stop codon UAA) and gives thereby an estimation of readthrough background level (≈ 0.1 %) in the dual reporter assay, which confirms the TR background levels found earlier (Namy et al. 2001; Harrell et al. 2002).
Five out of 13 tested genomic constructs that are shown in Figure 10 on the following page exhibit TR rates that exceed the background significantly. These constructs are pDRVL-LENG1,

pDRVL-HGNC	SCC (stop codon underlined)	TR	SD	p-values
pDRVL-LENG1	CCTTACTCAC TGA CTCCTGAGGG	0.26	0.01	0.008
pDRVL-ZNF574	GATCAGTGGC TGA CTCTGCCCGA	0.31	0.02	0.006
pDRVL-EDEM3	GGATGAGCTA TGA CTTGCTAAAC	0.66	0.03	0.001
pDRVL-EDN1	AGCACATTGG TGA CAGACCTTCG	0.25	0.01	0.028
pDRVL-PPP1R3F	ATTCTCCCAA TAA AGCTTTACAG	0.18	0.01	0.024

Table 10: The genomic candidates selected to expand the training data for in silico regression model LINiter show partially TR higher than background levels. These dual reporter constructs are shown with their respective SCC, TR values and p-values. P-values were computed with two sided, unpaired, equal variance Student's t-tests. SD, standard deviation.

pDRVL-ZNF574, pDRVL-EDEM3, pDRVL-EDN1, pDRVL-PPP1R3F, which showed TR of 0.25 to 0.66 % and exceeded background levels significantly (Table 10).
Naturally also the other constructs displayed in Figure 10 on the following page and two further constructs, which will be elucidated in Paragraph 3.5 on page 37, contributed to the new

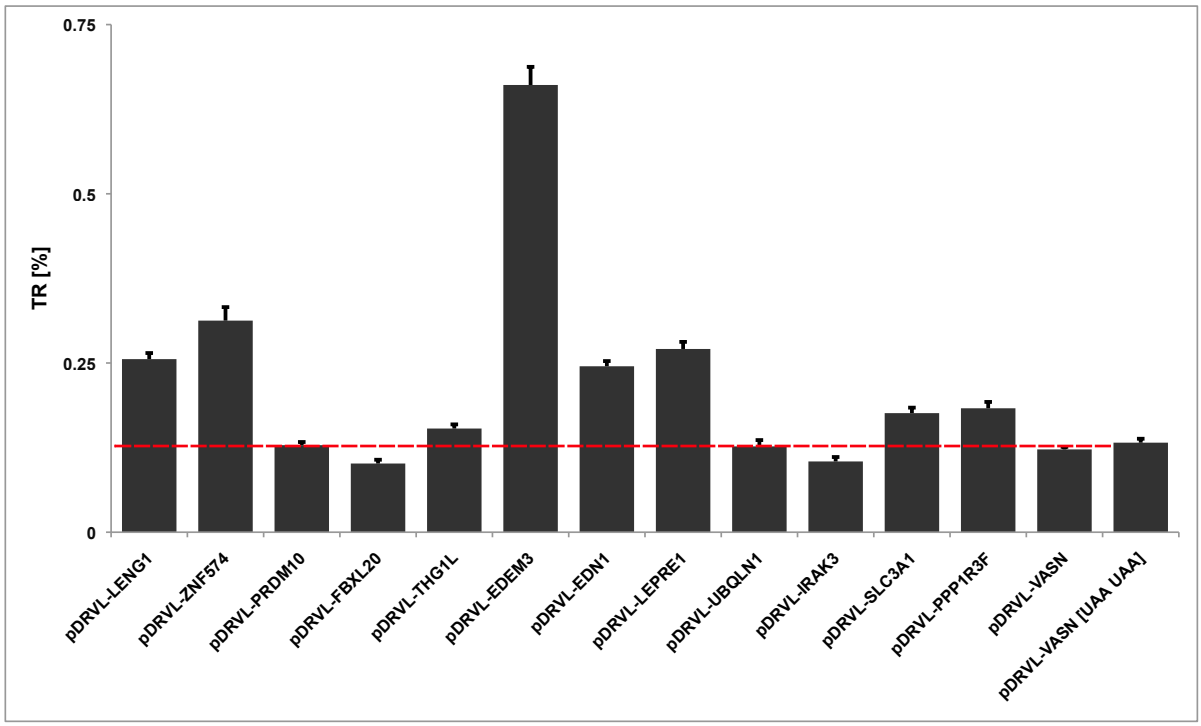


Figure 10: The translational readthrough (TR) [%] rates shown for the selected candidates of the genomic wide search list were obtained using a dual reporter assay [Venus-hRluc] (Table 15 on page 78). The weighted mean was calculated from a number of n constructs with n being $3 \leq n \leq 7$, all constructs were transfected in HeLa cells and analysed in 12-well plates. Ratios normed to positive control (100 % TR), pDRVL-VASN [UAA UAA] with tandem stop codon serves as negative control, thus the red line indicates background level. Error bars, SD.

training set. We added the new experimental data in form of these measured readthrough levels to the training data to obtain an iterative and extended RTP model (LINiter). The regression factors of the LINiter model are listed in Table 13 on page 76. Again, we applied this model to all human transcripts, see Dataset 1 (Schueren et al. 2014) for listed values. The iterative model was refined two times using feature selection (fs), i.e. those positions that contributed least to the prediction were taken out of the model successively (Schueren et al. 2014). The refined models consider less positions of the SCCs. These models LINfs5, and LINfs3 comprise contexts of 5 positions (−6, +4 to +7) and 3 positions (+4 to +6), respectively. The measured data shows correlation to the prediction of TR (Schueren et al. 2014), i.e. the RTP score (Table 15 on page 78), as the Pearson correlation coefficients for TR [%] vs LINiter, and TR [%] vs LINfs3 are $r = 0.34$ ($p = 0.002$), and $r = 0.41$ ($p = 0.0001$), respectively. So the rational refinement of the model led to a better correlation between model and measured data and more importantly revealed the consensus UGA CUA (G) (stop codon underlined) for context-driven high TR in *Homo sapiens*. Only 38 (19) genes actually carry this consensus, see Table 14 on page 77 in the appendix.

As mentioned in Paragraph 1.2 on page 1 drugs from the group of aminoglycoside antibiotics and their derivatives can induce TR. The aminoglycoside antibiotic geneticin (G418) increased

TR between 3.25 (±0.41) and 40.38 (±5.33) fold (Figure 11).

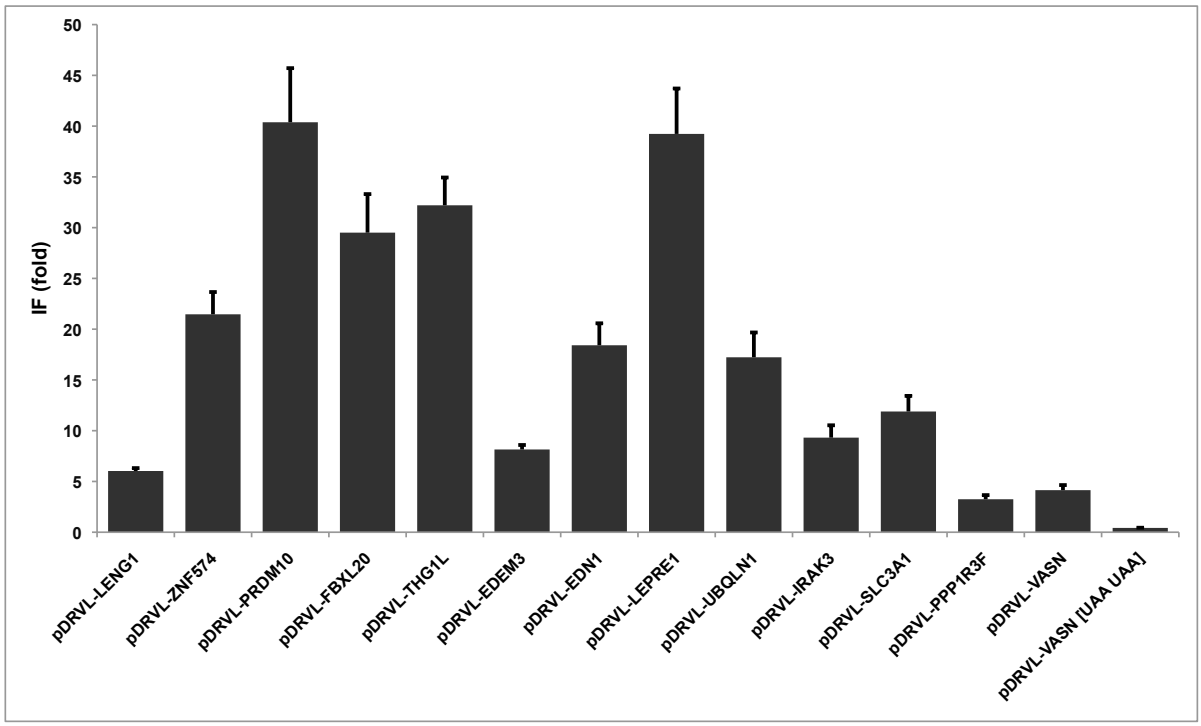


Figure 11: The induction factor (IF) is shown for selected candidates of the genomic wide search list, it was obtained using the dual reporter assay and the comparison between untreated and treated samples (treated with geneticin (G418), 100 µg/ml), the values are listed in Table 15 on page 78. The weighted mean was calculated from a number of n constructs with n being $4 \leq n \leq 7$, constructs were transfected in HeLa cells and analysed in 12-well plates. The construct pDRVL-VASN [UAA UAA] serves as a negative control, and it shows no response to treatment with aminoglycoside geneticin (G418). Error bars, SD

Geneticin only increased TR (i.e. the ratio luciferase-signal/Venus-signal) when a stop codon separated Venus and luciferase. This is a requirement for my assay, as the control's signal should not be altered by TR inducing drugs. The finding that experimental TR could be increased by treatment with aminoglycosides excludes alternative mechanisms such as RNA editing or splicing that might explain the relative increase of the luciferase over the Venus signal. Thereby I confirmed that the dual reporter assay really measures translational readthrough and does so faithfully.

3.4 Ranking of translational readthrough candidates by a product of readthrough propensity and peroxisomal targeting signal 1 posterior probability

As the goal of the study was not to find high TR candidates but to detect FTR candidates in the human genome, the next step involved an *in silico* scan for a functional domain. While I

first scanned the dataset by hand for different possible functional domains, now a predictive algorithm was used. The scan in our database focused on the peroxisomal targeting signal 1 (PTS1) (Paragraph 1.6 on page 11). A PTS1 posterior probability score was calculated for the 42,000 unique human SCCs using a PTS1 algorithm, it was adapted from plant PTS1 to human PTS1 for this purpose (Lingner et al. 2011; Schueren et al. 2014). Given that we scanned the database for PTS1 in the extensions of TR candidates, we had to combine the scores of RTP and PTS1 posterior probability. The RTP score indicated the level to which the extension was appended to a protein and the PTS1 score is an indicator for the probability that the appended signal resulted in peroxisomal targeting of the protein. Thus a protein would be most likely to target to peroxisomes by a cryptic PTS1 when the product of the RTP and the extension's PTS1 scores are high. To avoid negative scores we used the product of positively scaled RTP ($\rightarrow$RTP$^+$) LINiter scores and PTS1 posterior probabilities (RTP$^+\times$ PTS1) as a predictor of functional peroxisomal targeting by a hidden PTS1 in the extension (Schueren et al. 2014). The 42,000 unique SCCs were ranked by the product (RTP$^+\times$ PTS1) (Figure 12).

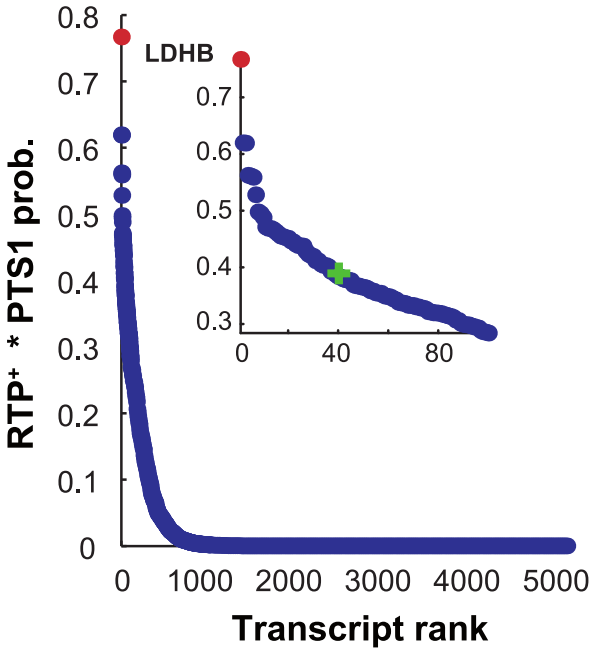


Figure 12: The product of positively scaled RTP and PTS1 probability score (RTP$^+\times$ PTS1) is shown in its genomic distribution. The scores of ranks 5,015-42,069 are zero. LDHB exhibits the highest product score and exceeds rank 2 by 24 %. The green cross indicates 50 % of the highest product score (LDHB). The figure was designed by T. Lingner (Schueren et al. 2014).

We identified LDHB, one of the two human lactate dehydrogenase (LDH) subunits, at the top (position 1 of 42, 069 entries) of our sorted list of multiplied RTP$^+$ and PTS1 scores, see dataset 1 (Schueren et al. 2014). The RTP$^+\times$ PTS1 product score shows a striking distribution. It drops by 50 % over the first 40 of 42, 069 transcripts and transcripts 5, 015 to 42, 069 have a score of zero. This rapid decline of the product score indicates that other candidates must have considerably lower RTP scores and/or targeting efficiencies and that there might be only few

candidates for FTR involving PTS1 in the human genome. Malate dehydrogenase 1 (MDH1) was found on rank 175, it's score being only 27 % of LDHB's product score. Still both enzymes, which were already part of the set of 15 TR candidates I used to expand the database, sparked my interest and I characterised them further in terms of TR and IF as well as possible secondary structures in their respective extensions.

3.5 Dual reporter assay for stop codon contexts of LDHB and MDH1

Lactate dehydrogenase subunit B and malate dehydrogenase 1 were chosen for further elucidation. Therefore their SCCs (positions −10 to +13) were cloned into the dual reporter and TR was measured. Additionally the response upon treatment with aminoglycoside geneticin (G418) was tested (Figure 14 on the next page). LDHB undergoing TR appends six amino acids to the protein plus the one inserted at the over-read termination site. MDH1 is extended by a sequence of 18 amino acids if read through, see Figure 13.

LDHB C-terminus TLWDIQKDLKDLSTOPLVSSRLSTOP

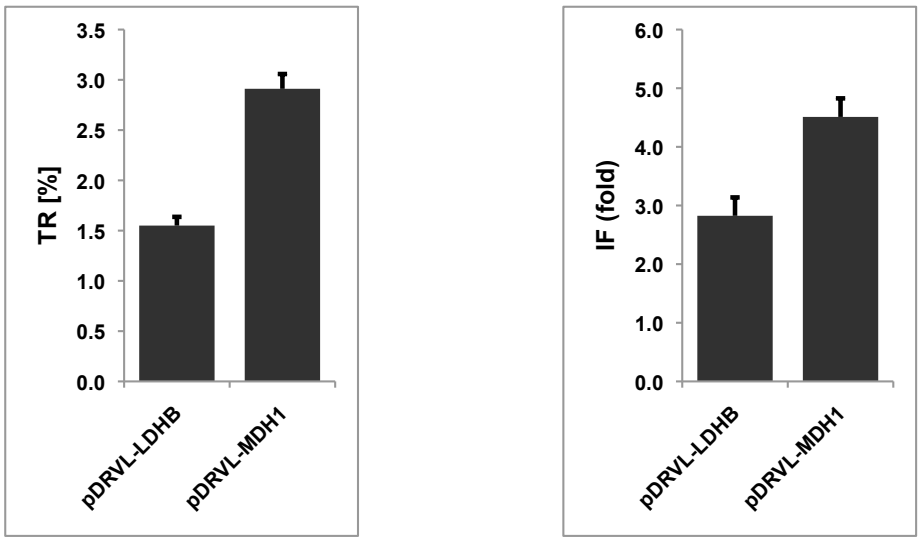
MDH1 C-terminus EKESAFEFLSSASTOPLDNDVTKCFKAEESKCRLSTOP

Figure 13: The extensions of LDHB and MDH1 shown as an amino acid sequence differ in length. Six amino acids are appended to LDHB undergoing TR and 18 amino acids are attached C-terminally to MDH1 after stop supression.

Both constructs showed readthrough rates exceeding the background level by far. LDHB exhibited TR of 1.55 % (±0.09 %) and geneticin treatment induced readthrough rates 2.82 (±0.31) fold. MDH1 showed readthrough rates of 2.91 % (±0.15 %) and drug treatment induced TR 4.51 (±0.32) fold. TR of MDH1-WT and LDHB-WT shown in Figure 14 on the next page differs between the constructs significantly, pDRVL-MDH1 exhibits roughly twice as much TR as pDRVL-LDHB.

The feature selection that revealed the consensus UGA CUA (G) (stop codon underlined) illustrates the importance of the stop codon and position +4 for TR, thus these were the positions first analysed for LDHB (Figure 15 on page 39).

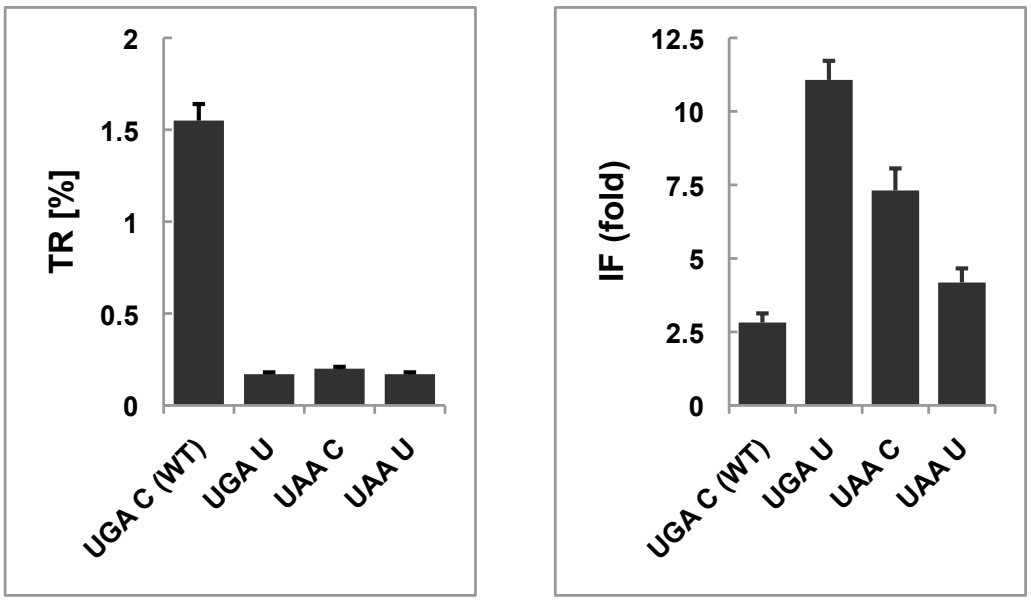
LDHB-WT was compared with three altered constructs, the first has a U instead of C at position +4, I exchanged the stop codon of LDHB-WT for the assumed tightest stop codon (UAA) to design the second construct, and the third construct exhibits the tight stop codon UAA and U at position +4. The measurement of LDHB construct and three mutants in the dual reporter assay revealed a strong dependence of TR of LDHB on the leaky stop codon UGA and the cytosine residue at position +4 (Figure 15 on page 39). The TR value of wild-type construct pDRVL-LDHB is significantly higher than TR of the altered constructs pDRVL-LDHB-[UGA U], pDRVL-LDHB-[UAA C], and pDRVL-LDHB-[UAA U]. Mutation of the stop codon or the first position downstream (pos. +4) strongly suppresses TR.



(a) (b)

Figure 14: **(a)** TR values for pDRVL-LDHB, pDRVL-MDH1 obtained with DRVL-assay. TR of pDRVL-MDH1 is significantly higher than TR of pDRVL-LDHB (p = 0.022, two sided, unpaired, equal variance Student's t-test). **(b)** IF calculated as the ratio of the construct's TR values from treated/untreated cells. The weighted mean of TR and IF were calculated from a number of n constructs with n being $5 \leq n \leq 7$. Constructs were transfected in HeLa cells in 12-well plates. Error bars, SD.

In order to further analyse the influence of nucleotides surrounding the stop codons of LDHB and MDH1, dual reporter constructs were designed with altered nucleotides at positions +4 to +6 for LDHB and exchanged stop codons and nucleotides at positions +4 to +6 for MDH1.
The readthrough rates measured for these constructs are displayed (Figure 16 on page 40). The wild-type construct (pDRVL-LDHB) exhibits significantly higher TR than any construct displayed in Figure 16 on page 40 (a) with $p < 0.002$ (two sided, unpaired, equal variance Student's t-tests). TR of constructs pDRVL-LDHB-[UGA UUA] and pDRVL-LDHB-[UGA GUA] differs significantly with G at position +4 allowing higher readthrough than U at position +4, see marker ($*$) in Figure 16 on page 40. This finding is not supported by the *in silico* derived regression factors (Table 13 on page 76). TR of constructs pDRVL-LDHB-[UGA CAA] and pDRVL-LDHB-[UGA CCA] differs significantly (Figure 16 on page 40, ($**$)). The residue A at position +5 allows higher readthrough than C at that position, which is supported by the *in silico* derived regression factors (Table 13 on page 76).
The wild-type construct (pDRVL-MDH1) exhibits significantly higher TR than any construct displayed in Figure 16 on page 40 (b) with $p < 0.000005$ (two sided, unpaired, equal variance Student's t-tests). TR of constructs pDRVL-MDH1-[UAA CUA] and pDRVL-MDH1-[UAA UUA] differs significantly, compare marker ($*$) in Figure 16 on page 40. The residue C at position +4 mediates higher readthrough than residue U at position +4, which is supported by the *in silico* derived regression factors (Table 13 on page 76). TR rates differ significantly between the constructs pDRVL-MDH1-[UAA UUA] and pDRVL-MDH1-[UGA UUA], com-

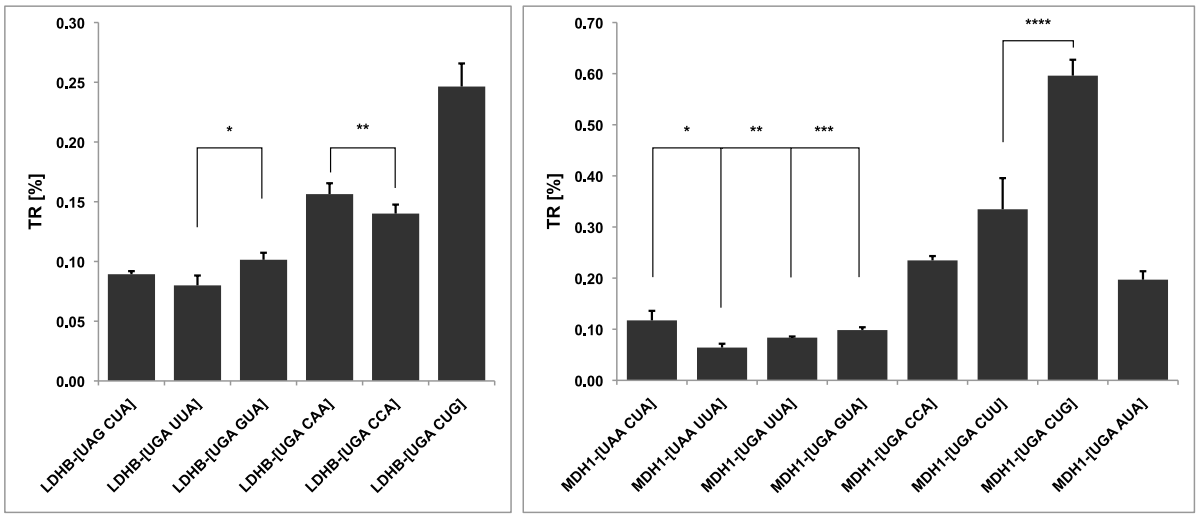


(a) (b)

Figure 15: **(a)** TR values for pDRVL-LDHB, pDRVL-LDHB-[UGA U], pDRVL-LDHB-[UAA], and pDRVL-LDHB-[UAA U] obtained with DRVL-assay. The TR of wild-type construct pDRVL-LDHB differs significantly from TR of altered constructs pDRVL-LDHB-[UGA U], pDRVL-LDHB-[UAA C], and pDRVL-LDHB-[UAA U] (p = 0.0006, p = 0.00003, p = 0.00001 respectively, two sided, unpaired, equal variance Student's t-tests). **(b)** IF calculated as the ratio of the construct's TR values from treated/untreated cells. The weighted mean of TR and IF were calculated from a number of n constructs with n being $4 \leq n \leq 8$. Constructs were transfected in HeLa cells in 12-well plates. Error bars, SD.

pare marker (∗∗) in Figure 16 on the next page. The stop codon UGA at positions +1 to +3 allows higher TR than the tight stop codon UAA. This finding meets the expected outcome (Table 13 on page 76). TR of constructs pDRVL-MDH1-[UGA UUA] and pDRVL-MDH1-[UGA GUA] differs significantly with G at position +4 allowing higher readthrough than U, see marker (∗ ∗ ∗) in Figure 16 on the next page. This is not in agreement with the *in silico* derived regression factors (Table 13 on page 76). TR of constructs pDRVL-MDH1-[UGA CUU] and pDRVL-MDH1-[UGA CUG] differs significantly with G at position +6 allowing higher readthrough than U (Figure 16 on the next page, (∗ ∗ ∗ ∗)), which is supported by the *in silico* derived regression factors (Table 13 on page 76). TR of constructs pDRVL-MDH1-[UGA AUA] and pDRVL-MDH1-[UGA GUA] differs significantly with A at position +4 allowing higher readthrough than G. The effect of A>G on TR is confirmed by the *in silico* derived regression factors (Table 13 on page 76). Also TR of constructs pDRVL-MDH1-[UGA AUA] and pDRVL-MDH1-[UGA UUA] differs significantly with A at position +4 allowing higher readthrough than U, which is supported by the *in silico* derived regression factors (Table 13 on page 76).

The analysis of context driven TR of LDHB and MDH1 leads to an order of nucleotides for high TR, which is UGA>UAA>UAG (C>A>G>U) (U>A>C) (A>G>U) (stop codon under-



(a) TR of LDHB-pos. +4 to +6 mutants

(b) TR of MDH1-pos. +4 to +6 mutants

Figure 16: TR of DRVL-assay for LDHB and MDH1 mutated in positions +4 to +6. TR was calculated as weighted mean from n constructs with $n = 4$. **(a)** shows the TR [%] of LDHB derivatives. TR differs significantly for construct pair $(*)$ and $(**)$ with $p = 0.005$, and $p = 0.03$, respectively. **(b)** shows the TR [%] of MDH1 derivatives. TR differs significantly for construct pair $(*)$, $(**)$, $(***)$, and $(****)$ with $p = 0.00001$, $p = 0.0006$, $p = 0.027$, and $p = 0.000006$, respectively. TR of pDRVL-MDH1-[UGA AUA] is significantly higher than TR of pDRVL-MDH1-[UGA GUA], and pDRVL-MDH1-[UGA UUA] with $p = 0.001$, and $p = 0.0005$, respectively. The p-values were calculated using two sided, unpaired, equal variance Student's t-tests. The constructs were transfected in HeLa cells in 96-well plates. Error bars, SD.

lined), this indicates the consensus UGA CUA for high context driven TR. Except for the order of G>U for position +4 and +6 this found influence of nucleotides/ stop codons in position +1 to +6 is in agreement with our models LINiter, and LINfs3, compare Table 13 on page 76 and Table 2 (Schueren et al. 2014).

The feature selection of the *in silico* regression model's coefficients identified positions −6, and +7 as important beside the now analysed positions +4, +5, +6 (Schueren et al. 2014). LDHB, MDH1 and also my pDRVL constructs of these genes have a G residue at position +7, thus I didn't test the position's influence here. However they differ in position −6, which is G, U for LDHB, MDH1 respectively. To test if position −6 might influence readthrough values as predicted by the model given that the other relevant positions (positions +1 to +7) bear identical nucleotides, I compared LDHB and MDH1 constructs with identically mutated SCCs of Figure 16. Two pairs were identified that do not show a significant effect of position −6 on TR and two pairs exhibit a significant difference of TR between one another. TR of construct pair pDRVL-LDHB-[UGA UUA] and pDRVL-MDH1-[UGA UUA], and construct pair pDRVL-LDHB-[UGA GUA] and pDRVL-MDH1-[UGA GUA] differs not significantly ($p = 0.5$, $p = 0.57$, two sided, unpaired, equal variance Student's t-tests). Translational readthrough of constructs pDRVL-LDHB-[UGA CCA] and pDRVL-MDH1-[UGA CCA] differs significantly ($p = 0.02$, two sided, unpaired, equal variance Student's t-test) with pDRVL-MDH1-[UGA

CCA] allowing higher readthrough than its LDHB counterpart. Thus this indicates an effect of U>G for position −6, which is in agreement with the *in silico* derived regression factors (Table 13 on page 76). TR of constructs pDRVL-LDHB-[UGA CUG] and pDRVL-MDH1-[UGA CUG] differs significantly ($p = 1.5\times10^{-7}$, two sided, unpaired, equal variance Student's t-test) with pDRVL-MDH1-[UGA CUG] allowing higher readthrough than its LDHB counterpart, again this is supported by the *in silico* derived regression factors (Table 13 on page 76). The model LINiter predicts that U at position −6 (as in the MDH1 constructs) allows higher readthrough than G (as in the LDHB constructs), however the comparison is not entirely clean as the less influential positions of my dual reporter constructs were influencing the measured TR rates although the important positions had the same nucleotides in the compared constructs. To exclude that distal elements in the extensions of LDHB and MDH1, which are conserved in mammals, and vertebrates respectively, exist and influence TR, the extensions were subjected to a prediction programme. The secondary structures of mRNA parts (nucleotide position -10 to the second in-frame stop codon) were predicted for LDHB and MDH1 in several species, including *Homo sapiens*, *Felis catus* and *Mus musculus* (Gruber et al. 2008). The predicited RNA secondary structures are shown in Figure 17 on the next page. The structures suggest no conserved element in the TR extension of LDHB and MDH1. This justifies the focus of the analysis on context driven TR, although a distal elements downstream of the extension is not excluded.

So far I described the *in silico* regression model and how I selected candidates to expand the database of the model LIN. The refined models led to the consensus for high context-driven TR and allowed us to rank our extensive gene list according to the product score of RTP^+ and PTS1 posterior probability. Within the highest ranks of the sorted list *LDHB* and *MDH1* were found with *LDHB* on the top position. I used the dual reporter assay to characterise TR of these two genes and the influence of the SCC on their TR.

3.6 Translational readthrough in multiple cell types

As mentioned in the introduction, evidence was found indicating that TR can undergo regulation due to developmental state of the organism. This was observed for genes in *Drosophila melanogaster*, compare Paragraph 1.3 on page 4, and Paragraph 1.4 on page 7. It was also observed that the TR gene VEGFAx was down-regulated in grade 2 or 3 adenocarcinoma cells of colon (human xenografttumor in mice) (Eswarappa et al. 2014). To test if TR might be differentially regulated in different cell types as well, first TR and IF of pDRVL-LDHB and one mutant of pDRVL-LDHB with a tight stop codon were analysed in COS-7, HEK and U118 cells (Figure 18 on page 43). Subsequently TR of pDRVL-MDH1 and pDRVL-LDHB was compared in HeLa, and glioblastoma cell lines (U373, and U118) (Figure 19 on page 44). This analysis was expanded to more constructs and further cell lines (Figure 20 on page 45).

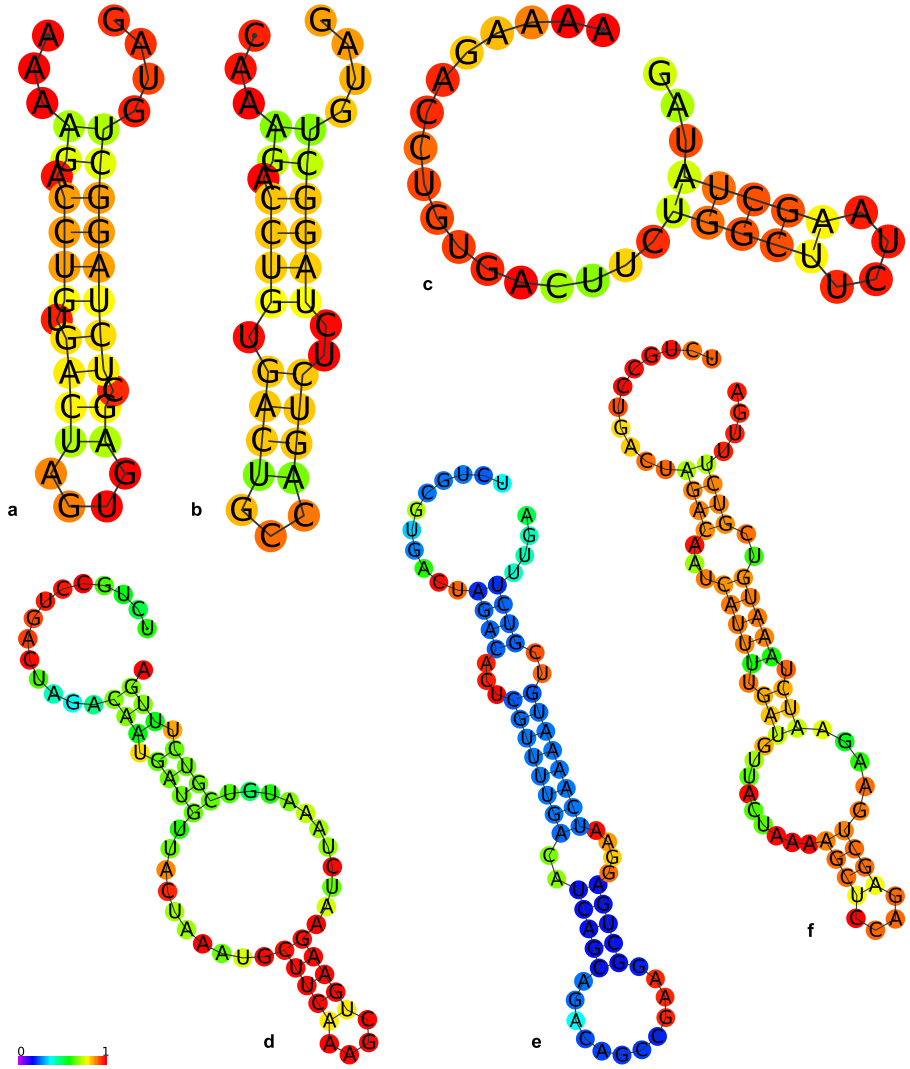


Figure 17: The RNA secondary structures do not indicate conserved structures in the hidden TR extensions of LDHB, and MDH1. From 13 analysed species three examples are shown for LDHB, and MDH1. Shown examples are LDHB ((**a**) *Homo sapiens*, **(b)** *Mus musculus*, **(c)** *Felis catus*), and MDH1 ((**d**) *Homo sapiens*, **(e)** *Mus musculus*, **(f)** *Felis catus*). mRNA structure prediction was done with RNAfold Web Server (Gruber et al. 2008), the sequence fragment was entered from nucleotide position -10 until the second in-frame stop codon. The colour code indicates the pairing probability of the nucleotides.

Similarly to the assessment of TR of pDRVL-LDBH and pDRVL-LDHB-[UAA U] in HeLa cells (Figure 15 on page 39) the first analysis of TR in different cells was carried out with these two constructs in fibroblast-like cells (COS-7), human embryonic kidney cells (HEK 293) and glioblastoma cells (U118). TR and IF by treatment with geneticin (G418) are visualised in Figure 18 on the following page. It seems apparent that readthrough rates are at different levels in each of the cell lines, however, TR of wild-type construct pDRVL-LDHB differs significantly only between HeLa and U118 cells, COS and U118 cells, and HEK and U118 cells. The high TR in glioblastoma cells (U118) in comparison to the other cell types might be noteworthy. The effect of the stop codon (UAA) on TR is as striking in COS, HEK, and U118 cells as in the HeLa cells.

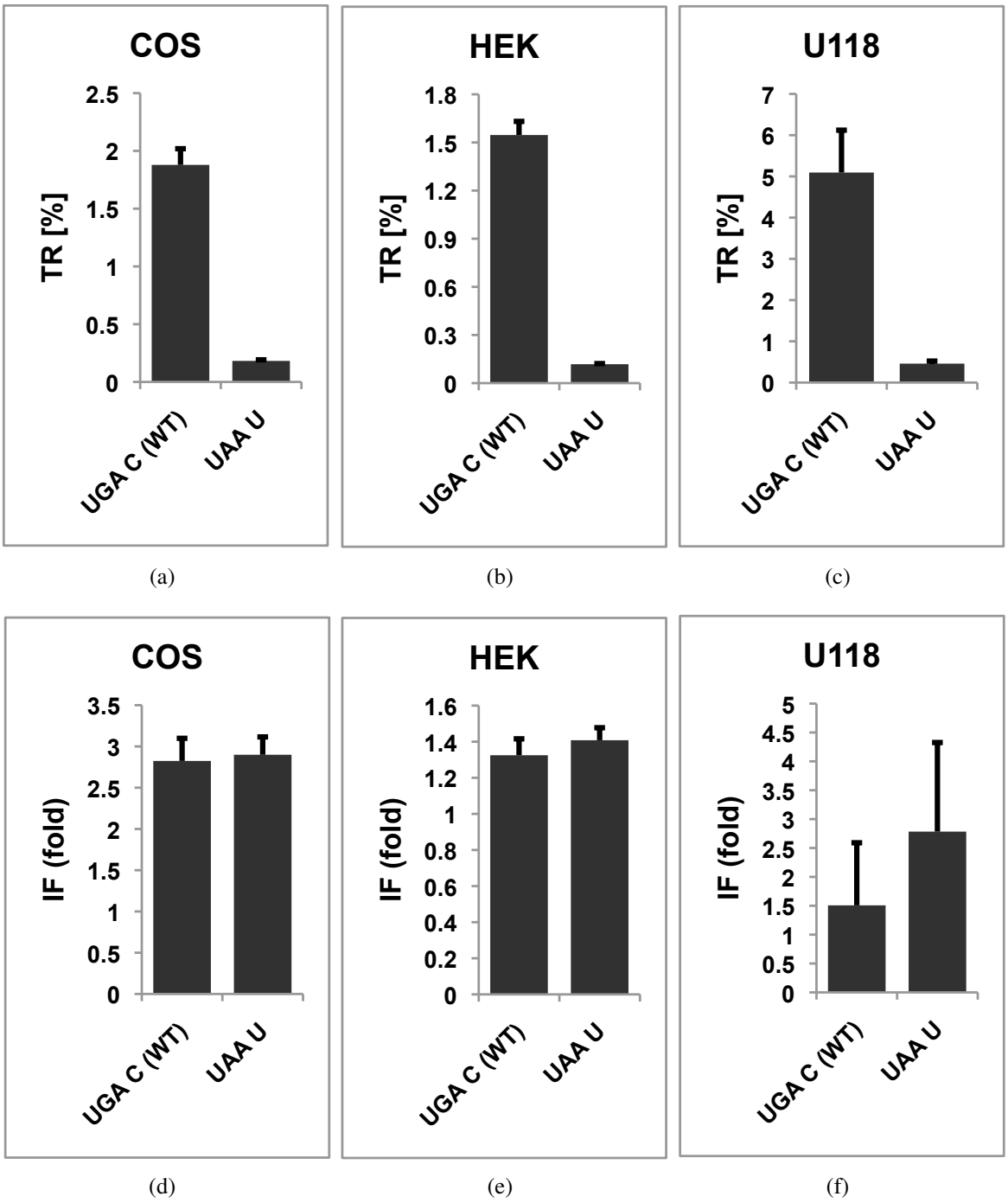


Figure 18: TR values for pDRVL-LDHB and pDRVL-LDHB-[UAA U] obtained with DRVL-assay. TR of construct pDRVL-LDHB differs significantly only between HeLa (data of Figure 14 on page 38) and U118 cells, COS and U118 cells, and HEK and U118 cells (with $p = 0.00004$, $p = 0.009$, and $p = 0.003$, respectively, two sided, unpaired, equal variance Student's t-test). IF calculated as the ratio of RT values from treated/untreated cells. The weighted mean, and IF were calculated from a number of n constructs with n being $3 \leq n \leq 10$. **(a,d)** Constructs were transfected in COS cells, **(b,e)** constructs were transfected in HEK cells, **(c,f)** constructs were transfected in U118 cells in 12-well plates. Error bars, SD.

LDHB and MDH1 were the genes of highest interest in this study and the high TR rates ob-

served for LDHB in U118 were interesting, thus I compared the TR rates of both constructs (pDRVL-LDHB/-MDH1) in HeLa cells with their TR rates in the glioblastoma cell lines U118 and U373 (Figure 19).

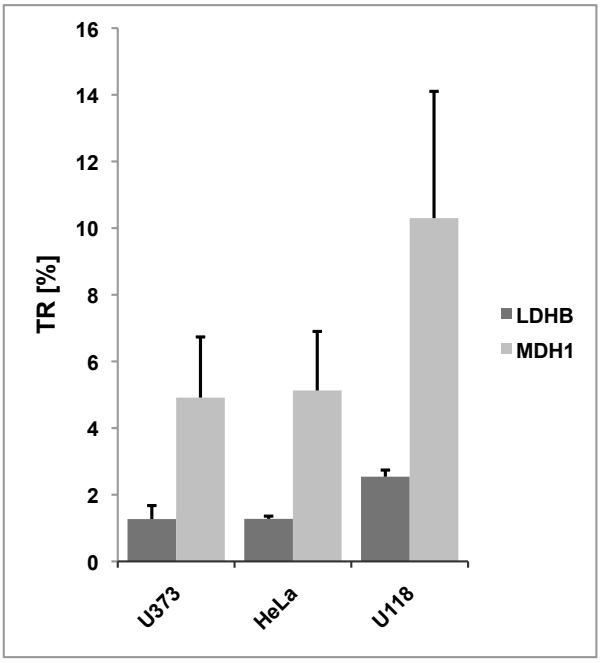


Figure 19: TR values of pDRVL-LDHB, and pDRVL-MDH1 obtained with DRVL-assay in HeLa, U118, and U373 cells. TR of pDRVL-LDHB in U118 cells is significantly higher than in HeLa, and U373 cells ($p = 0.0000003$, $p = 0.00008$). TR of pDRVL-MDH1 in U118 cells is significantly higher than in HeLa, and U373 cells ($p = 0.013$, $p = 0.011$). TR of pDRVL-MDH1 is higher than TR of pDRVL-LDHB in all three cell lines (U373, HeLa, U118, with $p = 0.002$, $p = 0.001$, $p = 0.0005$, respectively). The significance was tested using two sided, unpaired, equal variance Student's t-tests. The mean was calculated from a number of n constructs with $n = 5$ or $n = 6$. Constructs were transfected in HeLa, U373, U118 cells in 96-well plates. Error bars, SD.

Readthrough of pDRVL-LDHB in HeLa and U373 does not differ significantly between the cell lines, but expression of pDRVL-LDHB in U118 cells shows significantly higher TR than in HeLa and U373 cells. A very similar pattern is evident for pDRVL-MDH1 (Figure 19), as its expression results in TR rates that are significantly higher in U118 cells than in HeLa and U373 cells. Thus pDRVL-LDHB and pDRVL-MDH1 exhibit significantly higher TR in U118 cells than in the other two cell lines, a pattern that was observed earlier for LDHB (Figure 18 on the previous page). The construct pDRVL-MDH1 exhibits TR rates that are higher than TR of pDRVL-LDHB in all three cell lines (U373, HeLa, U118).

Further analysis of cell dependent readthrough regulation was carried out with two constructs (pDRVL-ZNF574 and pDRVL-EDEM3) that had shown significant readthrough above background in the genomic candidate analysis in HeLa cells and naturally pDRVL-LDHB, pDRVL-MDH1 and one low readthrough construct (pDRVL-IRAK3). These constructs were tested in five distinct cell lines (HeLa, HEK, U373, U118 and HT1080 cells) (Figure 20 on the following page).

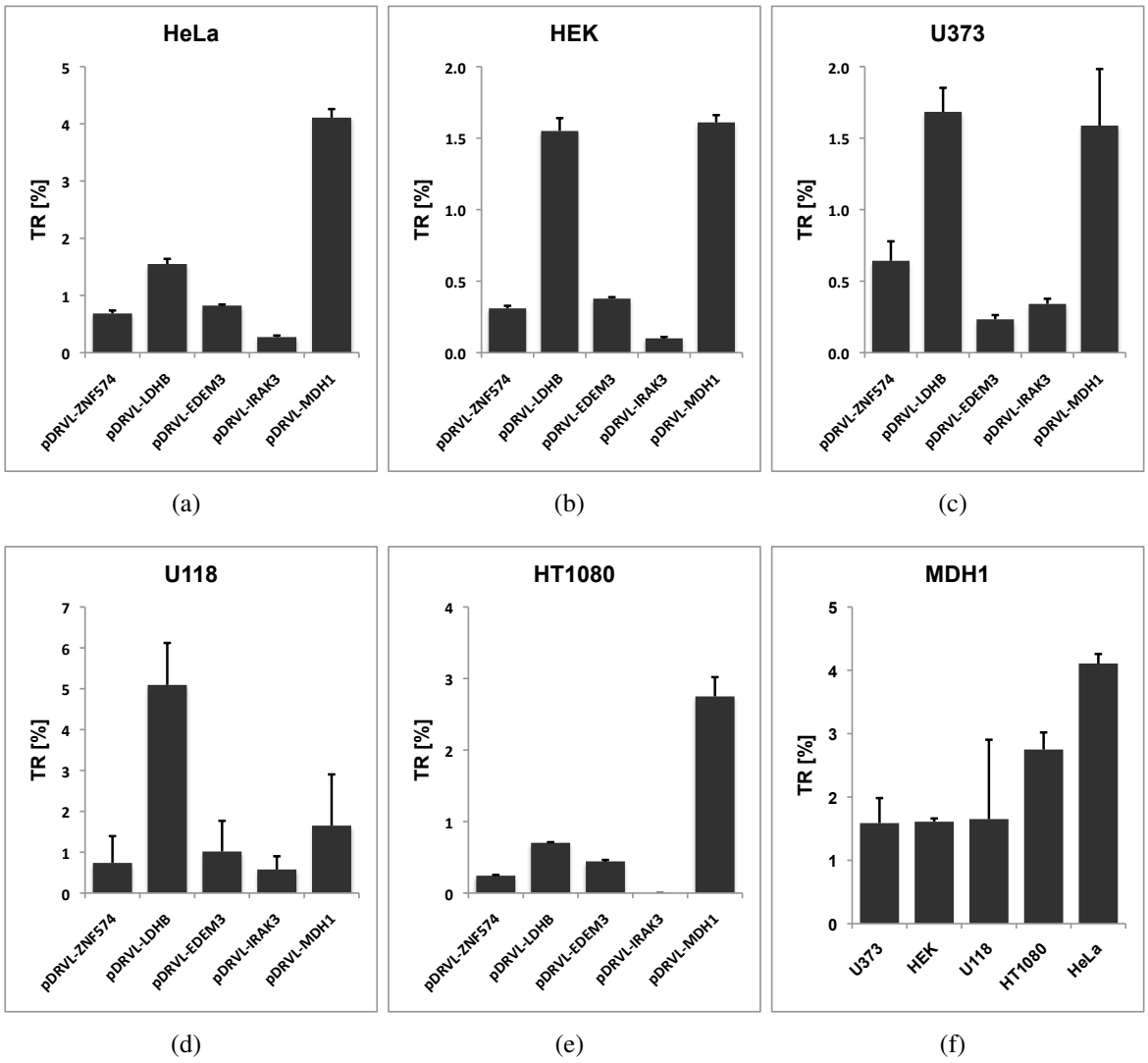


Figure 20: TR values for pDRVL-ZNF574, pDRVL-LDHB, pDRVL-EDEM3, pDRVL-IRAK3, pDRVL-MDH1 obtained with DRVL-assay. Measurements were undertaken in 12-well plates. The weighted mean was calculated from a number of $n = 3$ constructs, except for LDHB, which was measured in HeLa/HEK 7/5 times, respectively. **(a)** Constructs were transfected in HeLa cells, **(b)** constructs were transfected in HEK cells, **(c)** constructs were transfected in U373 cells, **(d)** constructs were transfected in U118 cells, **(e)** constructs were transfected in HT1080 cells, pDRVL-IRAK3 in HT1080 was compromised. The HT1080 set of constructs showed significantly lower TR than the HeLa set ($p = 0.03$, two sided, paired, equal variance Student's t-test). **(f)** TR values of pDRVL-MDH1 grouped for the different cell types as additional overview. Error bars, SD.

Of the sets of constructs in different cell lines only the HT1080 group showed significant lower TR values as a group in comparison to the HeLa group. The pattern observed before, that constructs expressed in the U118 glioblastoma cell line exhibit higher TR than in the same constructs expressed in other cell lines, was not observed in a comparison of the groups displayed in Figure 20. Thus this might be an effect of single high TR constructs (LDHB, and MDH1), but not necessarily an effect that could be generalised.

Translational readthrough of pDRVL-MDH1 is significantly lower in HEK and HT1080 than

in HeLa cells ($p = 0.0003$, $p = 0.03$, two sided, unpaired, equal variance Student's t-tests). According to this measurement readthrough of pDRVL-LDHB is significantly higher in U118 and lower in HT1080 than in HeLa cells ($p = 0.002$, $p = 0.02$, two sided, unpaired, equal variance Student's t-tests). No evidence was found for a significant difference in readthrough rates of pDRVL-ZNF574 between HeLa and any of the other cell lines (HEK, U373, U118, HT1080). Construct pDRVL-EDEM3 exhibits significantly lower TR in HEK cells and HT1080 cells than in HeLa cells ($p = 0.005$, $p = 0.01$, two sided, unpaired, equal variance Student's t-tests). Translational readthrough of pDRVL-IRAK3 is significantly lower in HEK cells in comparison to HeLa cells ($p = 0.01$, two sided, unpaired, equal variance Student's t-test).
After I characterised the influence of the nucleotides around the stop codon on TR or LDHB and MDH1, I tested if TR might be regulated differently in distinct cell lines. The constructs pDRVL-LDHB and pDRVL-MDH1 showed similar patterns in HeLa cells and glioblastoma cell lines (U118, U373). The MDH1 construct exhibits higher TR than LDHB in all cell lines tested and the cell line U118 allowed higher TR of LDHB and MDH1 than the other cell lines. A more expanded analysis of five constructs was carried out in five cell lines. All five constructs showed TR rates that were lower in HT1080 cells than in HeLa cells.

3.7 Influence of stop codon context on expression in absence of the stop codon

So far it was shown in the preceding paragraphs that the SCC including the stop codon can influence TR, this might indicate an effect exerted directly by the nucleotides neighbouring the termination site on the ribosome. Consequently I asked three questions, such as whether the influence of the stop codon can be isolated, whether the SCC is able to influence expression of luciferase in the dual reporter without the stop codon, and whether the codon frequencies downstream the stop codon are possibly relevant for TR. In other words, I wanted to know if the sequence that evolved around the stop codon had an effect on the genes expression even in absence of the stop codon and to possibly find a clue hinting the mechanism that underlies context-driven TR.
The effect of the stop codon was isolated already by exchange of UGA with one of the other two termination codons, as displayed in Figure 15 on page 39.
To check if the SCC without a stop codon is able to influence expression of the luciferase in the pDRVL constructs these SCCs were designed with a tryptophan codon (UGG) at the former termination site. Tryptophan was chosen, because it is one of the amino acids that were found to be incorporated most often for UGA termination sites in mammalian translation systems (Paragraph 1.5 on page 9). The HGNC symbols of the constructs, the nucleotide sequences and the amino acid sequences of these constructs are shown in Table 11 on page 49. The luciferase/Venus-signal of these constructs with respect to the control pDRVL is shown in

Figure 21 and listed as *SCC influence* in Table 11 on page 49. The respective constructs derived

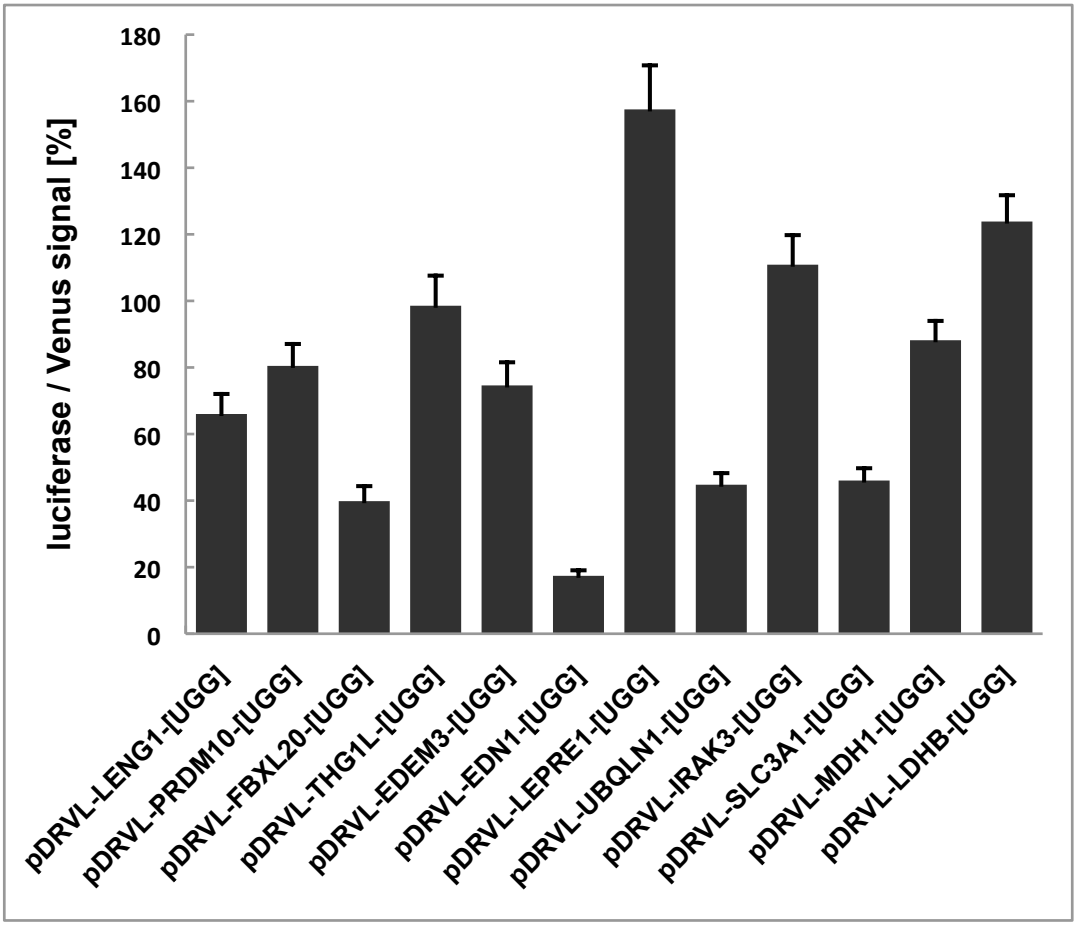


Figure 21: Results of the pDRVL-X-[UGG] measurement, with $n = 4$. The stop codons of the constructs for genomic TR measurement were exchanged for a tryptophan codon (UGG). The venus signal served as expression control, the hRluc signal is proportional to the expression of luciferase, the ratios hRluc/Venus are shown as bars in % of the positive control's ratio. The measurement was carried out in 12-well plates. Error bars, SD.

from genes *LENG1*, *PRDM10*, *FBXL20*, *EDEM3*, *EDN1*, *UBQLN1*, and *SLC3A1* exhibit a signal ratio (*SCC influence*, Table 11 on page 49) that is significantly lower than the signal of positive control pDRVL ($p_i \leq 0.024$, two sided, unpaired, equal variance Student's t-tests).The results displayed (Figure 21) indicate that the SCC without the stop codon still influences the expression of luciferase relative to that of Venus.

Table 11 on page 49 also lists TR_2, which is the signal of a pDRVL-X construct divided by its equivalent pDRVL-X-[UGG] construct. Although the controls (either pDRVL or pDRVL-X-[UGG]) that serve as a norm are quite distinct from each other, TR (pDRVL-X/pDRVL) and TR_2 correlate well with Pearson correlation coefficient $r = 0.98$. This might confirm that the stop codon's influence by far exceeds the influence of the surrounding nucleotides.

To understand the differences between the pDRVL-X-[UGG] constructs and pDRVL shown in Figure 21, the codons in the vicinity of the former stop codon were analysed. Could the low expression level of luciferase of some pDRVL-X-[UGG] constructs be the result of rare codons in the SCC? The Pearson correlation coefficient for SCC influence [%] and codon frequency

(i.e. the lowest codon frequency in the SCC) is $r = -0.869$ and it is not significant with error probability $\alpha = 5\,\%$. This suggests that it is not rare codons that influence the luciferase/Venus signal of the pDRVL-X-[UGG] constructs.

To test whether TR might be influenced by rare codons located downstream the stop codon, i.e. the abundance of tRNAs that decode certain codons might influence the speed of the ribosome and thereby the probability of mis-incorporation of a near-cognate tRNA at a termination site, the codon usage frequencies in human genome (Table 16 on page 79) were used to create a score that evaluated the context downstream of the termination site. The score simply shows the rarest codon within the first four codons downstream the stop codon ($codon\ frequency_{min}$, Table 11 on the following page). There was no apparent correlation between TR and the codon frequency score, though interestingly, leucine is coded for with a rare codon (CUA) in LDHB and MDH1, the two high TR genes. However, if the rare codon (for leucine) was really facilitating readthrough, the construct pDRVL-LDHB-[UGA UUA] would have shown a similar readthrough rate as the wild-type construct pDRVL-LDHB-[UGA CUA] (Figure 16 on page 40), because UUA is also a rarely used codon for leucine, see table of codon usage in the appendix (Table 16 on page 79). TR of pDRVL-LDHB-[UGA UUA] is by far lower than TR of pDRVL-LDHB-[UGA CUA], thus the frequency with which codons are used in the genome (i.e. abundance of respective tRNAs) might not influence TR this obviously.

The next paragraphs focus on FTR, so on the physiological effect TR has on LDHB in human cells.

HGNC	nt-sequence	Aa-sequence	TR [%]	SCC influence [%]	TR_2 [%]	ratio	codon $frequency_{min}$
LENG1	C CTT ACT CAC TGA CTC CTG AGG G	R L T H Stop L L R V	0.26	66.02	0.036	27.84	11.4
PRDM10	C ACC AAA CCA TGA CTT CCA CCC T	R T K P Stop L P P F	0.13	80.47	0.015	66.49	12.8
FBXL20	A ATG GCT ACA TAA CTC TCC AAC T	R Met A T Stop L S N F	0.1	39.84	0.018	56.40	17.4
THG1L	A GCC AGG CTT TGA CGG AAG AGT C	R A R L Stop R K S L	0.15	98.55	0.040	25.08	11.9
EDEM3	G GAT GAG CTA TGA CTT GCT AAA C	R D E L Stop L A K L	0.66	74.61	0.190	5.26	12.8
EDN1	A GCA CAT TGG TGA CAG ACC TTC G	R A H W Stop Q T F V	0.25	17.40	0.399	2.50	14.6
LEPRE1	A AGT AGC AGC TGA GAG ATA GCG A	R S S S Stop E I A I	0.27	157.54	0.136	7.35	7.1
UBQLN1	C CAG CCA TCA TAG CAG CAT TTC T	R Q P S Stop Q H F F	0.13	44.78	0.006	155.15	10.4
IRAK3	T ATG GAG ACG TGA TTT CTG CAA C	R Met E T Stop F L Q L	0.1	110.88	0.003	299.53	11.8
SLC3A1	A CAC TCA CCC TGA ATG TGT CTA A	R H S P Stop Met C L I	0.18	46.03	0.027	37.60	6.9
MDH1	T TCC TCT GCC TGA CTA GAC AAT G	R S S A Stop L D N V	2.91	88.18	1.534	0.65	6.9
LDHB	A AAA GAC CTG TGA CTA GTG AGC T	R K D L Stop L V S F	1.55	123.85	0.761	1.31	6.9

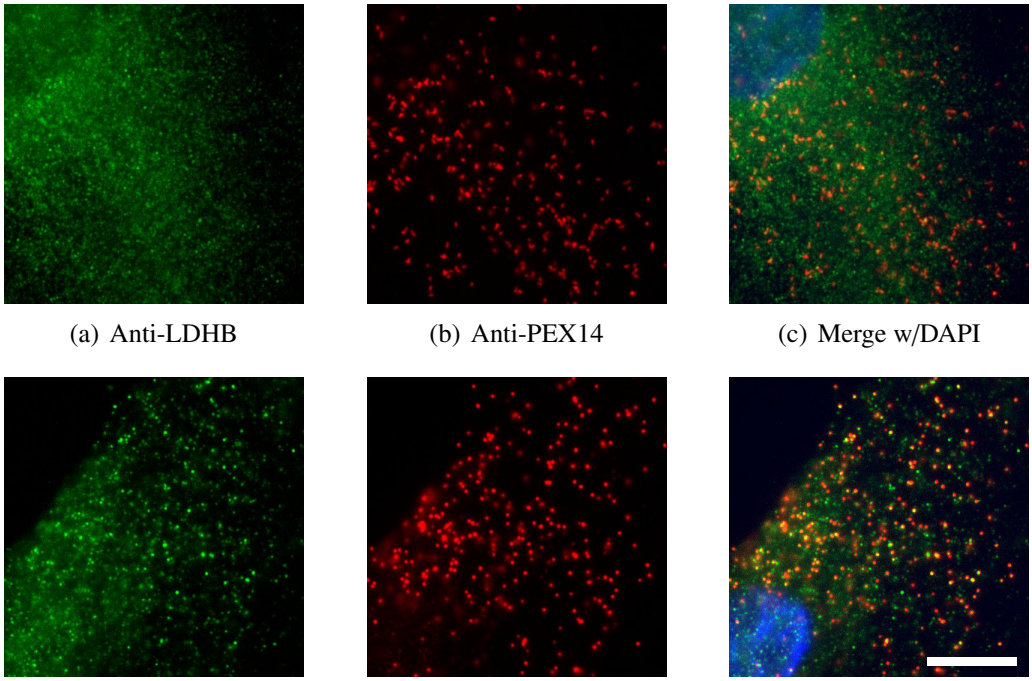
Table 11: A list of pDRVL-X and pDRVL-X-[UGG] constructs. The column *nt-sequence* lists the nucleotide environment (positions −10 to +13) of the construct's stop codons. Column *Aa-sequence* contains these environments in form of the respective amino acids. The column TR [%] displays the TR measured with the dual reporter assay for the pDRVL-X constructs, the values are normed to the positive control pDRVL, i.e. $TR = \frac{pDRVL-X}{pDRVL} \cdot 100$. The column *SCC influence* [%] displays the luciferase/Venus signal of the pDRVL-X-[UGG] constructs measured with the dual reporter assay, the values are normed to the positive control pDRVL, i.e. $\frac{pDRVL-X-[UGG]}{pDRVL} \cdot 100$. The column TR_2 [%] displays the TR measured with the dual reporter assay for the pDRVL-X constructs, the values are normed to their respective counterpart of the pDRVL-X-[UGG] constructs, i.e. $TR_2 = \frac{pDRVL-X}{pDRVL-X-[UGG]} \cdot 100$. The column *ratio* lists the signal of pDRVL-X-[UGG] with respect to the corresponding pDRVL-X construct, i. e. $\frac{1}{TR_2} = \frac{pDRVL-X-[UGG]}{pDRVL-X} \cdot 100$. The last column *codon* $frequency_{min}$ contains the lowest codon frequency found for a codon within the first four codons following downstream of the stop codon of the construct/gene, the frequencies can be found in Table 16 on page 79.

3.8 Peroxisomal targeting of endogenous LDHBx analysed quantitatively by immunofluorescence experiments

After I used the dual reporter assay to expand the database for our *in silico* regression model, to analyse TR of mainly LDHB and MDH1, and to test the influence of the nucleotides in the vicinity of their stop codons (the SCC), my further interest was to analyse the function hidden in the extension of LDHBx (extended LDHB), see Figure 22. MDH1x (extended MDH1) was shown to be peroxisomal using fluorescence imaging and cytosol bleaching (Stiebler et al. 2014).

Figure 22: The extensions of LDHB and MDH1 shown as an amino acid sequence differ in length. Six amino acids are appended to LDHB undergoing TR and 18 amino acids are attached C-terminally to MDH1 after stop suppression. The hidden PTS1 are boxed in green.

While the qualitative experiments to prove FTR of LDHB were done by coworkers (Schueren et al. 2014), I focused on the quantitative assessment of FTR using immunofluorescence microscopy. Instead of truncated regions of interest that were analysed in the dual reporter assay, here the endogenous LDHB in untransfected wild-type cells was used for FTR analysis. This analysis relies on the samples created by C. Dickel and an assay developed by R. George. The assay depletes a cell's cytosol by perforation of its membrane using digitonin and subsequent washing steps (Schueren et al. 2014). Only the content of organelles remains intact, thereby any background signal of fluorophores residing in the cytosol is drastically reduced, which in turn allows undisturbed observation of fluorophores trapped in organelles.
As pDRVL-LDHB was analysed in different cell types using the dual reporter assay, the following figures display endogenous LDHB and LDHBx in different cell types (Figure 23, Figure 24, and Figure 37). The peroxisomal enzyme LDHBx was unmasked using the digitonin-assay to remove the cytosol. FTR of endogenous LDHB is shown using immunofluorescence microscopy in four cell types, and the qualitative prove of visual co-localisation is supported by Pearson correlation coefficients that I computed for each of the images displayed in the following figures using ImageJ (plug-in JACoP). See figure legends of immunofluorescence images (Figure 23-26, Figure 37) for the Pearson correlation coefficient that indicates the correlation between the Alexa488 (LDHB) fluorophore and the Cy3 (PEX14) fluorophore. LDHBx was observed endogenously in HeLa cells (Figure 37 on page 75 in the appendix), human skin fibroblasts (Figure 23 on the next page), COS-7 cells, and glioblastoma cells (U118 cell line) (Figure 24 on page 52).
My measurements of TR of pDRVL-LDHB (shortened constructs) in HeLa cells with the dual reporter assay revealed TR rates of 1.55 (±0.09 %) for LDHB. I wanted to check this result of

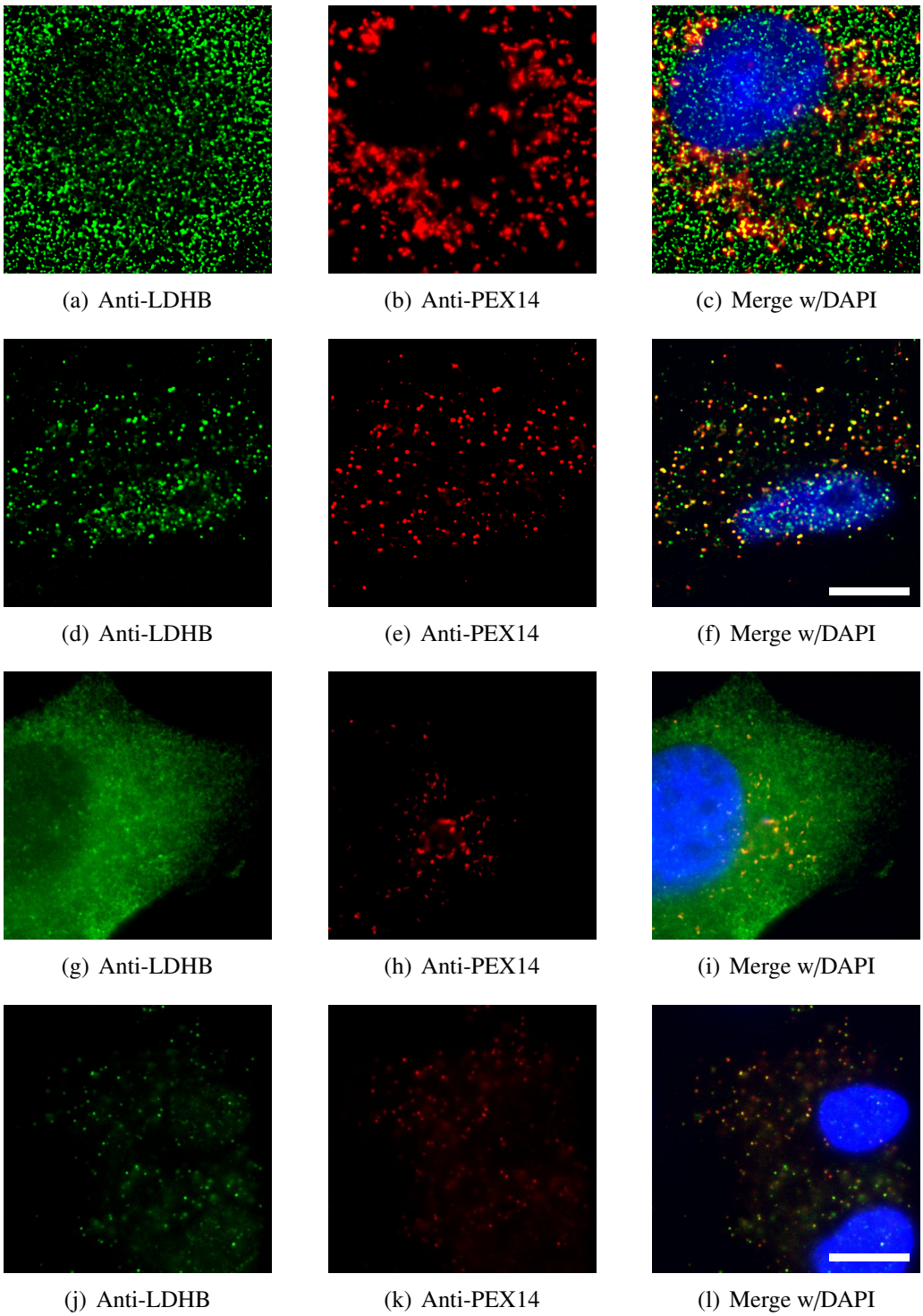


(a) Anti-LDHB (b) Anti-PEX14 (c) Merge w/DAPI

(d) Anti-LDHB (e) Anti-PEX14 (f) Merge w/DAPI

Figure 23: Endogenous LDHB in wild-type human skin fibroblasts localises to peroxisomes (PEX14), cells with (**a-c**)/ without (**d-f**) cytosol. Immunofluorescence with secondary antibodies Alexa488-coupled and Cy3-coupled. Co-localisation: Pearson correlation coefficient $r_{d-f} = 0.71$. Slides were prepared by C. Dickel. Bar 10 µm.

the dual reporter assay using immunofluorescence experiments of full length LDHB, i.e. the images of endogenous LDHB in untreated wild-type HeLa cells. These cells were not treated with any readthrough stimulating antibiotic. The analysis contained two groups of images, group 1 consists of 18 cells without cytosol removal (Figure 25 on page 53, (**a-f**)) and group 2 consists of 28 images of cells that were treated with digitonin (Figure 25 on page 53, (**g-l**)). The images of the two groups were analysed for signal of anti-LDHB antibodies against the endogenous LDHB with ImageJ. The residual intensity observed in peroxisomes of cells with removed cytosol was divided by the signal intensity in cells with complete cytosol. The ratio that is equal to the relative amount of LDHBx to cytosolic LDHB gives thereby a TR of endogenous LDHB in untreated HeLa cells. The measurements reveals endogenous TR of 2.85 (±1.22 %) ($p = 1.18E - 25$, two sided, unpaired, equal variance Student's t-test). Therefore the alternative way to measure the TR using endogenous full length LDHB in wild-type cells supports the TR rate of pDRVL-LDHB (1.55 %) measured with the dual reporter assay.
In an attempt to support the qualitative evidence that LDHB partially localises to peroxisomes in a quantitative way I used the plug-in JACoP developed for ImageJ to analyse the degree of co-localisation between the peroxisomes and the LDHB signal in cytosol or peroxisomes, the Pearson correlation coefficients of the displayed images are given in their figure legends.



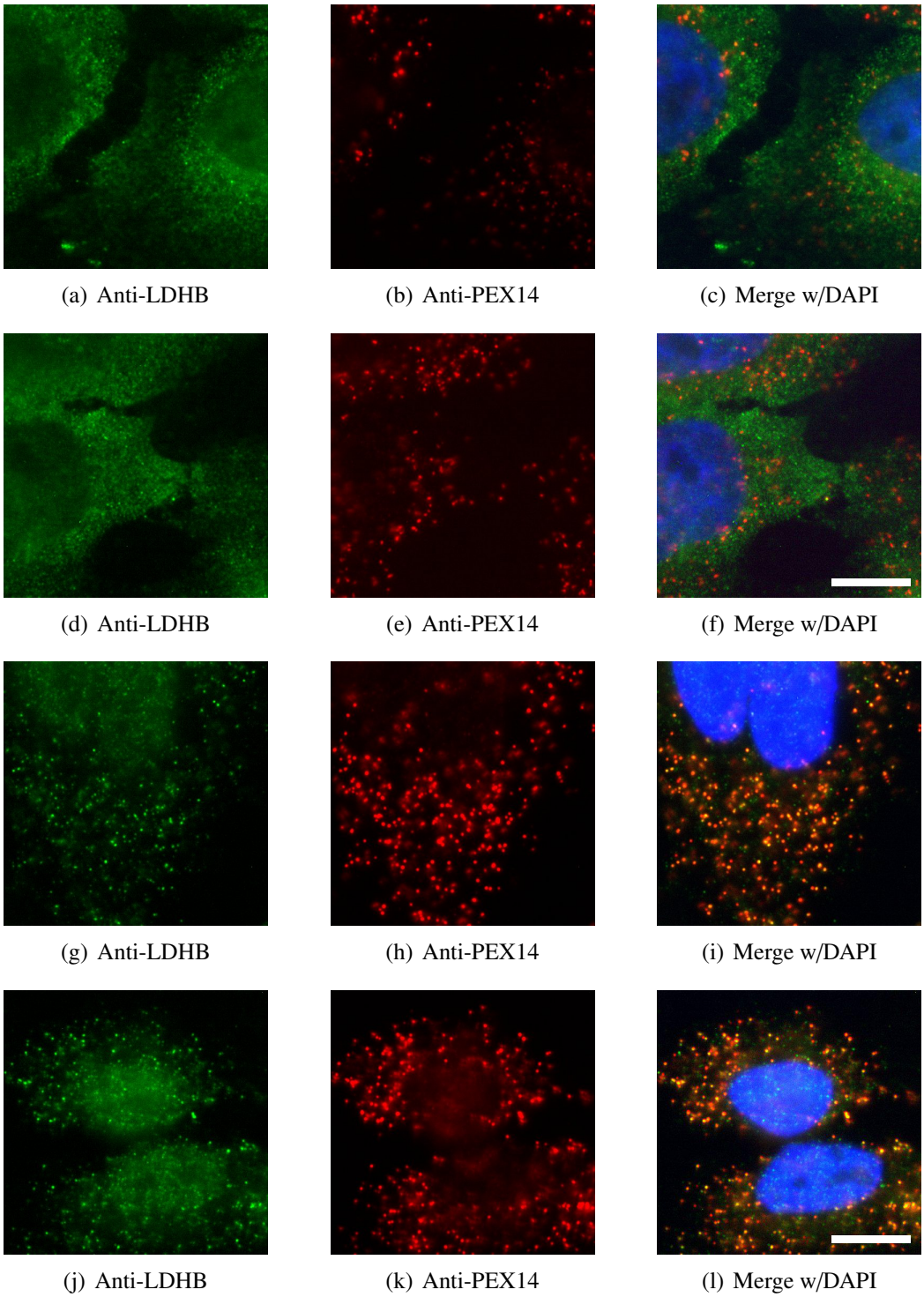
(a) Anti-LDHB (b) Anti-PEX14 (c) Merge w/DAPI

(d) Anti-LDHB (e) Anti-PEX14 (f) Merge w/DAPI

(g) Anti-LDHB (h) Anti-PEX14 (i) Merge w/DAPI

(j) Anti-LDHB (k) Anti-PEX14 (l) Merge w/DAPI

Figure 24: Endogenous LDHB in wild-type fibroblast cells (COS-7) localises to peroxisomes, cells with (**a-c**)/ without (**d-f**) cytosol. Endogenous LDHB in wild-type glioblastoma cells (U118) localises to peroxisomes, cells with (**g-i**)/ without (**j-l**) cytosol. Immunofluorescence with secondary antibodies Alexa488-coupled and Cy3-coupled. Colocalisaton: Pearson correlation coefficient $r_{d-f} = 0.70$ (COS-7), and $r_{j-l} = 0.82$ (U118). Slides were prepared by C. Dickel. Bar 10 µm.



(a) Anti-LDHB (b) Anti-PEX14 (c) Merge w/DAPI

(d) Anti-LDHB (e) Anti-PEX14 (f) Merge w/DAPI

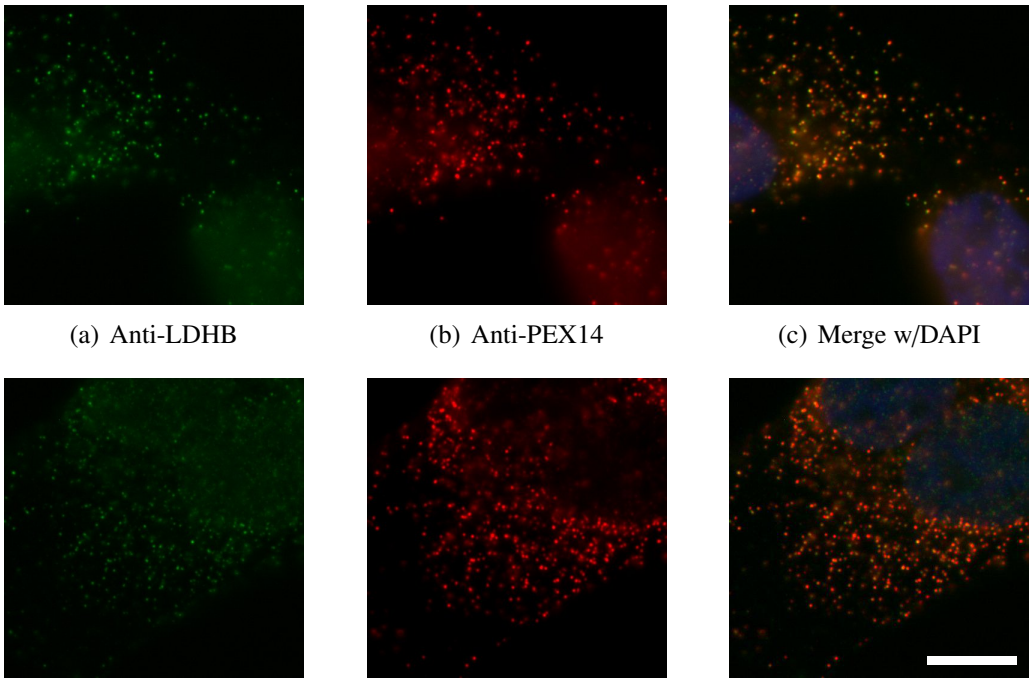
(g) Anti-LDHB (h) Anti-PEX14 (i) Merge w/DAPI

(j) Anti-LDHB (k) Anti-PEX14 (l) Merge w/DAPI

Figure 25: (**a-f**) Endogenous LDHB in wild-type HeLa cells is distributed in cytosol and peroxisomes. Eightteen images of this kind were analysed of cells without removal of cytosol to observe the overall signal of anti-LDHB antibodies in the cell. (**g-l**) Endogenous LDHB in wild-type HeLa cells is localised to peroxisomes. Twenty-eight images of this kind were analysed of cells with depleted cytosol to observe the residual signal of anti-LDHB antibodies in peroxisomes. Immunofluorescence with secondary antibodies Alexa488-coupled and Cy3-coupled. Co-localisation: Pearson correlation coefficient $r_{g-i} = 0.84$ and $r_{j-l} = 0.87$. Slides were prepared by C. Dickel. Bar 10 µm.

3.9 Influence of readthrough induction on peroxisomal localisation of LDHB

To test whether beside TR of endogenous LDHB in HeLa cells also the mechanism's induction by drug treatment with G418 could be observed, several images of cells with removed cytosol and with (Figure 26) or without (Figure 25 on the preceding page, (**g-l**)) drug treatment (G418) were analysed. I expected to see evidence of elevated residual LDHBx levels after the removal of cytosol.



(a) Anti-LDHB (b) Anti-PEX14 (c) Merge w/DAPI

(d) Anti-LDHB (e) Anti-PEX14 (f) Merge w/DAPI

Figure 26: Endogenous LDHB in wild-type HeLa cells localises to peroxisomes. Twenty-eight images of this kind of wild-type HeLa cells with depleted cytosol that underwent drug treatment (G418, geneticin 50 mg/ml, dilution 1 : 500) were analysed to estimate the residual signal of of anti-LDHB antibodies in peroxisomes. Immunofluorescence with secondary antibodies Alexa488-coupled and Cy3-coupled. Co-localisation: Pearson correlation coefficient $r_{a-c} = 0.88$ and $r_{d-f} = 0.86$. Slides were prepared by C. Dickel. Bar 10 µm.

The quantitative analysis of TR induction by application of G418 to HeLa was carried out using ImageJ. The images of HeLa cells with and without G418 treatment were acquired after removal of cytosol, for each of the two conditions 28 images were analysed. It is apparent that the treatment with G418 leads to higher residual intensity in peroxisomes after cytosol removal (Figure 27 on the next page).

The difference of residual signal between treated and untreated cells is significant and results in an induction factor of 1.86 (±1.01) fold ($p = 0.0001$, two sided, unpaired, equal variance Student's t-test) for readthrough of endogenous LDHB in HeLa cells. This IF supports the IF

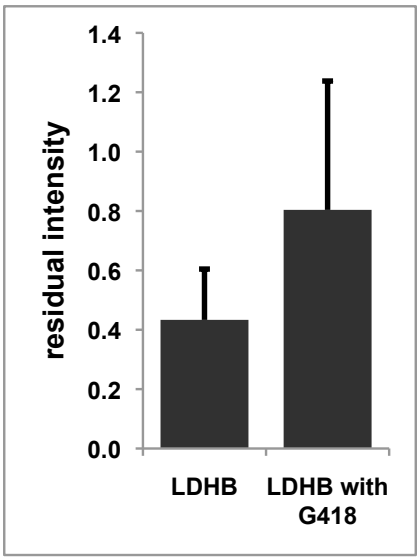
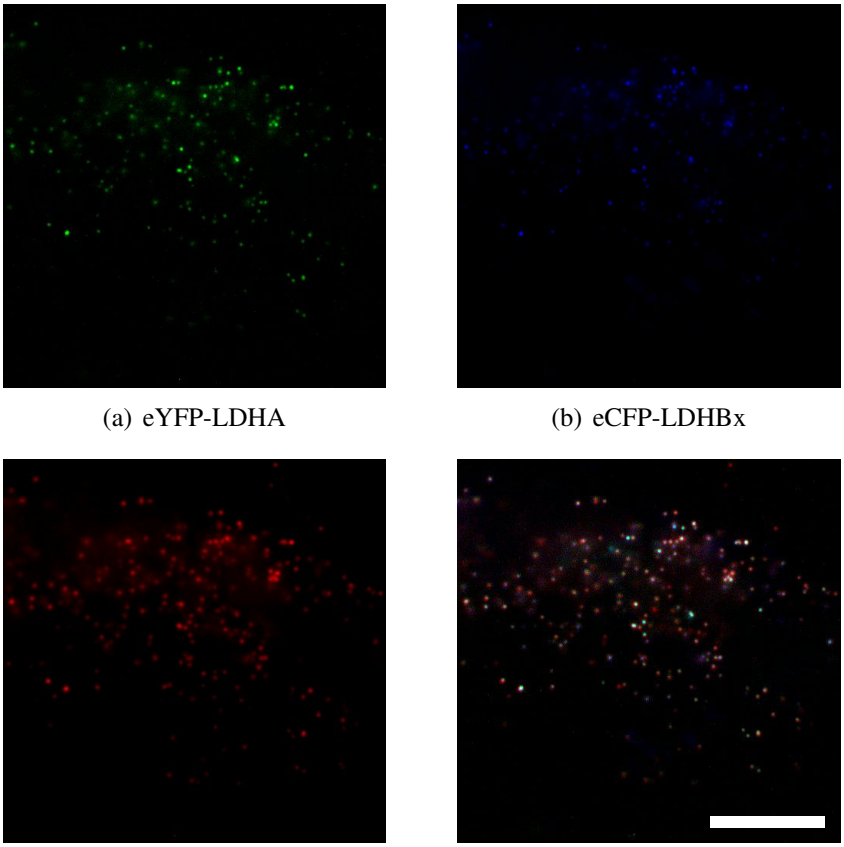


Figure 27: Image acquisition with Zeiss Axio Cam, channels adjusted equally for each image, then measurement of LDHB-Marker intensity with ImageJ. Analysis to test if HeLa cells without G418 treatment ($n = 28$) and with G418 treatment ($n = 28$) exhibit different residual intensity in peroxisomes after removal of cytosol. The analysed pictures were acquired from slides prepared by C. Dickel. Error bars, SD.

measured for pDRVL-LDHB ($IF_{pDRVL-LDHB} = 2.82$ fold) using the dual reporter assay, refer to Paragraph 3.5 on page 37.

3.10 Co-import of LDHA with LDHBx into peroxisomes

As elucidated in the introduction (Paragraph 1.6 on page 11) LDHB is one of the two subunits of LDH. While the mechanism used to import LDHB into peroxisomes is now apparent, it remained unclear how LDHA, which was found in peroxisomes as well, enters the organelle. As peroxisomes are able to import oligomers, we concluded that LDHA is imported into peroxisomes by piggy-pack co-import via LDHBx (Schueren et al. 2014). To prove this hypothesis, both genes (LDHBx, LDHA) were over-expressed as N-terminally fluorophore coupled proteins (Figure 28 on the following page). The images display the co-localisation of direct fluorescence signals from eYFP-LDHA and eCFP-LDHBx with the peroxisomal marker PEX14, which was detected using immunofluorescence. This indicates peroxisomal localisation of both LDHBx and LDHA. The visual appearance of co-localisation is striking, however to support our hypothesis I computed Pearson correlation coefficients r (ImageJ, JACoP) for the three signals. The result proves the co-localisation of both enzymes with peroxisomes and indicates strongly that both reside in the organelle, as they were already found within it by other studies. The signals of LDHBx and PEX14 show $r = 0.86$, the signals of LDHA and PEX14 show $r = 0.85$, and the signals of LDHB and LDHA show $r = 0.93$, so all signals are highly correlated with one another.



(a) eYFP-LDHA (b) eCFP-LDHBx

(c) Anti-PEX14 (d) Merge wo/DAPI

Figure 28: The co-import of LDHA by overexpressed LDHBx in HeLa cell localises eCFP-LDHA and eYFP-LDHBx to peroxisomes, cytosol removed. Secondary antibody used to stain peroxisomes was Alexa647 coupled. Slides were prepared by C. Dickel. Bar 10 µm.

3.11 Mass spectrometry analysis of LDHB and its peroxisomal isoform LDHBx

I attempted to isolate endogenous LDHBx in rat tissue samples with the ultimate goal to identify the amino acid that is incorporated for the stop codon using mass spectrometry. The concentration of protein contained in the rat tissue lysates was measured (BCA-assay, Interchim, France) (Table 12 on the next page).

The lysates were analysed using western blot, per pocket 20 µg of protein were loaded (according to BCA assay results) (Figure 29 on the following page). The high amount of fat in the fat tissue sample led to a high uncertainty in the BCA assay and during the handling of the sample. Although the lines were loaded with equal amounts of protein the bands of the heart muscle, liver, kidney, and skeletal muscle show slightly different amounts of LDHB. The result is roughly in agreement with the distribution of isozymes shown for adult rat tissues (Figure 6 on page 14), as the liver sample shows less LDHB that the heart and skeletal muscle sample.

An immuno precipitation (IP) (monoclonal mouse anti-LDHB antibody 2H6, Abnova) was

Sample	c_{mean} [µg/µl]	SD
heart muscle	20.02	1.07
liver	22.43	0.18
kidney	17.06	0.15
skeletal muscle	11.28	0.52
fat	6.06	0.18

Table 12: Lists the protein concentration of rat tissue lysates. The mean concentration values are arithmetic means of two samples. SD, standard deviation.

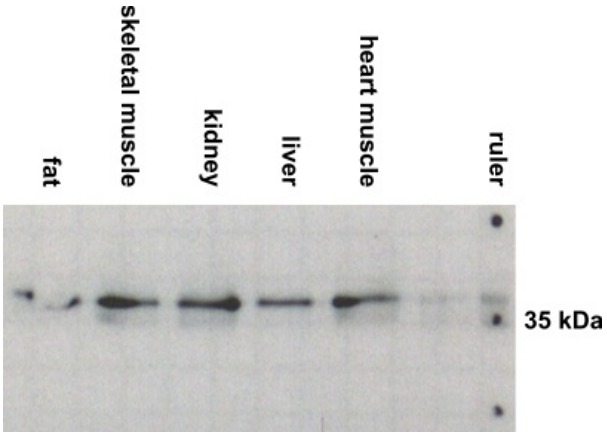


Figure 29: Equal amounts of protein from rat tissue lysates were loaded, the tissues contain varying amounts of LDHB. LDHB was stained using a monoclonal mouse α-LDHB-antibody (2H6, Abnova). The image was taken with 240 s for integration. Ruler: PageRuler Prestained Protein Ladder, Thermo Scientific. kDa, kilo Dalton.

carried out with the tissue lysates to detect endogenous LDHB and LDHBx (WB, Figure 30). As the IPs done by me did not enable me to detect the endogenous LDHBx in rat tissue using western blots or silver stained SDS gels (not shown), C. Dickel repeated the experiments. The IP performed with α-LDHB (mouse, monoclonal, 60H11, Abnova) led to a tissue dependent pull down of LDHA as well, which might be related to the oligomer composition pattern (Figure 6 on page 14). I used an α-LDHA antibody (polyclonal, epitope 216-228, Sigma-Aldrich) to check the WB for pull downs for LDHA.

The work is ongoing and we hope that the pull down of the IP will enable PD Dr. Olaf Jahn (MPI for experimental medicine) to detect and further characterise endogenous LDHBx using mass spectrometry. It would allow the identification of the amino acid inserted at the stop codon's position. To achieve this the amount of precipitated protein has to be increased, to this end the procedure needs to be optimised.

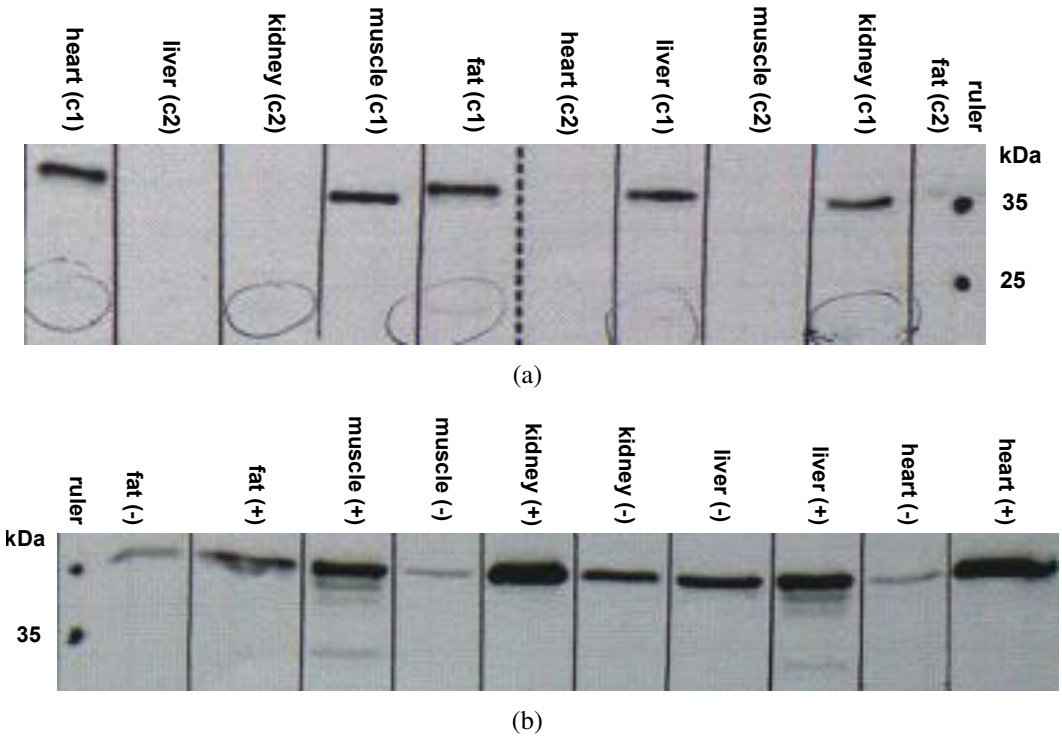


Figure 30: **(a)** Controls (c1 = input, c2 = supernatant) of immunoprecipitation (IP) of endogenous LDHB and LDHBx in rat tissue. Integration time 8 min. **(b)** IP of rat tissue lysates in search for endogenous LDHBx, samples are shown with (+) and without (-) antibody. Integration time 2 min. LDHB was stained using a monoclonal mouse α-LDHB-antibody (2H6, Abnova). Ruler: PageRuler Pre-stained Protein Ladder, Thermo Scientific. kDa, kilo Dalton.

3.12 Purification of LDHB and LDHBx to analyse their enzymatic activity

To be able to test LDHB and LDHBx for functional differences not concerning their localisation in the cell but concerning their enzymatic activity, the full length (FL-) proteins were overexpressed in *E. coli*. After transfection of *E. coli* cells with FL-LDHB-WT and FL-LDHB-[UGG] constructs, expression was induced using IPGT (Table 2 on page 17). At time points t_0 (prior to induction), $t_1 = 1\,\text{h}$ (after induction), $t_2 = 2\,\text{h}$ (after induction), $t_3 = 3\,\text{h}$ (after induction) samples were taken and subsequently compared on an SDS-Gel. Sample t_2 showed the highest amount of bacteria (Figure 31 on the following page).

The purification of the cells' lysates was done with an imidazole gradient for LDHB-WT and LDHB-[UGG] (Figure 35 on page 73) using an Äkta Purifier (Table 3 on page 19). The purification process resulted in purified protein, that was in solution in Buffer A (Table 7 on page 21) with additional imidazole from the eluting gradient. The device divided the lysate in fractions, these were checked with a commassie stained gel (Figure 36 on page 74). Of each construct three fractions that showed the highest concentration were chosen. These fractions were pooled and enriched using a filter device (Amicon Ultra-4, Table 3 on page 19). The

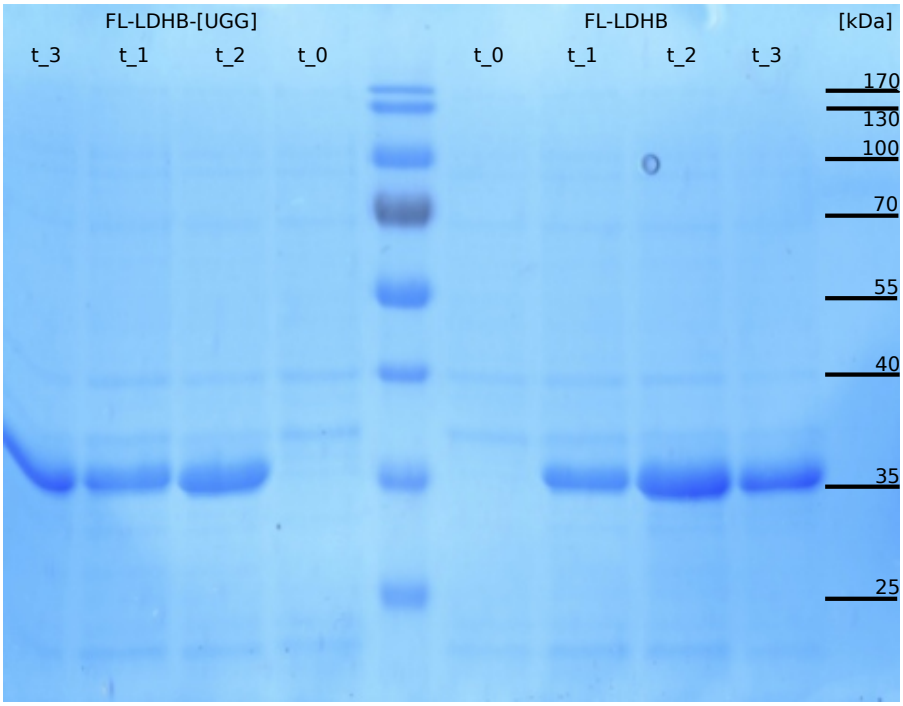


Figure 31: Samples taken at certain time points prior to (t_0) and after induction ($t_i = i\,[\mathrm{h}]$) from the growing *E-coli* in culture. The bacteria were transfected with the constructs FL-LDHB (wild-type) and FL-LDHB-[UGG]. They were grown for purification and subsequent enzymatic activity analysis and mass spectroscopy analysis. t_i, timepoints. kDa, kilo Dalton.

pooled fractions were loaded onto the filter device and spun in a precooled centrifuge (4 °C, 4,000 g, 30 min). The enriched solutions were pooled again and stored at 4 °C.

Samples of the enriched solutions were depleted of imidazole, and the buffer solution was switched from a PBS based buffer that contains imidazole to a sodium phosphate buffer (Table 7 on page 21) using a disposable PD-10 desalting column (GE Healthcare, UK). The protein samples in sodium phosphate buffer were given to PD Dr. Olaf Jahn for mass spectrometric analysis (Figure 32 on the next page).

The two enzymes exhibit distinct spectra. Though I could not precipitate endogenous LDHBx and determine the physiological amino acid(s) that decode(s) the stop codon of LDHB, I did analyse the over-expressed enzyme LDHBx. As the two enzymes are clearly distinguishable, the result suggests that the endogenous extension could be detected using mass spectrometry.

Additionally the enzymatic activity of FL-LDHB (pooled fractions 17, 18, 19) and FL-LDHB-[UGG] (pooled fraction 15, 16, 17) was analysed (Figure 33 on the following page). For this purpose samples with a similar concentration of protein were created, their concentration was measured with a BCA assay, this allowed me to reach a concentration that would result in activities within the optimal range of the central laboratory's device ($\approx 200\,\mathrm{U/l}$). The catalytic activity of the enzymes was assessed in central lab of UMG Göttingen. Two samples with different concentrations of both FL-LDHB and FL-LDHB-[UGG] were analysed, the mean activities [U/mg] are shown (Figure 33 on the next page).

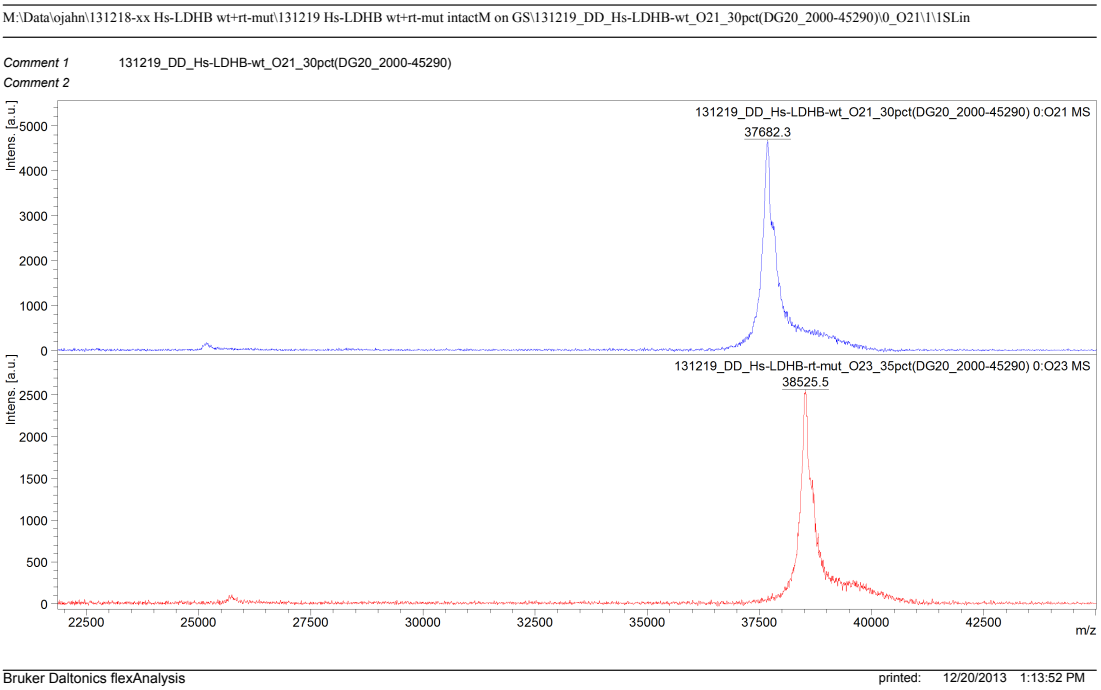


Figure 32: Mass spectroscopy from the constructs FL-LDHB and FL-LDHB-[UGG] that where expressed in *E. coli* and purified with imidazole gradient purifier. The difference between the constructs though only a six amino acids is obvious. The analysis was carried out by Dr. Olaf Jahn, MPI for experimental medicine.

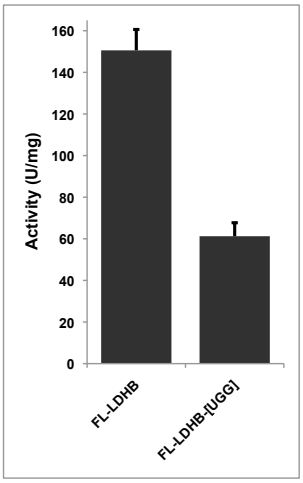


Figure 33: The enzymatic activity in U/mg of FL-LDHB and the 100 % readthrough mutant FL-LDHB-[UGG] was measured in central lab of UMG Göttingen and is shown as the weighted mean of two sets of measurement from 2014-01-22 and 2014-02-29 with each $n = 10$. Interference by stability problems due to residual imidazole in the purified LDHB constructs. Error bars, SD.

The activity of FL-LDHB-[UGG] appeared to be significantly lower than the enzymatic activity of FL-LDHB ($p = 1E - 19$, two sided, unpaired, equal variance Student's t-test), see

Figure 33 on the preceding page. The striking difference in activity might be a result of the extension that is appended to the parental protein, however, it is also possible that the imidazole used in the gradient to purify the protein, the temperature it was stored at, or other factors destabilised the protein and thereby disturbed the activity analysis of LDHB and LDHBx. Thus a second purification process and further experiments are required.

4 Discussion

The study presented here was undertaken to detect and characterise physiological translational readthrough (TR) and functional translational readthrough (FTR) in *Homo sapiens*. The focus was set on gain or alteration of function by FTR for proteins in form of dual localisation due to hidden peroxisomal targeting signals 1 (PTS1). This particular focus was motivated by findings of TR driven dual localisation in fungi (Freitag et al. 2012). The results of this study can be seen against a background of interesting findings and approaches on TR and FTR in various organisms including *Homo sapiens*. In general viruses are known to use TR to maximise their coding capacity (Atkins et al. 1990). System biology approaches have been employed to detect TR and analyse stop codon environments in yeast (Namy et al. 2003; Williams et al. 2004) and *Drosophila melanogaster* (Jungreis et al. 2011). In mammals the first TR genes were identified by serendipity (Geller and Rich 1980; Chittum et al. 1998; Yamaguchi et al. 2012), though recently systems biology approaches have been used to reveal TR in mammals.

A genome wide *in silico* scan for translational readthrough and the design of a high TR consensus

The database scan of the human genome (Ensembl database), the extraction of 42,000 unique stop codon contexts (SCCs) and their ranking with respect to their readthrough propensity prediction was an iterative process. For this the *in silico* regression model was utilised (Schueren et al. 2014). The model was based on a training data set of sequences and their respective experimental TR values. The first model (LIN) was trained with 66 sequences of which 65 originate from human disease related nonsense mutations (Floquet et al. 2012).
The data base was expanded for a second model (LINiter) using TR values acquired through a dual reporter assay. The shortness of the dual reporter constructs excluded mRNA secondary structures that might influence TR. The selected genes of which the dual reporter constructs were derived comprised SCCs with high, intermediate and low RTPs (based on LIN model). The TR of constructs derived from genes LENG1, ZNF574, EDEM3, EDN1, and PPP1R3F exceeded background levels significantly ($p<0.03$). Feature selection (fs) led to reduction of considered nucleotide positions and to refined models (LINfs5, LINfs3) (Schueren et al. 2014). These models consider 5 and 3 nucleotide positions respectively. As the reduction of positions taken into account led to higher correlation between experimental TR values and predicted RTP scores, it can be concluded that the nucleotide positions influence TR non-linearly. A larger database of experimental TR values is needed to set up a non-linear *in silico* model (Schueren et al. 2014), which could model TR more accurately. A high throughput assay employed to measure TR values is advisable in case non-linear approximative models are to be designed.

The non-linear contribution of certain nucleotide positions to RTP scores is a precondition for the consensus we rationally derived for context-driven high TR in *Homo sapiens* (Schueren et al. 2014). This consensus reads UGA CUA (G) (stop codon underlined). 38 genes comprising the consensus UGA CUA plus 19 genes comprising the consensus UGA CUA G are listed in the database created by the *in silico* regression model (Schueren et al. 2014). So far only five of these genes namely *AQP4*, *LDHB*, *MDH1*, *OPRK1*, and *OPRL1* are characterised in terms of TR (Dunn et al. 2013; Schueren et al. 2014; Loughran et al. 2014; Stiebler et al. 2014).

The consensus we derived differs from the motifs found for some viruses (Skuzeski et al. 1991) and yeast (Namy et al. 2001), however, it reminds of the motif found for alphavirus SINV, which is UGA/UAA CGG/CUA (stop codon underlined) (Li and Rice 1993; Beier and Grimm 2001; Firth et al. 2011).

Besides our *in silico* regression model used to predict TR candidates in the human transcriptome, two complementary systems biology approaches to TR (partially in mammals) have been employed. Phylogenetic approaches create a comparative metric using alignments of several species and evaluate thereby the coding potential of sequences (Lin et al. 2011). Thus also the coding potential of sequences downstream of annotated stop codons can be estimated. Still verification of TR candidates requires experiments. In total this approach was used to detect 283 putative TR genes in *Drosophila melanogaster* (Lin et al. 2007; Jungreis et al. 2011), and some of them were confirmed as TR events using transgenic flies and mass spectrometry (Jungreis et al. 2011). Ribosome profiling (Ingolia et al. 2009, 2011) combines next generation deep sequencing with the fact that translating ribosomes protect a fragment of mRNA from digestion by nucleases (Takanami et al. 1965; Steitz 1969). This allows the detection of recoding events including TR, by detecting sequences downstream of stop codons that show translating ribosomes. The systems biology approach ribosome profiling revealed 350 TR events in *Drosophila melanogaster* embryos and S2 cells, derived from this organism (Jungreis et al. 2011). Remarkably, the approach also identified putative TR genes in yeast and human skin fibroblasts (Dunn et al. 2013). Indeed the methods appear to be complementary and show little redundancy in their results (Figure 34 on the next page), which originates from their distinct angles used to approach TR. Phylogenetic approaches concentrate on evolutionary conserved TR events, our *in silico* model detects only context-driven TR and therefore overlooks other forms of TR and recoding events that occur downstream the stop codon. Both methods can detect TR genes, which are expressed at low levels or TR restricted to few tissues. Ribosome profiling on the other hand is a powerful experimental method and is less focused in its detection of recoding events, however, it might miss short TR extensions or TR that occurs only in few tissues. Ribosome profiling is possibly unable to differentiate between TR and low level re-initiation of translation in 3'UTRs. Additional experiments are required for all three methods to verify TR events.

Figure 34 on the following page gives an overview on systems biology approaches used to

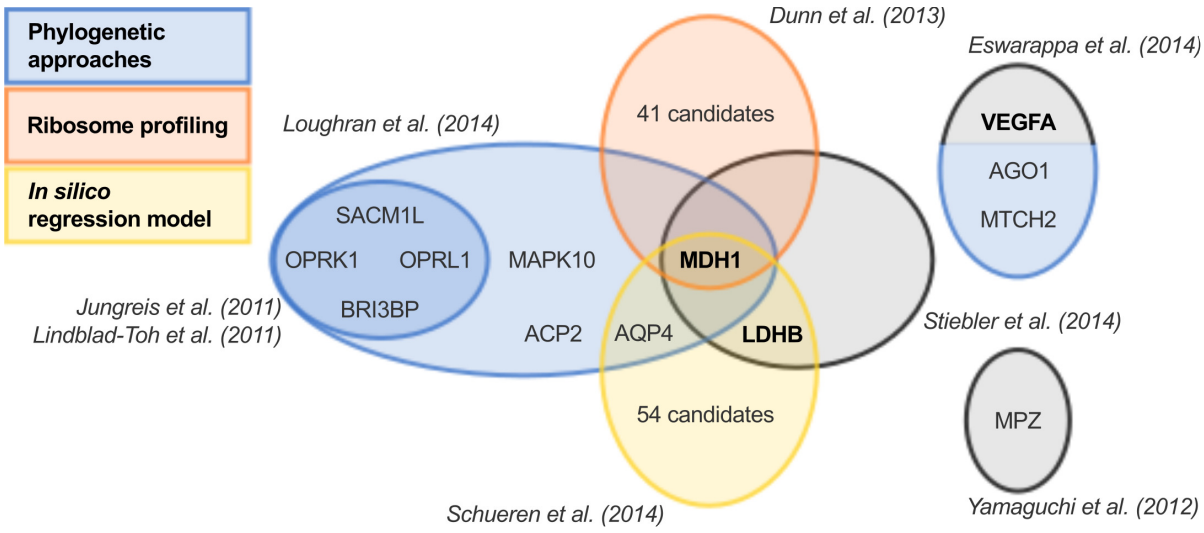


Figure 34: Systems biology approaches to translational readthrough in mammals. Readthrough genes have identified with varying levels of experimental confirmation. Gene symbols of gene products known to undergo functional translational readthrough (FTR) are depicted in bold. Circle sizes do not correspond to number of analysed genes. Black cycles refer to other than systems biology approaches.

detect TR in mammals and recent findings of TR and FTR. While TR of the rabbit beta-globin was detected 35 years ago (Geller and Rich 1980), recently the MPZ gene (myelin protein zero P0) was found to be expressed as L-MPZ via stop suppression. MPZ is displayed here because its readthrough form is probably a case of FTR in mammals. The appended extension of L-MPZ contains antigenic sites for neuropathy-associated antibodies, and may be involved in adhesion processes of PNS myelin (Yamaguchi et al. 2012). Phylogenetic approaches identified 10 possible TR genes (Lindblad-Toh et al. 2011; Jungreis et al. 2011; Eswarappa et al. 2014; Loughran et al. 2014), of these candidates *OPRK1*, *OPRL1*, *AQP4*, *MAPK10*, *MDH1*, *MTCH2*, and *AGO1* were confirmed as TR genes experimentally (Loughran et al. 2014; Eswarappa et al. 2014; Schueren et al. 2014). 42 TR events were identified using ribosome profiling (Dunn et al. 2013). Our *in silico* regression model predicted 57 TR genes based on the immediate environment of their stop codons, three of these (AQP4, LDHB, and MDH1) were confirmed so far, all three exhibit the consensus UGA CUA G (stop codon underlined) (Schueren et al. 2014). The human MDH1 was independently studied in four of the recent studies and thereby is the most robust of the new TR genes (Dunn et al. 2013; Loughran et al. 2014; Schueren et al. 2014; Stiebler et al. 2014).

To detect FTR the search for functional domains or signals in the extensions downstream the annotated stop codon was focused on PTS1 using an algorithm (Schueren et al. 2014). The gene database was sorted according to the product of positively scaled RTP (RTP^+) and PTS1 posterior probability, thus $RTP^+ \times PTS1$. We identified LDHB at position 1 out of 42,069 genes (Schueren et al. 2014). The gene of the related enzyme MDH1 was found at position 175, its score at this position is only 27% of LDHB's product score. The general distribution of the product score hints that only few genes possess both high TR scores and PTS1 posterior

probabilities as the product score drops by 50% within the first 40 candidates (Figure 12 on page 36). However, this does not exclude these candidates (rank > 40) from being targeted to peroxisomes. The two enzymes LDH and MDH are well known and sparked my interest, therefore I characterised the TR of LDHB and MDH1, their response to treatment with aminoglycoside geneticin, and partly functional aspects of their appended extensions.

Both LDHB and MDH1 exhibit TR exceeding background levels in the dual reporter assay, while the latter shows roughly 1.9 times higher TR than the former. MDH1 displays a higher response (IF=4.51) to geneticin treatment as well. Full length constructs were analysed by Western blotting and showed TR, which could be elevated by aminoglycoside treatment (Schueren et al. 2014). My analysis of the influence of the stop codons themselves and nucleotides at positions +4 to +6, the positions with the highest influence as found by feature selection, confirmed the consensus UGA CUA found *in silico* for high context dependent TR in *Homo sapiens* (Schueren et al. 2014). Additionally the analysis confirmed the order of influence of U>G on TR in position -6 as predicted by the *in silico* model (LINfs5). The focus of the analysis on context driven TR is justified by the lack of conserved secondary structures of mRNA (i.e. position -10 to the 2^{nd} in frame stop codon). Distal elements downstream the conserved extension might exist, however, functional distal elements would present as conserved sequences and are thereby unikely.

Evidence was found indicating that TR can undergo regulation due to developmental state of the organism. This was observed for genes in *Drosophila melanogaster*, compare Paragraph 1.3 on page 4, and Paragraph 1.4 on page 7. The TR gene VEGFAx is down-regulated in grade 2 or 3 adenocarcinoma cells of colon (human xenografttumor in mice) (Eswarappa et al. 2014). In order test if TR might be differentially regulated in different cell types as well, TR of LDHB was assessed in several cell lines (HeLa, COS-7, HEK, U118, U373, and HT1080). Translational readthrough of MDH1 was observed in cell lines HeLa, HEK, U118, U373, and HT1080. In general this analysis suggests that TR of LDHB is not restricted to one tissue and there might exist a varying regulation of its TR in different tissues. The comparison of TR rates between the cell lines exhibited higher TR in U118 glioblastoma cells for LDHB and MDH1. The high TR in the glioblastoma cell line hinted that TR rates might be elevated in cancer cell lines, possibly because cancer cell lines have less regulated and controlled metabolism and protein synthesis. Also glial cells might require higher levels of LDH in peroxisomes as it might be involved in neuronal/glial lactate metabolism (Schueren et al. 2014). Therefore TR of MDH1 and LDHB was tested in another glioblastoma cell line (U373), the constructs did not show elevated TR rates in U373 compared to TR in HeLa cells. Additionally TR of three further constructs (pDRVL-ZNF574/EDEM3/IRAK3) was measured in U118 and four other cell lines (incl. U373), the set of constructs did not show significantly elevated TR rates in U118 in comparison to other cell lines. Thus it can be concluded that firstly it is probably neither the cancer cell's metabolism nor the specific glial metabolism that explains the effect of elevated TR of LDHB and MDH1 in U118 glioblastoma cells as it wasn't observed for U373

cells, and secondly that the effect observed for LDHB and MDH1 does not extend to other TR genes (pDRVL-ZNF574/EDEM3/IRAK3).
I tried to gather information on the mechanism of context driven TR and its most influential elements. The stop codon itself is highly important and the effect it exerts on TR was isolated. The SCC did influence expression of luciferase in the dual reporter assay even in absence of the stop codon, though this influence is not correlated to TR. No effect of codon frequencies (and thereby tRNA abundance) on TR was observedfor the tested constructs. This suggests that the influence of SCC (especially pos. +4 to +6) on TR might be mediated directly by the nucleotides following the stop codon, which might possibly work via conformational changes in the ribosome invoked by the nucleotides.

Functional translational readthrough

Beside TR the focus of this study is the physiological FTR found for LDHB. LDHB is known to reside in peroxisomes, however, it was not known how it reached peroxisomal matrix. Qualitatively it was shown that LDHBx (extended LDHB) resides in peroxisomes due to TR and its appended PTS1 using overexpression of an N-terminally tagged LDHBx. The importance of TR and PTS1 for the targeting process was tested and confirmed using mutant constructs (Schueren et al. 2014). Interestingly an analysis of LDHB orthologs in vertebrates revealed strict conservation of the hidden PTS1 in mammals (Schueren et al. 2014), which highlights the importance of TR in an evolutionary perspective. Using untransfected HeLa cells, human skin fibroblasts, COS-7 cells and glioblastoma cells (U118) it was shown that endogenous LDHBx exhibits dual localisation in cytosol and peroxisomes of these cells. I present quantitative results that suggest the extent to which endogenous LDHBx occurs in peroxisomes (2.85 %) is in agreement with pDRVL-LDHB's TR rate (1.55 %) measured in the dual reporter assay using shortened constructs. Due to time limitation I could not test the effect of high (F)TR in U118 cells for endogenous or transfected full length LDHB as it was observed for it's pDRVL construct. Endogenous LDHBx in HeLa cells exhibits a response to G418 treatment (IF= 1.86 fold) that resembles the induction factor of the pDRVL-LDHB construct (IF= 2.82 fold). Interestingly, this hints a new effect of treatment with antibiotics (aminoglycosides), i.e. a treatment with aminoglycosides or their derivatives might raise LDH levels in peroxisomes of human cells.
High correlation coefficients support the visual assumption of co-localisation between LDHBx and peroxisomal marker PEX-14. The evidence that TR of LDHB exists in several human cell types and that LDHBx is expressed in various human cell types hints that LDHBx is possibly expressed in all human tissues which generally express LDHB. The measured TR of 1.55 % (dual reporter assay) and 2.85 % (fluorescence imaging) is in agreement with the 1.2 % activity of LDH found in peroxisomes (see Paragraph 1.6 on page 11) and possibly ensures a cytosolic concentration of LDH in peroxisomes considering that peroxisomes comprise 2 % of a cell's

volume (Moody and Reddy 1976). It indicates as well that TR of LDHB is the only mechanism that targets the subunits of LDH to peroxisomes and thereby results in peroxisomal LDH. The targeting of LDHBx being dependent on TR and the hidden PTS1, LDHBx belongs to the first examples of this kind of expansion of our genome's coding potential on a post-transcriptional level.

Interestingly, the second subunit LDHA of lactate dehydrogenase was also found in peroxisomes but does not comprise a hidden PTS1. As peroxisomes are capable of importing oligomers it was tested if LDHBx imports LDHA via a piggy-pack mechanism (Schueren et al. 2014). This seems possible as the structure of homotetramer LDHBx does show the PTS1 signal at the margin of the oligomer, see Figure 38 on page 76. Additionally, the oligomerisation is not disturbed by the extension (Schueren et al. 2014). Expression of a fluorophore tagged LDHA did result in cytosolic localisation, however co-expression of a tagged LDHA and LDHBx led to presence of LDHA in peroxisomes (Schueren et al. 2014). The signals of fluorophores attached to LDHB, LDHA and PEX-14 reveal correlation coefficients of $r \geq 0.85$, which support the visual appearance of co-localisation between peroxisomal marker and the two subunits of LDH. These results explain the long known presence of LDH subunits LDHB and LDHA in peroxisomes (McGroarty et al. 1974; Volkl and Fahimi 1985; Baumgart et al. 1996; Gronemeyer et al. 2013). They also suggest that peroxisomes contain up to 9 new isoforms of LDH, though those containing a maximum of one LDHBx subunit are most likely to occur as LDHBx is expressed at a low level. Interestingly, the isoform $LDHA_3B$ was found in peroxisomes to a higher percentage than in the cytosolic fraction of rat hepatocytes Baumgart et al. (1996). The isoform $LDHA_4$ was assumed to be found in peroxisomes due to cytosolic contamination (Baumgart et al. 1996). However it is not ruled out that this isoform does exist in peroxisomal matrix (Baumgart et al. 1996), especially as LDHA was found to be associated to peroxisomes (Gronemeyer et al. 2013) and a piggy-pack import mechanism for LDHA was revealed (Schueren et al. 2014). The question arises whether the oligomers might be reorganised in the peroxisome and therefore also isoforms of LDH without any LDHBx subunit (e.g. $LDHA_4$) exist in peroxisomes.

The function of peroxisomal LDH is still a matter of speculation. Peroxisomes exhibit several metabolic processes including fatty acid oxidation, which leads to production of NADH, which in turn needs to be re-oxidated to NAD^+ for fatty acid oxidation to continue. The balance of the peroxisomal NAD^+/NADH pool requires a shuttle system as peroxisomal membranes are impermeable to NAD^+/NADH (van Roermund et al. 1995). This system could involve lactate/pyruvate shuttling either directly (Visser et al. 2007) or via transporters over the peroxisomal membrane (Baumgart et al. 1996; McClelland et al. 2003), either way the system could involve cytosolic and peroxisomal LDH. Given the size of NADH molecules and monocarboxylates it seems logical that the smaller ones are used for the shuttle over the membrane. Peroxisomal LDH might also contribute to peroxisomal glyoxylate metabolism (Schueren et al. 2014). Thus peroxisomal LDH is probably involved in the redox balance of

peroxisomes, allowing thereby the organelle's metabolism to occur.
So far the only other known FTR gene in mammels - besides *LDHB*, and *MDH1* - is *VEGFA*, to which FTR appends an extension of 22 amino acids. The gene undergoes FTR of $7 - 25\,\%$ and the extension changes its function form proangiogenic to antiangiogenic (Eswarappa et al. 2014). The level of VEGFA in cells is strictly regulated, a two-fold increase of its amount is lethal to embryonic cells. In high-grade colon adenocarcinomaa reduced expression levels were found (Eswarappa et al. 2014). The extension of VEGFA is conserved in mammals and contains a sequence, which is known to mediate antiangiogenic activity of VEGFA's alternative splice variant VEGFAb (Harper and Bates 2008). The readthrough form of VEGFA is the first example of stop codon independent and distal element dependent FTR in mammals.
In contrast to VEGFA, *LDHBx* and *MDH1x* are the first examples of context-driven FTR, as their extensions (6, and 18 aa, respectively) are too short to allow extensive conserved mRNA secondary structures, see also Figure 17 on page 42 (Dunn et al. 2013; Stiebler et al. 2014; Schueren et al. 2014; Loughran et al. 2014). MDH1x might fulfil a similar function in peroxisomes as LDHBx, it appends a hidden PTS1 using TR and it targeted to peroxisomes by it (Stiebler et al. 2014).
In general TR or FTR might be an affordable way for an organisms to test new functions of proteins on an evolutionary scale by appending new C-termini, as FTR diverts only a small fraction of the parental protein to the new function. Thereby the parental protein's function is not disturbed. However, it is doubtful that FTR in human is as abundant as it is assumed to occur in e.g. *D. melanogaster* (Jungreis et al. 2011).

Further analysis of LDHBx

In an attempt to characterise the amino acid that is incorporated at the stop codon's position during TR rat tissues were used. The different tissues (heart and skeletal muscle, liver, kidney and fat tissue) were lysed and subjected to immuno precipitation (IP) with the aim to employ a α-LDHB antibody to pull down LDHB residing in cytosol and peroxisomes and to detect endogenous LDHBx using a western blot. Unfortunately the immunoprecipitation did not yet yield sufficient LDHB to detect LDHBx in the samples. The detection of LDHBx and its further characterisation is planned with the help of mass spectrometry.
Overexpressed LDHB and LDHBx were purified using an imidazole gradient and subjected to catalytic activity analysis to test whether the appended extension of LDHBx effects the enzymes activity. The result of this analysis requires verification and retesting is in order as apparently LDHBx shows lower catalytic activity than that of LDHB, however, several disturbing factors were present. Imidazole used for the gradient and still present in the solution of the purified proteins might destabilise the enzymes. Additionally the enzymes were stored at $4\,^{\circ}\mathrm{C}$, a temperature that might not be optimal for LDH and might be a cause of instability, thus

the results presented for LDHBx's activity may be seen as a possibility but are not reliable yet. A purification that does not involve imidazole and new activity analysis that minimise storage time are required.

5 Outlook

Translational readthrough was studied for a long time in lower organisms, now it is apparent that it plays a physiological role in higher organisms as well. The mechanisms of TR are not yet fully understood, though there is new insight into the mechanism of termination of translation (Brown et al. 2015). Even though many TR candidates were found recently, they need still need to be verified. Databases of mass spectrometry experiments might contain information that would allow confirmation of TR candidates. Also flow cytometry experiments might be suitable to analyse TR candidates time efficiently to detect the most promising genes. The possible physiological functions of the extended proteins await characterisation. More specifically the function of peroxisomal LDH and MDH might be hypothesised but is not yet proven, though the function of LDHBx is revealed. It is known now that MDH1 targets to peroxisomes (Stiebler et al. 2014), but it is still of interest to further analyse the gene in terms of the extension's conservation and to test if TR of MDH1 varies with the tissue endogenous MDH1 is expressed in. Even though TR and FTR is probably not as abundant in complex organisms such as *Homo sapiens* as it is said to be in viruses or Drosophila, we just started to uncover the extensome, the set of gene products expressed using TR.

Additionally, there is the chance to utilise TR in medical research and clinical application. Already the induction of TR by aminoglycosides of their derivatives (often less toxic than aminoglycosides) is a strategy to treat genetic disorders caused by premature stop codon (Bidou et al. 2012; Keeling et al. 2014). Results from studies on aminoglycosides describe firstly the problems of the drugs toxicity and thereby the impossibility of long term application and secondly the variable effect in patients, i.e. the extent to which the full length proteins are restored (Keeling et al. 2014). I did observe very variable responses to drug treatment and it might be fruitful to use a high throughput assay to analyse different SCCs in terms of their response to different drugs. A promising experiment could be designed using flow cytometry, to analyse many transfected cell samples efficiently. The data could be used to build a training set for the model we used, this time it would model the SCCs responses to drug treatment to predict the outcome of treatments of premature stop codons in human patients. Such a predictor would allow rational and targeted therapies for patients and avoid subjecting patients with non-responsive SCCs to the *side-effects* of the drugs.

Remarkably, the same SCC we derived as the high TR consensus, was found to reduce the otherwise lethal effect of a premature stop codon in the *LAMA3* gene in a patient with junctional epidermolysis bullosa (Pacho et al. 2011). This exemplifies the clinical potential of the presented findings.

6 Summary

The presented work deals with translational readthrough and functional translational readthrough. Translational readthrough is a phenomenon that is not yet fully understood. In most cases a ribosome stops and detaches from the mRNA after the encounter with a stop codon. However at low rates amino acids are incorporated at termination sites and the ribosome continues translation into the extension downstream the stop codon. The term translational readthrough describes ribosomes that translate into mRNA sequences downstream of a stop codon at a higher ratio than the normal error rate. Functional translational readthrough is defined as translational readthrough that appends extensions to parental proteins that change the function of the resulting protein from its original one. Translational readthrough is employed by viruses and more complex organisms to expand the coding potential of their limited genome, however, also in more complex organisms examples of readthrough and even functional readthrough have been found. The study contains a description of an *in silico* regression model that was used to derive a consensus for nucleotide context based high readthrough in mammals and to detect 57 candidates for translational readthrough in a genome wide scan in a human transcriptome. The results of this genome wide scan for TR are discussed against a background of eight recent studies that applied systems biology approaches (PhyloCSF, ribosome profiling, and *in silico* regression model) to mammalian genomes/transcriptomes.

A dual reporter assay was used to test and verify the stop codon context of some of these 57 transcripts. The results of 15 constructs were used to expand the database used for the model. The resulting database of 42.000 unique stop codon contexts was sorted according to their readthrough potential and a score developed to estimate the probability for hidden PTS1 (peroxisomal targeting signal). The subunit LDHB of lactate dehydrogenase was detected at the first position, exceeding rank two by far. *MDH1*, the gene coding for cytosolic malate dehydrogenase was detected at rank 175. Both enzymes are evolutionary related and were analysed in detail in terms of their translational readthrough. The proteins showed high readthrough rates and responded to treatment with aminoglycosides with elevated readthrough rates. Analysis of readthrough in multiple cell types suggests that the readthrough forms of LDHB (LDHBx) and MDH1 (MDH1x) are expressed in probably all human tissues at varying levels. Their extensions are conserved in mammals, but show no conserved secondary structure. Next to immediate stop codon contexts also elaborate mRNA secondary structures can influence and mediate translational readthrough. The influence of the stop codon and of the stop codon context on TR was tested as well as codon usage frequencies.

The hidden PTS1 found in the extensions of LDHBx was tested. It was found that endogenous LDHBx is imported to peroxisomes by means of translational readthrough and the hidden PTS1. Translational readthrough was quantified and analysed using immunofluorescence imaging and image analysis tools (ImageJ). LDHBx was found to co-import the other subunit LDHA into peroxisomes allowing theoretically nine new isoforms to occur in the peroxisomal

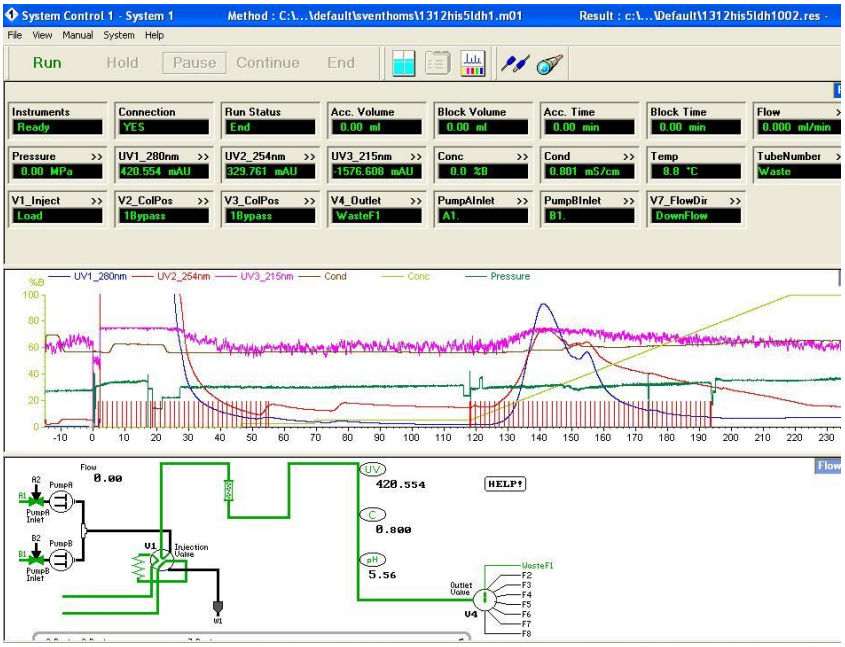
matrix.

The potential of the findings lies in the used methods themselves e.g. the *in silico* regression model, and clinical application in form of drug treatment of genetic disorders caused by premature stop codons. An outline for experiments that could allow the design of a predictor model for the response of stop codon contexts to drug treatment (aminoglycosides and their derivatives) is given.

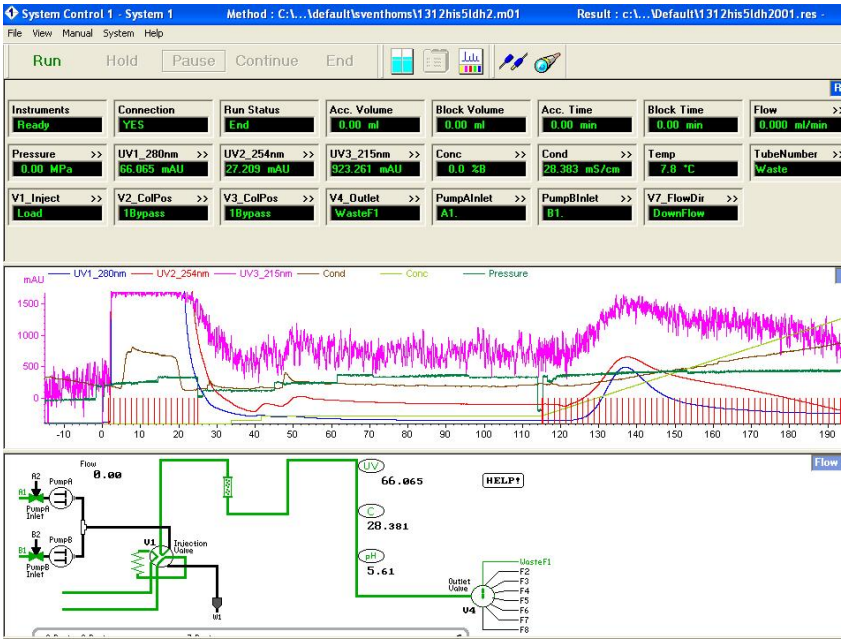
A Appendix

A.1 Appendix - Figures

A.1.1 Purification graphs of full length LDHB and full length LDHB-[UGG]



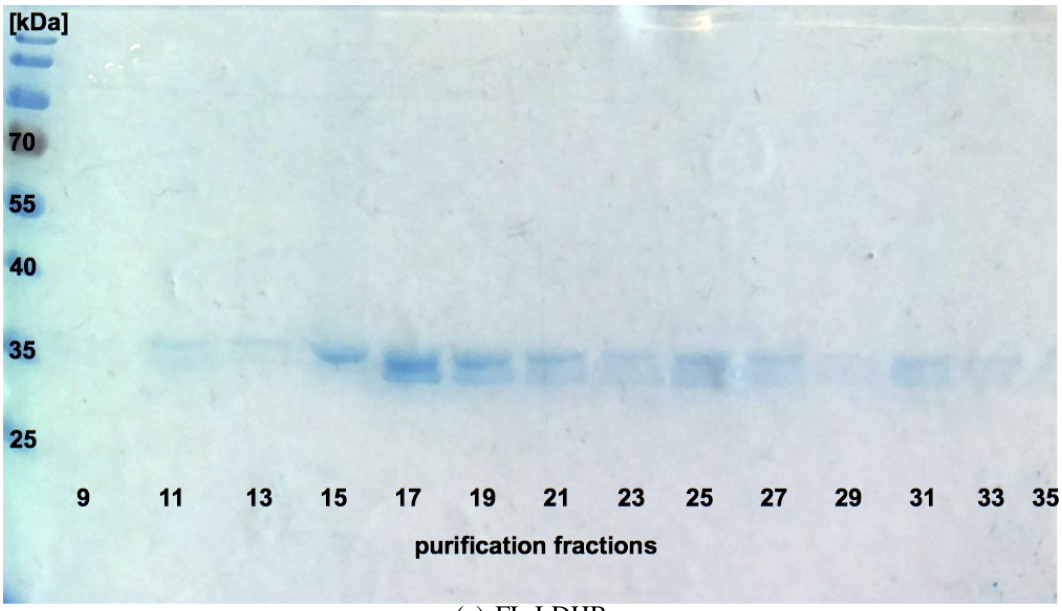
(a) LDHB-WT-Purification



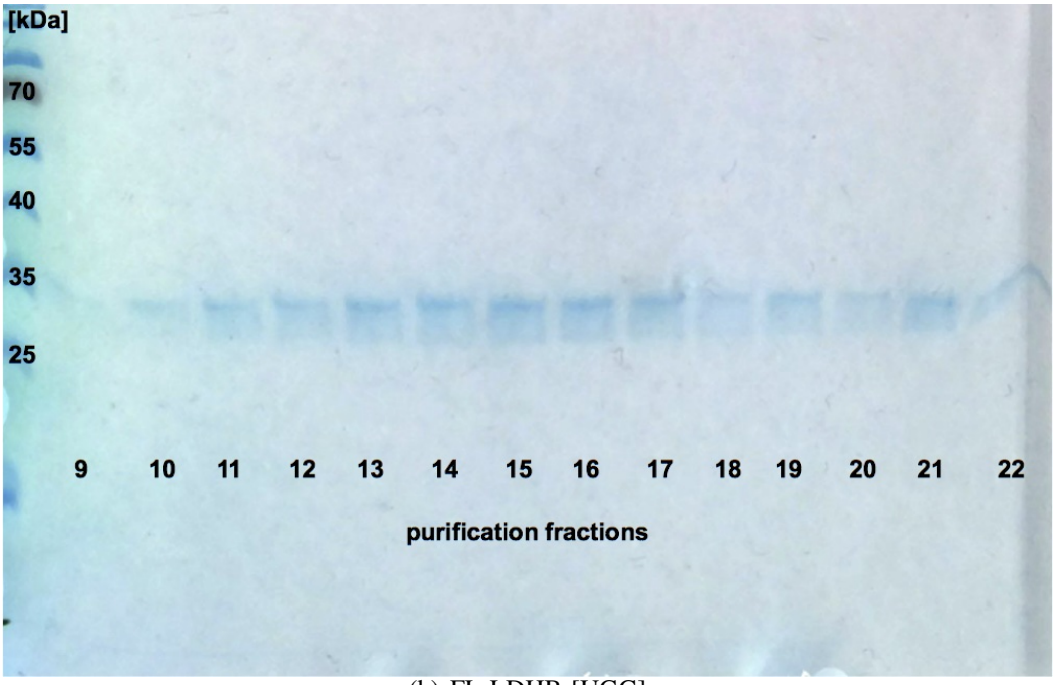
(b) LDHB-[UGG]-Purification

Figure 35: Report of purification of LDHB-WT and LDHB-[UGG]. The purification was carried out with the support of PD Dr. S. Thoms.

A.1.2 Commassie gel with purified fractions of FL-LDHB and FL-LDHB-[UGG]



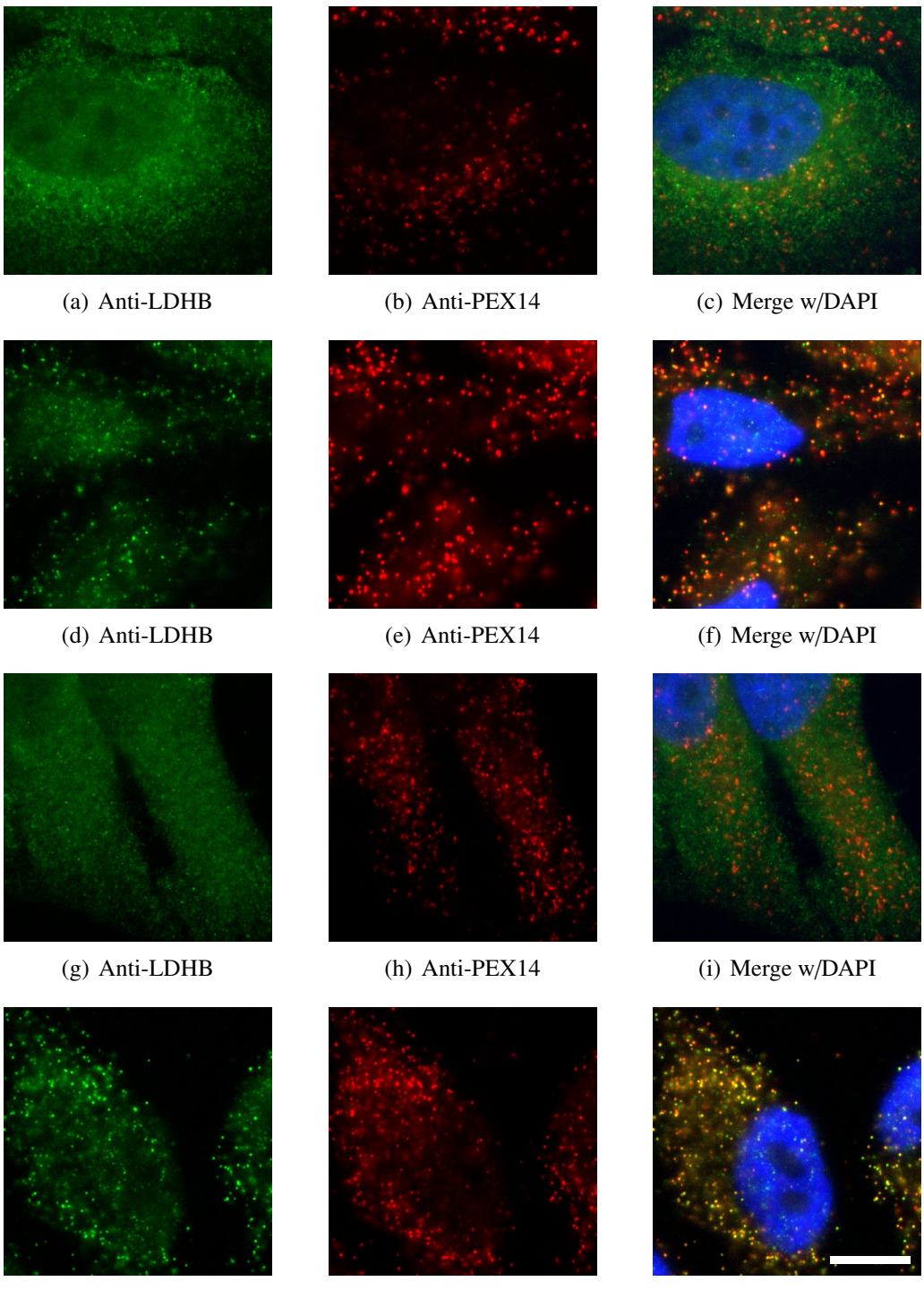
(a) FL-LDHB



(b) FL-LDHB-[UGG]

Figure 36: The purification process of FL-LDHB and FL-LDHB-[UGG] resulted in fractions with different concentrations of protein, to decide which fractions were to be used in further analysis, they were run on a commassie stained SDS gel (12 %). A gene ruler allows an estimation of the protein's weight.

A.1.3 Endogenous LDHB in HeLa cells under various conditions



(a) Anti-LDHB (b) Anti-PEX14 (c) Merge w/DAPI

(d) Anti-LDHB (e) Anti-PEX14 (f) Merge w/DAPI

(g) Anti-LDHB (h) Anti-PEX14 (i) Merge w/DAPI

(j) Anti-LDHB (k) Anti-PEX14 (l) Merge w/DAPI

Figure 37: Endogenous LDHB in wild-type HeLa cells localises to peroxisomes, cells with (**a-c**, **g-i**)/ without (**d-f**, **j-l**) cytosol and with (**g-l**)/ without (**a-f**) drug treatment (G418). Immunofluorescence with secondary antibodies Alexa488-coupled and Cy3-coupled. Co-localisaton: Pearson correlation coefficient $r_{d-f} = 0.85$ and $r_{j-l} = 0.94$. Slides were prepared by C. Dickel. Bar 10 µm.

A.1.4 LDHBx tetramer structure

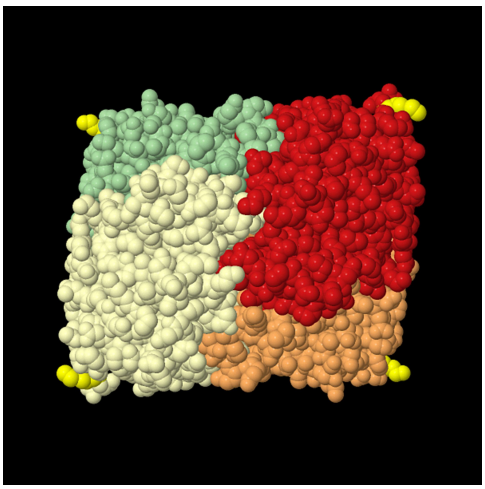


Figure 38: The structure shows four LDHx subunits that form together a homotetramer (Schueren et al. 2014). The C-termini of the subunits are shown to be on the margin of the enzyme and the appended extensions of the LDHBx subunits are displayed in yellow. It is likely that the PTS1 that are located at the very c-terminal end are accessible to PEX5.

A.2 Appendix - Tables

A.2.1 Model weights used for the *in silico* regression model

Base/position	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**4**
A	-0.00041	0.00130	-0.00028	-0.00073	-0.00071	0.00016	-0.00037
C	-0.00105	0.00164	0.00075	-0.00004	0.00133	0.00109	0.00375
G	0.00060	-0.00077	-0.00041	0.00193	-0.00048	0.00043	-0.00156
U	0.00200	-0.00103	0.00108	-0.00002	0.00100	-0.00054	-0.00067
Base/position	**+5**	**+6**	**+7**	**+8**	**+9**	**Stop**	
A	-0.00068	0.00276	-0.00020	0.00105	-0.00081	-0.00026	UAA
C	-0.00097	-0.00026	-0.00062	-0.00017	0.00148	-0.00103	UAG
G	-0.00008	-0.00059	0.00245	-0.00058	0.00014	0.00243	UGA
U	0.00287	-0.00076	-0.00049	0.00084	0.00032		

Table 13: The regression factors listed here are model weights of the LINiter model. The training data set used for the model includes the experimental TR values measured with the dual reporter assay, see Paragraph 3.3 on page 33. The RTP score is a sum of weights representing the stop codon and the corresponding nucleotides of the SCC surrounding the termination site. The weights were computed by the model build by T. Lingner, for more information please refer to Table 2 in Schueren et al. (2014).

A.2.2 Human high TR candidates comprising consensus UGA CUA (G)

Genes with consensus UGA CUA	Genes with consensus UGA CUA G
APH1A	AL121581.1
BIVM	AQP4
C12orf74	CDH23
C15orf57	CGGBP1
CCNL2	DIO2
DUS2	DUS4L
DYRK1A	EPT1
ELK1	HCLS1
EMC8	KCNB2
ESRRG	LDHB
FOXP1	MAPK10
FUZ	MDH1
GSKIP	MS4A5
HPS6	OPRK1
KLK15	OPRL1
KLRB1	PHF19
KRTAP13-3	SPATA32
LMAN1	TMEM86B
MAPK8	VDR
MED18	
METTL4	
NPS	
NUBP2	
OPRM1	
OR2Z1	
PAN2	
PLA2G10	
PLCE1	
PRRG4	
RGS13	
SHCBP1	
SKA2	
STON1-GTF2A1L	
STRIP2	
TFAP4	
TTC4	
ZNF548	
ZPLD1	

Table 14: 38 human high TR candidates comprising consensus UGA CUA and 19 human high TR candidates comprising consensus UGA CUA G are listed above. The gene symbols were extracted from the database created by the *in silico* regression model (Schueren et al. 2014).

A.2.3 Results of dual reporter assay of genomic candidates

PST#	pDRVL-HGNC	TR	SD	Induce factor	SD	LIN	LINiter	LINfs5	LINfs3
1435	pDRVL-MDH1	2.91	0.147	4.51	0.32	0.23	0.816	0.917	0.939
1385	pDRVL-LDHB	1.55	0.087	2.82	0.31	0.27	0.609	0.816	0.939
1422	pDRVL-EDEM3	0.66	0.027	8.15	0.44	0.15	0.426	0.631	0.655
1418	pDRVL-LENG1	0.26	0.009	6.03	0.29	0.10	0.372	0.472	0.714
1384	pDRVL-ZNF574	0.31	0.020	21.46	2.19	0.27	0.367	0.465	0.714
1419	pDRVL-PRDM10	0.13	0.005	40.38	5.33	0.24	0.276	0.424	0.655
1420	pDRVL-FBXL20	0.10	0.006	29.51	3.80	0.10	0.245	0.394	0.646
1423	pDRVL-EDN1	0.25	0.008	18.41	2.16	0.17	0.230	0.238	0.330
1430	pDRVL-LEPRE1	0.27	0.010	39.23	4.47	0.08	0.209	0.260	0.330
1421	pDRVL-THG1L	0.15	0.006	32.20	2.74	0.22	0.168	0.264	0.394
1424	pDRVL-UBQLN1	0.13	0.009	17.23	2.44	0.02	0.125	0.047	0.107
1425	pDRVL-IRAK3	0.10	0.007	9.32	1.22	0.14	0.090	0.054	0.204
1426	pDRVL-SLC3A1	0.18	0.008	11.89	1.54	0.02	0.029	0.109	0.070
1437	pDRVL-VASN	0.12	0.004	4.14	0.50	0.01	0.024	0.002	-0.112
1387	pDRVL-PPP1R3F	0.18	0.009	3.25	0.41	-0.21	-0.024	-0.076	-0.215

Table 15: A list of the 15 TR candidates derived from a scan of the human genome, it comprises experimental TR values [%], experimentally assessed IF (fold) values, and RTP scores predicted by *in silico* regression models: LIN, LINiter, LINfs5, and LINfs3. The construct pDRVL-VASN was used to derive a negative control. SD, standard deviation.

A.2.4 Codon frequencies in the human genome

Table 16: Codon frequencies in the human genome (Genscript 2014). The columns *triplett* show the nucleotides of the codons, the colums *abbr.* list the single letter abbreviation of the resulting amino acids. The columns *fration* displays the relative abundance of these codons in coding regions for each amino acid. The column *frequency* lists the resulting frequencies [10^{-3}].

triplett	abbr.	fraction	frequency	triplett	abbr.	fraction	frequency	triplett	abbr.	fraction	frequency	triplett	abbr.	fraction	frequency
TTT	F	0.45	16.9	TCT	S	0.18	14.6	TAT	Y	0.43	12	TGT	C	0.45	9.9
TTC	F	0.55	20.4	TCC	S	0.22	17.4	TAC	Y	0.57	15.6	TGC	C	0.55	12.2
TTA	L	0.07	7.2	TCA	S	0.15	11.7	TAA	*	0.28	0.7	TGA	*	0.52	1.3
TTG	L	0.13	12.6	TCG	S	0.06	4.5	TAG	*	0.2	0.5	TGG	W	1	12.8
CTT	L	0.13	12.8	CCT	P	0.28	17.3	CAT	H	0.41	10.4	CGT	R	0.08	4.7
CTC	L	0.2	19.4	CCC	P	0.33	20	CAC	H	0.59	14.9	CGC	R	0.19	10.9
CTA	L	0.07	6.9	CCA	P	0.27	16.7	CAA	Q	0.25	11.8	CGA	R	0.11	6.3
CTG	L	0.41	40.3	CCG	P	0.11	7	CAG	Q	0.75	34.6	CGG	R	0.21	11.9
ATT	I	0.36	15.7	ACT	T	0.24	12.8	AAT	N	0.46	16.7	AGT	S	0.15	11.9
ATC	I	0.48	21.4	ACC	T	0.36	19.2	AAC	N	0.54	19.5	AGC	S	0.24	19.4
ATA	I	0.16	7.1	ACA	T	0.28	14.8	AAA	K	0.42	24	AGA	R	0.2	11.5
ATG	M	1	22.3	ACG	T	0.12	6.2	AAG	K	0.58	32.9	AGG	R	0.2	11.4
GTT	V	0.18	10.9	GCT	A	0.26	18.6	GAT	D	0.46	22.3	GGT	G	0.16	10.8
GTC	V	0.24	14.6	GCC	A	0.4	28.5	GAC	D	0.54	26	GGC	G	0.34	22.8
GTA	V	0.11	7	GCA	A	0.23	16	GAA	E	0.42	29	GGA	G	0.25	16.3
GTG	V	0.47	28.9	GCG	A	0.11	7.6	GAG	E	0.58	40.8	GGG	G	0.25	16.4

A.2.5 List of plasmids used in this study

Table 17: List of plasmids used to for genomic candidates (pDRVL-X), position +4, +5, +6 mutants, pDRVL-X-[UGG] constructs, as well as full length constructs for immunofluorescence experiments (IF) and expression in bacteria.

Plasmid No. PST	plasmid name/ genotype	Source
genomic cand		
1360	pDRVL	R. George, F. Schüren
1384	pDRVL-ZNF574	R. George
1385	pDRVL-LDHB	R. George
1387	pDRVL-PPP1R3F	R. George
1393	pDRVL-LDHB-[UGG]	R. George
1394	pDRVL-LDHB-[UGA U]	R. George
1395	pDRVL-LDHB-[UAA]	R. George
1396	pDRVL-LDHB-[UAA U]	R. George
1418	pDRVL-LENG1	F. Schüren
1419	pDRVL-PRDM10	F. Schüren
1420	pDRVL-FBXL20	F. Schüren
1421	pDRVL-THG1L	F. Schüren
1422	pDRVL-EDEM3	F. Schüren
1423	pDRVL-EDN1	F. Schüren
1424	pDRVL-UBQLN1	F. Schüren
1425	pDRVL-IRAK3	F. Schüren
1426	pDRVL-SLC3A1	F. Schüren
1430	pDRVL-LEPRE1	F. Schüren
1435	pDRVL-MDH1	R. George, F. Schüren
1437	pDRVL-VASN	F. Schüren
1444	pDRVL-VASN-doubleTAA	F. Schüren
Pos 4, 5, 6		
1466	pDRVL-LDHB-[UAG CUA]	T. Wilke, F. Schüren
1467	pDRVL-LDHB-[UGA UUA]	T. Wilke, F. Schüren
1468	pDRVL-LDHB-[UGA GUA]	T. Wilke, F. Schüren
1469	pDRVL-LDHB-[UGA CAA]	T. Wilke, F. Schüren
1470	pDRVL-LDHB-[UGA CCA]	T. Wilke, F. Schüren
1472	pDRVL-LDHB-[UGA CUG]	T. Wilke, F. Schüren

continued on next page...

Table 17: *continued*

Plasmid No. PST	plasmid name/ genotype	Source
1473	pDRVL-MDH1-[UAA CUA]	T. Wilke, F. Schüren
1475	pDRVL-MDH1-[UAA UUA]	T. Wilke, F. Schüren
1476	pDRVL-MDH1-[UGA UUA]	T. Wilke, F. Schüren
1477	pDRVL-MDH1-[UGA GUA]	T. Wilke, F. Schüren
1479	pDRVL-MDH1-[UGA CCA]	T. Wilke, F. Schüren
1480	pDRVL-MDH1-[UGA CUU]	T. Wilke, F. Schüren
1481	pDRVL-MDH1-[UGA CUG]	T. Wilke, F. Schüren
1502	pDRVL-MDH1-[UGA AUA]	J. Hofhuis
SCC-Memory		
1445	pDRVL-LENG1-[UGG]	T. Wilke, F. Schüren
1446	pDRVL-PRDM10-[UGG]	T. Wilke, F. Schüren
1447	pDRVL-FBXL20-[UGG]	T. Wilke, F. Schüren
1448	pDRVL-THG1L-[UGG]	T. Wilke, F. Schüren
1449	pDRVL-EDEM3-[UGG]	T. Wilke, F. Schüren
1450	pDRVL-EDN1-[UGG]	T. Wilke, F. Schüren
1451	pDRVL-LEPRE1-[UGG]	T. Wilke, F. Schüren
1452	pDRVL-UBQLN1-[UGG]	T. Wilke, F. Schüren
1453	pDRVL-IRAK3-[UGG]	T. Wilke, F. Schüren
1454	pDRVL-SLC3A1-[UGG]	T. Wilke, F. Schüren
1455	pDRVL-MDH1-[UGG]	T. Wilke, F. Schüren
IF		
1434	pEYFP-C1-LDHA	R. George
1440	pECFP-C1-LDHBx-[UGG]	R. George
FL-Construct-Purification		
884	pET41a(+)	Merck Millipore, USA
1365	pOTB7-LDHB	Laboratory RG: S. Thoms
1389	pEYFP-C1-LDHB-[UGG]	R. George
1431	FL-LDHB	F. Schüren
1432	FL-LDHB-[UGG]	F. Schüren

A.2.6 List of oligonucleotides used in this study

Table 18: List of oligonucleotides used to construct plasmids for genomic candidates, position +4, +5, +6 mutants, SCC memory constructs, as well as full length constructs for immunofluorescence experiments (IF) and expression in bacteria.

OST No.	Name	Sequence 5' - 3'
genomic cand		
963	DR-MCS for	TCGAGCGGTCACCATCGATTCCGGACCG-TACGG
964	DR-MCS rev	TCGACCGTACGGTCCGGAATCGATGGTGAC-CGC
1081	DR-ZNF574 for	GTCACCATATCAGTGGCTGACTCTGCCCGAT
1082	DR-ZNF574 rev	CCGGATCGGGCAGAGTCAGCCACTGATATG
1083	DR-LDHB for	GTCACCAAAAAGACCTGTGACTAGTGAGCTT
1084	DR-LDHB rev	CCGGAAGCTCACTAGTCACAGGTCTTTTTG
1086	DR-PPP1R3F for	GTCACCATTGGTTCTCATAGGCTCTGCTTGT
1087	DR-PPP1R3F rev	CCGGACAAGCAGTGCCTATGAGAACCAATG
1123	DR-LDHB [UGG] for	GTCACCAAAAAGACCTGTGGCTAGTGAGCTT
1124	DR-LDHB [UGG] rev	CCGGAAGCTCACTAGCCACAGGTCTTTTTG
1117	DR-LDHB [UGA U] for	GTCACCAAAAAGACCTGTGATTAGTGAGCTT
1118	DR-LDHB [UGA U] rev	CCGGAAGCTCACTAATCACAGGTCTTTTTG
1119	DR-LDHB [UAA] for	GTCACCAAAAAGACCTGTAACTAGTGAGCTT
1120	DR-LDHB [UAA] rev	CCGGAAGCTCACTAGTTACAGGTCTTTTTG
1121	DR-LDHB [UAA U] for	GTCACCAAAAAGACCTGTAATTAGTGAGCTT
1122	DR-LDHB [UAA U] rev	CCGGAAGCTCACTAATTACAGGTCTTTTTG
1144	DR-LENG1 for	GTCACCGCCTTACTCACTGACTCCTGAGGGT
1145	DR-LENG1 rev	CCGGACCCTCAGGAGTCAGTGAGTAAGGCG
1148	DR-PRDM10 for	GTCACCGCACCAAACCATGACTTCCACCCTT
1149	DR-PRDM10 rev	CCGGAAGGGTGGAAGTCATGGTTTGGTGCG
1150	DR-FBXL20 for	GTCACCGCATCATCCTATGACAATGGAGGTT
1151	DR-FBXL20 rev	CCGGAACCTCCATTGTCATAGGATGATGCG
1152	DR-THG1L for	GTCACCGAGCCAGGCTTTGACGGAAGAGTCT
1153	DR-THG1L rev	CCGGAGACTCTTCCGTCAAAGCCTGGCTCG
1154	DR-EDEM3 for	GTCACCGGGATGAGCTATGACTTGCTAAACT
1155	DR-EDEM3 rev	CCGGAGTTTAGCAAGTCATAGCTCATCCCG
1156	DR-EDN1 for	GTCACCGAGCACATTGGTGACAGACCTTCGT

continued on next page...

Table 18: *continued*

OST No.	Name	Sequence 5' - 3'
1157	DR-EDN1 rev	CCGGACGAAGGTCTGTCACCAATGTGCTCG
1160	DR-UBQLN1 for	GTCACCGCCAGCCATCATAGCAGCATTTCTT
1161	DR-UBQLN1 rev	CCGGAAGAAATGCTGCTATGATGGCTGGCG
1162	DR-IRAK3 for	GTCACCGCAAAAAAGAATAAATTCTACCAGT
1163	DR-IRAK3 rev	CCGGACTGGTAGAATTTATTCTTTTTTGCG
1164	DR-SLC3A1 for	GTCACCGTACCTCGTGTTAGGCACCTTTATT
1165	DR-SLC3A1 rev	CCGGAATAAAGGTGCCTAACACGAGGTACG
1158	DR-LEPRE1 for	GTCACCGGGATGAGCTATGACAGCGTCCAGT
1159	DR-LEPRE1 rev	CCGGACTGGACGCTGTCATAGCTCATCCCG
1190	DR-MDH1 for	GTCACCGTTCCTCTGCCTGACTAGACAATGT
1191	DR-MDH1 rev	CCGGACATTGTCTAGTCAGGCAGAGGAACG
1198	DR-VASN for	GTCACCGGCCCTACATCTAAGCCAGAGAGAT
1199	DR-VASN rev	CCGGATCTCTCTGGCTTAGATGTAGGGCCG
1229	DR-VASN-dTAA for	GT CAC CGG CCC TAC ATC TAA TAA AGA GAG AT
1230	DR-VASN-dTAA rev	CCG GAT CTC TCT TTA TTA GAT GTA GGG CCG
Pos 4, 5, 6		
1449	DR-LDHB-[UAG CUA] for	GT CAC CGA AAA GAC CTG TAG CTA GTG AGC TT
1450	DR-LDHB-[UAG CUA] rev	CCG GAAGCTCACTAGCTACAGGTCTTTTCG
1451	DR-LDHB-[UGA UUA] for	GT CAC CGA AAA GAC CTG TGA TTA GTG AGC TT
1452	DR-LDHB-[UGA UUA] rev	CCG GAAGCTCACTAATCACAGGTCTTTTCG
1453	DR-LDHB-[UGA GUA] for	GT CAC CGA AAA GAC CTG TGA GTA GTG AGC TT
1454	DR-LDHB-[UGA GUA] rev	CCG GAAGCTCACTACTCACAGGTCTTTTCG
1455	DR-LDHB-[UGA CAA] for	GT CAC CGA AAA GAC CTG TGA CAA GTG AGC TT
1456	DR-LDHB-[UGA CAA] rev	CCG GAAGCTCACTTGTCACAGGTCTTTTCG

continued on next page...

Table 18: *continued*

OST No.	Name	Sequence 5' - 3'
1457	DR-LDHB-[UGA CCA] for	GT CAC CGA AAA GAC CTG TGA CCA GTG AGC TT
1458	DR-LDHB-[UGA CCA] rev	CCG GAAGCTCACTGGTCACAGGTCTTTTCG
1461	DR-LDHB-[UGA CUG] for	GT CAC CGA AAA GAC CTG TGA CTG GTG AGC TT
1462	DR-LDHB-[UGA CUG] rev	CCG GAAGCTCACCAGTCACAGGTCTTTTCG
1463	DR-MDH1-[UAA CUA] for	GT CAC CGT TCC TCT GCC TAA CTA GAC AAT G T
1464	DR-MDH1-[UAA CUA] rev	CCG GACATTGTCTAGTTAGGCAGAGGAACG
1467	DR-MDH1-[UAA UUA] for	GT CAC CGT TCC TCT GCC TAA TTA GAC AAT G T
1468	DR-MDH1-[UAA UUA] rev	CCG GACATTGTCTAATTAGGCAGAGGAACG
1469	DR-MDH1-[UGA UUA] for	GT CAC CGT TCC TCT GCC TGA TTA GAC AAT G T
1470	DR-MDH1-[UGA UUA] rev	CCG GACATTGTCTAATCAGGCAGAGGAACG
1471	DR-MDH1-[UGA GUA] for	GT CAC CGT TCC TCT GCC TGA GTA GAC AAT G T
1472	DR-MDH1-[UGA GUA] rev	CCG GACATTGTCTACTCAGGCAGAGGAACG
1475	DR-MDH1-[UGA CCA] for	GT CAC CGT TCC TCT GCC TGA CCA GAC AAT G T
1476	DR-MDH1-[UGA CCA] rev	CCG GACATTGTCTGGTCAGGCAGAGGAACG
1477	DR-MDH1-[UGA CUU] for	GT CAC CGT TCC TCT GCC TGA CTT GAC AAT G T
1478	DR-MDH1-[UGA CUU] rev	CCG GACATTGTCAAGTCAGGCAGAGGAACG
1479	DR-MDH1-[UGA CUG] for	GT CAC CGT TCC TCT GCC TGA CTG GAC AAT G T

continued on next page...

Table 18: *continued*

OST No.	Name	Sequence 5' - 3'
1480	DR-MDH1-[UGA CUG] rev	CCG GACATTGTCCAGTCAGGCAGAGGAACG
JH77	DR-MDH1-[UGA AUA] for	GTCACCGTTCCTCTGCCTGAATAGACAATGT
JH78	DR-MDH1-[UGA AUA] rev	CCGGACATTGTCTATTCAGGCAGAGGAACG
SCC		
1207	DR-LENG1-[UGG] for	GT CAC CGC CTT ACT CAC TGG CTC CTG AGG GT
1208	DR-LENG1-[UGG] rev	CCG GAC CCT CAG GAG CCA GTG AGT AAG GCG
1211	DR-PRDM10-[UGG] for	GT CAC CGC ACC AAA CCA TGG CTT CCA CCC TT
1212	DR-PRDM10-[UGG] rev	CCG GAA GGG TGG AAG CCA TGG TTT GGT GCG
1213	DR-FBLX20-[UGG] for	GT CAC CGC ATC ATC CTA TGG CAA TGG AGG TT
1214	DR-FBLX20-[UGG] rev	CCG GAA CCT CCA TTG CCA TAG GAT GAT GCG
1215	DR-THG1L-[UGG] for	GT CAC CGA GCC AGG CTT TGG CGG AAG AGT CT
1216	DR-THG1L-[UGG] rev	CCG GAG ACT CTT CCG CCA AAG CCT GGC TCG
1217	DR-EDEM3-[UGG] for	GT CAC CGG GAT GAG CTA TGG CTT GCT AAA CT
1218	DR-EDEM3-[UGG] rev	CCG GAG TTT AGC AAG CCA TAG CTC ATC CCG
1219	DR-EDN1-[UGG] for	GT CAC CGA GCA CAT TGG TGG CAG ACC TTC GT
1220	DR-EDN1-[UGG] for	CCG GAC GAA GGT CTG CCA CCA ATG TGC TCG
1221	DR-LEPRE1-[UGG] for	GT CAC CGG GAT GAG CTA TGG CAG CGT CCA GT
1222	DR-LEPRE1-[UGG] rev	CCG GAC TGG ACG CTG CCA TAG CTC ATC CCG

continued on next page...

Table 18: *continued*

OST No.	Name	Sequence 5' - 3'
1223	DR-UBQLN1-[UGG] for	GT CAC CGC CAG CCA TCA TGG CAG CAT TTC TT
1224	DR-UBQLN1-[UGG] rev	CCG GAA GAA ATG CTG CCA TGA TGG CTG GCG
1225	DR-IRAK3-[UGG] for	GT CAC CGC AAA AAA GAA TGG ATT CTA CCA GT
1226	DR-IRAK3-[UGG] rev	CCG GAC TGG TAG AAT CCA TTC TTT TTT GCG
1227	DR-SLC3A1-[UGG] for	GT CAC CGT ACC TCG TGT TGG GCA CCT TTA TT
1228	DR-SLC3A1-[UGG] rev	CCG GAA TAA AGG TGC CCA ACA CGA GGT ACG
1235	DR-MDH1-[UGG] for	GT CAC CGT TCC TCT GCC TGG CTA GAC AAT GT
1236	DR-MDH1-[UGG] rev	CCG GAC ATT GTC TAG CCA GGC AGA GGA ACG
IF		
1053	pEYFP-C1-LDHBx for	GCGCGAATTCTATGGCAACTCTTAAG-GAAAAAC
1054	pEYFP-C1-LDHBx rev	GCGCTCTAGACTACAGCCTAGAGCTCAC
FL-Construct-Purification		
1166	FL-LDHB for	CGC CATATG AAA CAT CAC CAT CAC CAT CAC CCC GCA ACT CTT AAG GAA AAA CTC
1167	FL-LDHB rev	CGCGC GAATTC TTA CAGGTCTTTTAGGTCCTTC
1166	FL-LDHB-[UGG] for	CGC CATATG AAA CAT CAC CAT CAC CAT CAC CCC GCA ACT CTT AAG GAA AAA CTC
1168	FL-LDHB-[UGG] rev	CGCGC GAATTC CTACAGCCTAGAGCTCAC

References

Atkins JF, Weiss RB, Gesteland RF (1990): Ribosome gymnastics—Degree of difficulty 9.5, style 10.0. Cell 62, 413–423

Baumgart E, Fahimi HD, Stich A, Völkl A (1996): L-Lactate Dehydrogenase A- and AB Isoforms Are Bona Fide Peroxisomal Enzymes in Rat Liver Evidence for involvement in intraperoxisomal NADH reoxidation. J Biol Chem 271, 3846–3855

Beier H, Grimm M (2001): Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res 29, 4767–4782

Berg JM, Tymoczko JL, Stryer L: Biochemistry: International Edition. 7th edition.; W. H. Freeman, Basingstoke 2011

Bergstrom DE, Merli CA, Cygan JA, Shelby R, Blackman RK (1995): Regulatory Autonomy and Molecular Characterization of the Drosophila Out at First Gene. Genetics 139, 1331–1346

Bidou L, Allamand V, Rousset J-P, Namy O (2012): Sense from nonsense: therapies for premature stop codon diseases. Trends Mol Med 18, 679–688

Bolte S, Cordelières FP (2006): A guided tour into subcellular colocalization analysis in light microscopy. J Microsc 224, 213–232

Bonetti B, Fu L, Moon J, Bedwell DM (1995): The Efficiency of Translation Termination is Determined by a Synergistic Interplay Between Upstream and Downstream Sequences in Saccharomyces cerevisiae. J Mol Biol 251, 334–345

Bossi L (1983): Context effects: Translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. J Mol Biol 164, 73–87

Brocard C, Hartig A (2006): Peroxisome targeting signal 1: Is it really a simple tripeptide? Biochim Biophys Acta BBA - Mol Cell Res 1763, 1565–1573

Brooks GA, Brown MA, Butz CE, Sicurello JP, Dubouchaud H (1999): Cardiac and skeletal muscle mitochondria have a monocarboxylate transporter MCT1. J Appl Physiol 87, 1713–1718

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

Brown CM, Stockwell PA, Trotman CNA, Tate WP (1990a): Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. Nucleic Acids Res 18, 6339–6345

Brown CM, Stockwell PA, Trotman CN, Tate WP (1990b): The signal for the termination of protein synthesis in procaryotes. Nucleic Acids Res 18, 2079–2086

Brown CM, Dinesh-Kumar SP, Miller WA (1996): Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. J Virol 70, 5884–5892

Bruyère A, Brault V, Ziegler-Graff V, Simonis MT, Van den Heuvel JF, Richards K, Guilley H, Jonard G, Herrbach E (1997): Effects of mutations in the beet western yellows virus readthrough protein on its expression and packaging and on virus accumulation, symptoms, and aphid transmission. Virology 230, 323–334

Cassan M, Rousset J-P (2001): UAG readthrough in mammalian cells: Effect of upstream and downstream stop codon contexts reveal different signals. BMC Mol Biol 2, 3

CBS (2013): TMHMM server v. 2.0. URL http://www.cbs.dtu.dk/services/TMHMM/. [Online; accessed 11-March-2013]

Chittum HS, Lane WS, Carlson BA, Roller PP, Lung F-DT, Lee BJ, Hatfield DL (1998): Rabbit β-globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps. Biochemistry (Mosc) 37, 10866–10870

Claros MG, von Heijne G (1994): TopPred II: an improved software for membrane protein structure predictions. Comput Appl Biosci CABIOS 10, 685–686

Dam EBT, Pleij CWA, Bosch L (1990): RNA pseudoknots: Translational frameshifting and readthrough on viral RNAs. Virus Genes 4, 121–136

De Duve C (1969): The peroxisome: a new cytoplasmic organelle. Proc R Soc Lond B Biol Sci 173, 71–83

Dettmer U, Folkerts M, Kächler E, Sönnichsen A: Biochemie Intensivkurs. 1st edition; Urban & Fischer Verlag/Elsevier GmbH. München 2005

Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. (2007): Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203–218

Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS (2013): Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. eLife 2, e01179

Engelberg-Kulka H (1981): UGA suppression by normal tRNA Trp in Escherichia coli: codon context effects. Nucleic Acids Res 9, 983–991

Eswarappa SM, Potdar AA, Koch WJ, Fan Y, Vasu K, Lindner D, Willard B, Graham LM, DiCorleto PE, Fox PL (2014): Programmed Translational Readthrough Generates Antiangiogenic VEGF-Ax. Cell 157, 1605–1618

Fearon K, McClendon V, Bonetti B, Bedwell DM (1994): Premature translation termination mutations are efficiently suppressed in a highly conserved region of yeast Ste6p, a member of the ATP-binding cassette (ABC) transporter family. J Biol Chem 269, 17802–17808

Felsenstein KM, Goff SP (1988): Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. J Virol 62, 2179–2182

Feng YX, Yuan H, Rein A, Levin JG (1992): Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence

immediately downstream of the gag termination codon followed by an RNA pseudoknot. J Virol 66, 5127–5132

Firth AE, Brierley I (2012): Non-canonical translation in RNA viruses. J Gen Virol 93, 1385–1409

Firth AE, Wills NM, Gesteland RF, Atkins JF (2011): Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. Nucleic Acids Res 39, 6679–6691

Floquet C, Hatin I, Rousset J-P, Bidou L (2012): Statistical Analysis of Readthrough Levels for Nonsense Mutations in Mammalian Cells Reveals a Major Determinant of Response to Gentamicin. PLoS Genet 8, e1002608

Freitag J, Ast J, Bölker M (2012): Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. Nature 485, 522–525

Geller AI, Rich A (1980): A UGA termination suppression tRNATrp active in rabbit reticulocytes. Nature 283, 41–46

Genscript (2014): Codon usage frequence table tool. URL http://www.genscript.com/cgi-bin/tools/codon_freq_table. [Online; accessed 2-September-2014]

Gesteland RF, Atkins JF (1996): Recoding: Dynamic Reprogramming of Translation. Annu Rev Biochem 65, 741–768

Gladden LB (2004): Lactate metabolism: a new paradigm for the third millennium. J Physiol 558, 5–30

Gronemeyer T, Wiese S, Ofman R, Bunse C, Pawlas M, Hayen H, Eisenacher M, Stephan C, Meyer HE, Waterham HR, et al. (2013): The Proteome of Human Liver Peroxisomes: Identification of Five New Peroxisomal Constituents by a Label-Free Quantitative Proteomics Survey. PLoS ONE 8, e57395

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008): The Vienna RNA websuite. Nucleic Acids Res 36, W70–74

Halestrap A, Price N (1999): The proton-linked monocarboxylate transporter (MCT) family: structure, function and regulation. Biochem J 343, 281–299

Harper SJ, Bates DO (2008): VEGF-A splicing. Nat Rev Cancer 8, 880–887

Harrell L, Melcher U, Atkins JF (2002): Predominance of six different hexanucleotide recoding signals 3′ of read-through stop codons. Nucleic Acids Res 30, 2011–2017

Hatfield D, Thorgeirsson SS, Copeland TD, Oroszlan S, Bustin M (1988): Immunopurification of the suppressor tRNA dependent rabbit. beta.-globin readthrough protein. Biochemistry (Mosc) 27, 1179–1183

Hayasaka K, Himoro M, Sawaishi Y, Nanao K, Takahashi T, Takada G, Nicholson GA, Ouvrier RA, Tachi N (1993): De novo mutation of the myelin Po gene in Dejerine–Sottas disease (hereditary motor and sensory neuropathy type III). Nat Genet 5, 266–268

Hayashi H, Miwa A (1989): The role of peroxisomal fatty acyl-CoA beta-oxidation in bile acid biosynthesis. Arch Biochem Biophys 274, 582–589

Hayashi H, Takahata S (1991): Role of peroxisomal fatty acyl-CoA beta-oxidation in phospholipid biosynthesis. Arch Biochem Biophys 284, 326–331

Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G (2007): Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450, 1026–1030

Hofmann K, Stoffel W (1993): Tmpred. URL http://www.ch.embnet.org/software/TMPRED_form.html. [Online; accessed 11- March-2013]

Hofstetter H, Monstein H-J, Weissmann C (1974): The readthrough protein A1 is essential for the formation of viable Qβ particles. Biochim Biophys Acta BBA - Nucleic Acids Protein Synth 374, 238–251

Honigman A, Wolf D, Yaish S, Falk H, Panet A (1991): cis Acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. Virology 183, 313–319

Howard MT, Shirts BH, Petros LM, Flanigan KM, Gesteland RF, Atkins JF (2000): Sequence specificity of aminoglycoside-induced stop condon readthrough: potential implications for treatment of Duchenne muscular dystrophy. Ann Neurol 48, 164–169

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009): Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 324, 218–223

Ingolia NT, Lareau LF, Weissman JS (2011): Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell 147, 789–802

Joosse S, Hannemann J (2013): Sequence conversion. URL http://in-silico.net/tools/biology/sequence_conversion. [Online; accessed 07-October-2013]

Jungreis I, Lin MF, Spokony R, Chan CS, Negre N, Victorsen A, White KP, Kellis M (2011): Evidence of abundant stop codon readthrough in Drosophila and other metazoa. Genome Res 21, 2096–2113

Juretić D, Zoranić L, Zucić D (2002): Basic charge clusters and predictions of membrane protein topology. J Chem Inf Comput Sci 42, 620–632

Kabran P, Rossignol T, Gaillardin C, Nicaud J-M, Neuvéglise C (2012): Alternative Splicing Regulates Targeting of Malate Dehydrogenase in Yarrowia lipolytica. DNA Res Int J Rapid Publ Rep Genes Genomes 19, 231–244

Keeling KM, Xue X, Gunn G, Bedwell DM (2014): Therapeutics Based on Stop Codon Readthrough. Annu Rev Genomics Hum Genet 15, 371–394

Kibbe WA (2007): OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35, W43–46

Klagges BR, Heimbeck G, Godenschwege TA, Hofbauer A, Pflugfelder GO, Reifegerste R, Reisch D, Schaupp M, Buchner S, Buchner E (1996): Invertebrate synapsins: a single gene codes for several isoforms in Drosophila. J Neurosci Off J Soc Neurosci 16, 3154–3165

Kopczynski JB, Raff AC, Bonner JJ (1992): Translational readthrough at nonsense mutations in the HSF1 gene of Saccharomyces cerevisme. Mol Gen Genet MGG 234, 369–378

Lanyon-Hogg T, Warriner SL, Baker A (2010): Getting a camel through the eye of a needle: the import of folded proteins by peroxisomes. Biol Cell 102, 245–263

Lazarow PB, De Duve C (1976): A fatty acyl-CoA oxidizing system in rat liver peroxisomes; enhancement by clofibrate, a hypolipidemic drug. Proc Natl Acad Sci 73, 2043–2046

Li G, Rice CM (1993): The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. J Virol 67, 5062–5067

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. (2011): A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478, 476–482

Lingner T, Kataya AR, Antonicelli GE, Benichou A, Nilssen K, Chen X-Y, Siemsen T, Morgenstern B, Meinicke P, Reumann S (2011): Identification of Novel Plant Peroxisomal Targeting Signals by a Combination of Machine Learning Methods and in Vivo Subcellular Targeting Analyses. Plant Cell 23, 1556–1572

Lin MF, Carlson JW, Crosby MA, Matthews BB, Yu C, Park S, Wan KH, Schroeder AJ, Gramates LS, St. Pierre SE, et al. (2007): Revisiting the protein-coding gene catalog of Drosophila melanogaster using 12 fly genomes. Genome Res 17, 1823–1836

Lin MF, Jungreis I, Kellis M (2011): PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27, i275–i282

Liu X, Ma C, Subramani S (2012): Recent advances in peroxisomal matrix protein import. Curr Opin Cell Biol 24, 484–489

Lodhi IJ, Semenkovich CF (2014): Peroxisomes: a Nexus for Lipid Metabolism and Cellular Signaling. Cell Metab 19, 380–392

Loughran G, Chou M-Y, Ivanov IP, Jungreis I, Kellis M, Kiran AM, Baranov PV, Atkins JF (2014): Evidence of efficient stop codon readthrough in four mammalian genes. Nucleic Acids Res 42, 8928–8938

Markert CL (1963): Lactate Dehydrogenase Isozymes: Dissociation and Recombination of Subunits. Science 140, 1329–1330

Martin R, Phillips-Jones MK, Watson FJ, Hill LS (1993): Codon context effects on nonsense suppression in human cells. Biochem Soc Trans 21, 846–851

Marzioch M, Erdmann R, Veenhuis M, Kunau WH (1994): PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes. EMBO J 13, 4908–4918

Maurer-Stroh S, Eisenhaber F (2005): Refinement and prediction of protein prenylation motifs. Genome Biol 6, R55

Maynard EL, Gatto GJ, Berg JM (2004): Pex5p binding affinities for canonical and noncanonical PTS1 peptides. Proteins Struct Funct Bioinforma 55, 856–861

McAlister-Henn L, Steffan JS, Minard KI, Anderson SL (1995): Expression and Function of a Mislocalized Form of Peroxisomal Malate Dehydrogenase (MDH3) in Yeast. J Biol Chem 270, 21220–21225

McCaughan KK, Brown CM, Dalphin ME, Berry MJ, Tate WP (1995): Translational termination efficiency in mammals is influenced by the base following the stop codon. Proc Natl Acad Sci 92, 5431–5435

McClelland GB, Khanna S, González GF, Eric Butz C, Brooks GA (2003): Peroxisomal membrane monocarboxylate transporters: evidence for a redox shuttle system? Biochem Biophys Res Commun 304, 130–135

McGroarty E, Hsieh B, Wied DM, Gee R, Tolbert NE (1974): Alpha hydroxy acid oxidation by peroxisomes. Arch Biochem Biophys 161, 194-210

McNew JA, Goodman JM (1996): The targeting and assembly of peroxisomal proteins: some old rules do not apply. Trends Biochem Sci 21, 54–58

Miller JH, Albertini AM (1983): Effects of surrounding sequence on the suppression of nonsense codons. J Mol Biol 164, 59–71

Moody, DE Reddy, JK (1976): Morphometric analysis of the ultrastructural changes in rat liver induced by the peroxisome proliferator SaH 42-348. J Cell Biol 71, 768–780

Mottagui-Tabar S, Björnsson A, Isaksson LA (1994): The second to last amino acid in the nascent peptide as a codon context determinant. EMBO J 13, 249–257

Mottagui-Tabar S, Tuite MF, Isaksson LA (1998): The influence of 5′ codon context on translation termination in Saccharomyces cerevisiae. Eur J Biochem 257, 249–254

Namy O, Hatin I, Rousset J-P (2001): Impact of the six nucleotides downstream of the stop codon on translation termination. EMBO Rep 2, 787–793

Namy O, Duchateau-Nguyen G, Rousset J-P (2002): Translational readthrough of the PDE2 stop codon modulates cAMP levels in Saccharomyces cerevisiae. Mol Microbiol 43, 641–652

Namy O, Duchateau-Nguyen G, Hatin I, Hermann-Le Denmat S, Termier M, Rousset J-P (2003): Identification of stop codon readthrough genes in Saccharomyces cerevisiae. Nucleic Acids Res 31, 2289–2296

OMIM® (2014a): Online mendelian inheritance in man. URL: http://omim.org/entry/150000. [Online; accessed 7-September-2014]

OMIM® (2014b): Online mendelian inheritance in man. URL: http://omim.org/entry/150100. [Online; accessed 7-September-2014]

Osmundsen H (1982): Factors which can influence beta-oxidation by peroxisomes isolated from livers of clofibrate treated rats. Some properties of peroxisomal fractions isolated in a self-generated Percoll gradient by vertical rotor centrifugation. Int J Biochem 14, 905–914

Pacho F, Zambruno G, Calabresi V, Kiritsi D, Schneider H (2011): Efficiency of translation termination in humans is highly dependent upon nucleotides in the neighbourhood of a (premature) termination codon. J Med Genet 48, 640–644

Pelham HRB (1978): Leaky UAG termination codon in tobacco mosaic virus RNA. Nature 272, 469–471

Poole ES, Brown CM, Tate WP (1995): The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. EMBO J 14, 151

Poole ES, Brimacombe R, Tate WP (1997): Decoding the translational termination signal: the polypeptide chain release factor in Escherichia coli crosslinks to the base following the stop codon. RNA 3, 974–982

Poole ES, Major LL, Mannering SA, Tate WP (1998): Translational termination in Escherichia coli: three bases following the stop codon crosslink to release factor 2 and affect the decoding efficiency of UGA-containing signals. Nucleic Acids Res 26, 954–960

Robinson DN, Cooley L (1997): Examination of the function of two kelch proteins generated by stop codon suppression. Development 124, 1405–1417

Salido E, Pey AL, Rodriguez R, Lorenzo V (2012): Primary hyperoxalurias: Disorders of glyoxylate detoxification. Biochim Biophys Acta BBA - Mol Basis Dis 1822, 1453–1464

Samuels ME, Schedl P, Cline TW (1991): The complex set of late transcripts from the Drosophila sex determination gene sex-lethal encodes multiple related polypeptides. Mol Cell Biol 11, 3584–3602

Schneider CA, Rasband WS, Eliceiri KW (2012): NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671–675

Schrader M, Grille S, Fahimi HD, Islinger M: Peroxisome Interactions and Cross-Talk with Other Subcellular Compartments in Animal Cells; in: Peroxisomes and their Key Role in Cellular Signaling and Metabolism; hrsg. v. Río LA del; Springer Netherlands 2013, 1–22

Schueren F, Lingner T, George R, Hofhuis J, Dickel C, Gärtner J, Thoms S (2014): Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. eLife 3, e03640

Skuzeski JM, Nichols LM, Gesteland RF, Atkins JF (1991): The signal for a leaky UAG

stop codon in several plant viruses includes the two downstream codons. J Mol Biol 218, 365–373

Smith D, Yarus M (1989): tRNA-tRNA interactions within cellular ribosomes. Proc Natl Acad Sci U S A 86, 4397–4401

Smith JJ, Aitchison JD (2013): Peroxisomes take shape. Nat Rev Mol Cell Biol 14, 803–817

Steitz JA (1969): Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature 224, 957–964

Steneberg P, Samakovlis C (2001): A novel stop codon readthrough mechanism produces functional Headcase protein in Drosophila trachea. EMBO Rep 2, 593–597

Stiebler AC, Freitag J, Schink KO, Stehlik T, Tillmann BAM, Ast J, Bölker M (2014): Ribosomal Readthrough at a Short UGA Stop Codon Context Triggers Dual Localization of Metabolic Enzymes in Fungi and Animals. PLoS Genet 10, e1004685

Takanami M, Yan Y, Jukes TH (1965): Studies on the site of ribosomal binding of f2 bacteriophage RNA. J Mol Biol 12, 761–773

Thoms S, Grønborg S, Gärtner J (2009): Organelle interplay in peroxisomal disorders. Trends Mol Med 15, 293–302

Thoms S, Harms I, Kalies K-U, Gärtner J (2012): Peroxisome Formation Requires the Endoplasmic Reticulum Channel Protein Sec61: ER Translocon Requirement for Peroxisome Formation. Traffic 13, 599–609

Urban C, Zerfass K, Fingerhut C, Beier H (1996): UGA suppression by tRNACmCATrp occurs in diverse virus RNAs due to a limited influence of the codon context. Nucleic Acids Res 24, 3424–3430

Urich K: Comparative Animal Biochemistry. Reprint of 1st edition. Springer Science & Business Media 2013

Van Roermund CW, Elgersma Y, Singh N, Wanders RJ, Tabak HF (1995): The membrane of peroxisomes in Saccharomyces cerevisiae is impermeable to NAD (H) and acetyl-CoA under in vivo conditions. EMBO J 14, 3480

Visser WF, van Roermund CWT, Ijlst L, Waterham HR, Wanders RJA (2007): Metabolite transport across the peroxisomal membrane. Biochem J 401, 365

Völkl A, Fahimi HD (1985): Isolation and characterization of peroxisomes from the liver of normal untreated rats. Eur J Biochem 149, 257–265

Walton PA, Hill PE, Subramani S (1995): Import of stably folded proteins into peroxisomes. Mol Biol Cell 6, 675–683

Wanders RJA, Waterham HR (2006a): Biochemistry of Mammalian Peroxisomes Revisited. Annu Rev Biochem 75, 295–332

Wanders RJA, Waterham HR (2006b): Peroxisomal disorders: The single peroxisomal enzyme deficiencies. Biochim Biophys Acta BBA - Mol Cell Res 1763, 1707–1720

Weiner AM, Weber K (1971): Natural Read-through at the UGA Termination Signal of Qβ Coat Protein Cistron. Nature 234, 206–209

Weller S, Gould SJ, Valle D (2003): Peroxisome Biogenesis Disorders. Annu Rev Genomics Hum Genet 4, 165–211

Williams I (2004): Genome-wide prediction of stop codon readthrough during translation in the yeast Saccharomyces cerevisiae. Nucleic Acids Res 32, 6605–6616

Wills NM, Gesteland RF, Atkins JF (1991): Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. Proc Natl Acad Sci U S A 88, 6991–6995

Yamaguchi Y, Hayashi A, Campagnoni CW, Kimura A, Inuzuka T, Baba H (2012): L-MPZ, a Novel Isoform of Myelin P0, Is Produced by Stop Codon Readthrough. J Biol Chem 287, 17765–17776

Yoshinaka Y, Katoh I, Copeland TD, Oroszlan S (1985a): Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc Natl Acad Sci 82, 1618–1622

Yoshinaka Y, Katoh I, Copeland TD, Oroszlan S (1985b): Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. J Virol 55, 870–873

Zerfass K, Beier H (1992): Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. Nucleic Acids Res 20, 5911–5918

Thanks to

I would like to thank everybody in the laboratory, all of you made it possible that I could experience exiting moments in research and you helped me with every question that came to mind, so thank you very much.
For their help and encouragement as well as their guidance and friendliness, I want to thank especially the following people:
Corinna Dickel
Judith Büntzel
Jutta Gärtner
Rosemol George
Thomas Lingner
Kareem Soliman
Sven Thoms
Tanja Wilke
I want to express my gratitude for the financial support offered by scholarship Go4med of UMG.