



IMLP, a predictor for internal matrix targeting-like sequences in mitochondrial proteins

Schneider, Kevin; Zimmer, David; Nielsen, Henrik; Herrmann, Johannes M.; Mühlhaus, Timo

Published in:
Biological Chemistry

Link to article, DOI:
[10.1515/hsz-2021-0185](https://doi.org/10.1515/hsz-2021-0185)

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Schneider, K., Zimmer, D., Nielsen, H., Herrmann, J. M., & Mühlhaus, T. (2021). IMLP, a predictor for internal matrix targeting-like sequences in mitochondrial proteins. *Biological Chemistry*, 402(8), 937-943.
<https://doi.org/10.1515/hsz-2021-0185>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Kevin Schneider, David Zimmer, Henrik Nielsen, Johannes M. Herrmann and Timo Mühlhaus*

iMLP, a predictor for internal matrix targeting-like sequences in mitochondrial proteins

<https://doi.org/10.1515/hsz-2021-0185>

Received March 8, 2021; accepted June 7, 2021;

published online July 5, 2021

Abstract: Matrix targeting sequences (MTSs) direct proteins from the cytosol into mitochondria. Efficient targeting often relies on internal matrix targeting-like sequences (iMTS-Ls) which share structural features with MTSs. Predicting iMTS-Ls was tedious and required multiple tools and webservices. We present iMLP, a deep learning approach for the prediction of iMTS-Ls in protein sequences. A recurrent neural network has been trained to predict iMTS-L propensity profiles for protein sequences of interest. The iMLP predictor considerably exceeds the speed of existing approaches. Expanding on our previous work on iMTS-L prediction, we now serve an intuitive iMLP webservice available at <http://iMLP.bio.uni-kl.de> and a stand-alone command line tool for power user in addition.

Keywords: deep learning; mitochondria; protein targeting; recurrent neural network; sequence analysis; webservice.

Introduction

Mitochondrial genomes code only for a handful of proteins. Most of the roughly 800 (yeast) to 1500 (human) mitochondrial proteins are encoded in the nucleus. These proteins are synthesized in the cytosol as precursor proteins

that have to be imported into the organelle (Avendaño-Monsalve et al. 2020; Chacinska et al. 2008), which is realized by targeting signals of characteristic features of their primary and secondary structures (Grevel et al. 2019; Heijne 1986; Vögtle et al. 2009)

Most proteins of the mitochondrial matrix and the inner membrane contain matrix targeting sequences (MTSs). These amino-terminal presequences are necessary and sufficient for the import of the respective protein into mitochondria. They are recognized by translocases in the outer mitochondrial membrane (the TOM complex) and the inner mitochondrial membrane (the TIM23 complex) which facilitate their translocation into mitochondria (Heijne 1986). MTSs are often removed from these precursors during the translocation process by the mitochondrial processing peptidase (MPP) (Calvo et al. 2017; Friedl et al. 2020; Vögtle et al. 2009). As recently discovered, many matrix proteins contain additional internal sequences which support protein translocation into mitochondria. These regions were called MTS-like sequences (iMTS-Ls) owing to their MTS-like structural properties; however, in contrast to the amino-terminal MTSs they are located within the mature region of the protein (Backes et al. 2018). While neither necessary nor sufficient on their own for translocation, iMTS-Ls can considerably increase import efficiency of the precursors. This stimulating effect is mediated by the mitochondrial surface receptor Tom70 in conjunction with the cytosolic chaperone machinery (Backes et al. 2020; Kreimendahl and Rassow 2020; Young et al. 2003). The interaction with Tom70 is particularly relevant for aggregation-prone or hydrophobic proteins and cooperative binding of several Tom70 receptors to one precursor might support the unfolding and import of precursor proteins (Backes et al. 2020; Backes and Herrmann 2017; Brix et al. 1999, 2000; Yamamoto et al. 2009). The iMTS-Ls-mediated protein targeting seems not to be restricted to MTS-containing proteins. For instance, an iMTS-L was found to target the ubiquitin ligase Ubx2 onto the mitochondrial surface (Mårtensson et al. 2019) or the cytosolic protein Orf9b of the SARS-CoV2 virus (Gordon et al. 2020; Samavarchi-Tehrani et al. 2020). The latter is particularly interesting as the COVID-19 causing coronavirus expresses an iMTS-L containing peptide in high amounts in order to inhibit Tom70 function in infected cells

Kevin Schneider and David Zimmer contributed equally to this work.

***Corresponding author: Timo Mühlhaus**, Computational Systems Biology, University of Kaiserslautern, Paul-Ehrlich-Strasse 23, D-67663 Kaiserslautern, Germany, E-mail: muehlhaus@bio.uni-kl.de. <https://orcid.org/0000-0003-3925-6778>

Kevin Schneider and David Zimmer, Computational Systems Biology, University of Kaiserslautern, Paul-Ehrlich-Strasse 23, D-67663 Kaiserslautern, Germany. <https://orcid.org/0000-0002-2198-5262> (K. Schneider). <https://orcid.org/0000-0002-7832-7658> (D. Zimmer)

Henrik Nielsen, Department of Health Technology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Johannes M. Herrmann, Cell Biology, University of Kaiserslautern, D-67663 Kaiserslautern, Germany. <https://orcid.org/0000-0003-2081-4506>

(Jiang et al. 2020; Miserey-Lenkei et al. 2021; Shi et al. 2014). The physiological relevance of iMTS-L sequences still has to be explored more generally but these examples already show that these Tom70-interaction regions are not necessarily restricted to mitochondrial proteins and can be highly relevant in the context of diseases.

MTSs and iMTS-L are helical regions of varying lengths, mostly between 10 and 70 residues. One face of these amphipathic helices is positively charged, one is hydrophobic; negative charges are absent and serine and threonine residues are frequent. MTSs can be detected by algorithms which either scan the sequences for these characteristic pre-defined features or which were trained on protein sequences of known localization. The TargetP algorithm is one of the machine-learning programs which is widely used in the community (Emanuelsson et al. 2000, 2007). It is well suited for the detection of MTSs and detects amino-terminal sequences with high accuracy.

Here we describe a user-friendly webservice implementation, coined iMLP, for iMTS-L prediction. This tool leverages the capabilities of sequence-based deep learning to offer a fast and reliable prediction of iMTS-L propensity profiles for any given input sequence through a convenient interface.

Results

The iMLP prediction workflow

The iMLP application presented here provides a direct and fast approach to predict iMTS-L propensity profiles from protein sequences of interest. This is achieved by an algorithm based on an artificial recurrent neural network architecture named long short-term memory (LSTM). These architectures are specially designed for feature detection in sequences and therefore well suited for the recognition of iMTS-Ls. In this setup, we trained our model in a supervised fashion, on a set of training sequences derived from the well-established TargetP algorithm. We exploited the fact shown previously, that iMTS-L stretches, supporting Tom70 binding, share the typical biochemical properties of mitochondrial targeting signals and can therefore be predicted classically using TargetP. However, to identify iMTS-L stretches, additional steps are necessary that are comprised of (i) preprocessing, (ii) profile generation, and (iii) a normalization step. During preprocessing, the restriction of analyzing only the N-terminus of a given protein is relieved. Due to the nature of the TargetP MTS prediction, a single TargetP score reflects the probability of the presence or absence of a transit peptide. This might lead

to sudden changes in the positional score along a sequence that are unreasonable when expecting contiguous motives in a sequence. Therefore, the vector of raw TargetP scores is smoothed using a Savitzky-Golay filter with the window size of the expected value of the distribution of known MTS lengths (Vögtle et al. 2009). To identify iMTS-Ls in the resulting iMTS profiles, the signal of the respective target sequence is normalized to background signal using artificial signals generated by analyzing 5000 random sequences. In our framework, we applied this procedure to the combined reference proteomes of yeast, mice, and human creating the dataset to train and evaluate our model. It can be directly consumed by our webservice (<http://impl.bio.uni-kl.de/>) or using a stand-alone version of the prediction tool (<https://github.com/CSBiology/iMLP>) and allows direct and fast identification of iMTS-Ls (Figure 1). Additional to the source code of the webservice (https://github.com/CSBiology/iMLP_Web), we provide the model in Open Neural Network Exchange (ONNX) standardized export format for reproducibility and validation independent of the machine learning framework. Exploiting a multi organism proteome as a learning data set, we could perform a train-test split in a 1:9 ratio, which enabled us to test our model using 2.500,000 iMTS-L predictions not encountered during the training process. Comparing model based iMTS-L propensity scores with those assigned to the test data set, we reached an accuracy of 99.10% determined by Pearson correlation. This seems sufficient considering a certain indistinctness of the indirect identification using TargetP followed by the processing procedure, indicating a successful direct mapping between protein sequence and iMTS-L propensity score (Figure 2A). This global quality assessment can be complemented by the in-depth analysis of well-studied candidates, such as the yeast ATP synthase alpha subunit *Atp1*. The comparative visualization of the iMTS-L propensity scores based on the TargetP based calculation and the LSTM based model shows that both prediction approaches agree in predicting the respective experimentally verified iMTS-L positions (Figure 2B). The comparison shows a different behavior of the TargetP2.0 algorithm, which is known to have a higher accuracy in predicting classical N-terminal MTS (Almagro Armenteros et al. 2019). However, this prediction does not match the iMTS-L propensity score expected according to the experimental evidence (Backes et al. 2018).

Analogously, we provide a plant model that was learned and evaluated on a combined reference proteomes of *Arabidopsis thaliana*, *Zea mays* and *Oryza sativa* with an accuracy of 99% determined by Pearson correlation (Supplementary Figure 1A). While the TargetP2.0-based prediction differs, iMLP is able to detect iMTS-Ls in

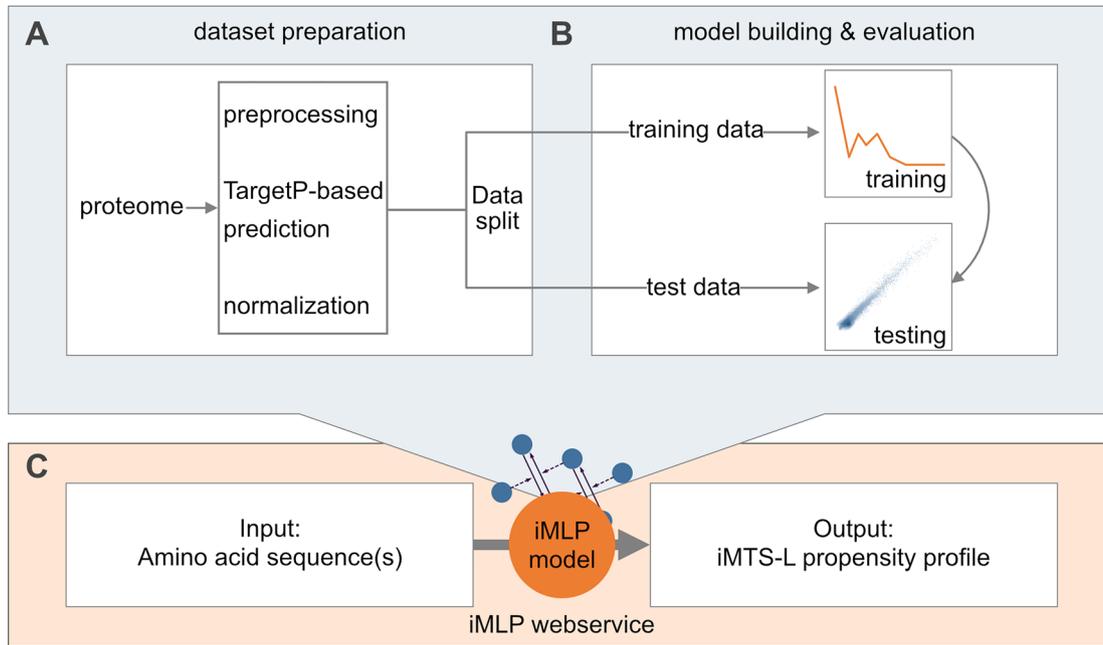


Figure 1: Schematic overview illustrating model building and consumption.

(A) Based on the protein sequences of the yeast, mice and human proteomes, we used the previously published workflow to calculate LSTM scores utilizing the independent calculation of TargetP scores followed by downstream normalization. The resulting data set consisting of 25,000,000 iMTS-L propensity scores was subsequently split in a 9:1 ratio into a test and training set used for training and evaluation, respectively (B) The multilayer LSTM based RNN was then trained to directly predict the iMTS-L propensity scores by minimizing the squared error between the model prediction and the scores assigned to the training set. The generalization of the model was verified by the correlation with the respective scores of the test set. (C) The resulting iMLP model can now be consumed using the iMLP web service, directly mapping a user defined amino acid sequence(s) to the corresponding iMTS-L propensity score profile(s).

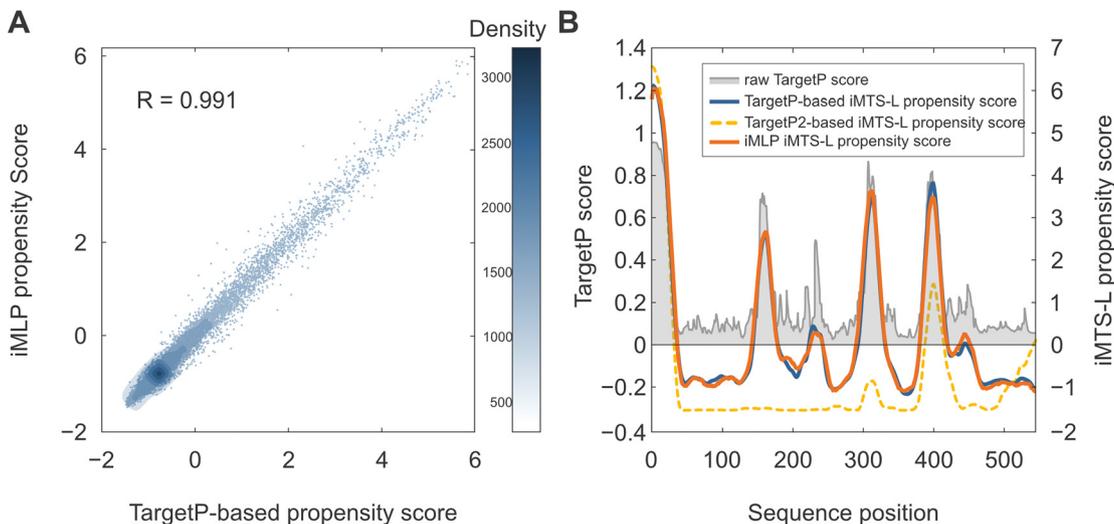


Figure 2: Validation of the deep learning based iMTS-L propensity score prediction.

(A) The genome wide model performance was validated comparing model-based iMTS-L propensity score predictions with previously unseen validation data consisting of 10% of the total data set with ~25 million entities. Considering the imperfect nature of the initial TargetP based iMTS-L propensity score calculation, the computed Pearson correlation of 0.991 indicates that the model is capable to create a direct mapping between protein sequence and iMTS-L propensity score (B) The exemplary comparison of scores analyzing the well-studied ATP1 protein sequence shows that the prediction approaches of iMLP and TargetP1 are in agreement predicting the respective iMTS-L positions (orange and blue), while TargetP2 shows a very different profile (yellow).

SDHB-RPS14 from rice flanking the cleavage site (G247) as expected according to Friedl et al. 2020 based on homology (Supplementary Figure 1B). Therefore, we offer the plant model as an experimental feature of our application that will need additional experimental validation.

Most recently, the interaction between Tom70 and Sars-Cov2 protein Orf9b was shown to be of high importance (Jiang et al. 2020). Using iMLP we could analyze the sequence of Orf9b and identify the interaction region, which is clearly separated from the rest of the sequence. The direct generation of the iMTS-L profile allows a more direct interpretation of the prediction result compared to the TargetP raw scores (Figure 3). In general, iMTS-L propensity scores above zero indicate iMTS-L stretches. It becomes apparent that the normalization versus random sequences increases the signal-to-noise ratio of the score derived by our model when compared to the raw TargetP output. The same holds true for the inspection of the well-studied Atp1 yeast protein (Figure 2B).

Web interface for the iMLP service

In order to use iMLP to predict iMTS-Ls for their own research, users do not have to undergo installation procedures or provide computational resources but can retrieve the prediction via an easy-to-use web interface

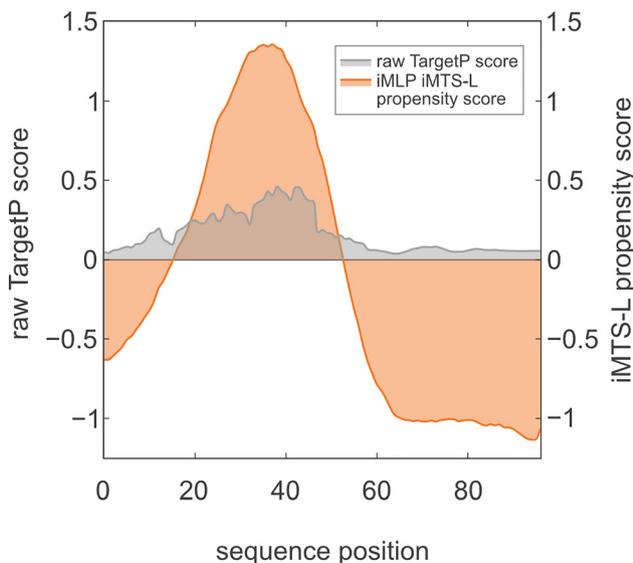


Figure 3: Normalization of the targetP score leads to an improved signal-to-noise ratio.

Analyzing the SARS-CoV2 protein Orf9b, a recently verified Tom70 interaction partner (Jiang et al. 2020), the iMTS-L propensity score clearly separates the predicted target sequence stretch (score > 0) from the background noise (score < 0).

(Figure 4). The iMLP webservice is available at <http://iMLP.bio.uni-kl.de>. Here, users can submit either individual protein sequences or files with multiple entries. The output is a detailed prediction report including: (i) a score heatmap of the input sequence, (ii) predicted iMTS-L propensity profile (P_a^{iMTS-L}). The graphics (ii) are generated with plotly.js and are therefore well suited for exploratory analysis due to their interactivity. These plots can be easily downloaded using the functionality above the respective plots. If multiple sequences are provided in a file, each sequence report is displayed in a separate tab on the page. Additionally, iMLP results can also be exported as tab separated files.

Discussion

Predicting iMTS-Ls is of great interest to identify protein sequences that determine the intracellular distribution of proteins. Such iMTS-L sequences are frequently found in mitochondrial proteins but also present in many cytosolic proteins, in some cases obviously to facilitate protein association with the mitochondrial surface (Backes et al. 2018). The relevance of iMTS-Ls was only experimentally analyzed for a small set of model proteins and a general, proteome-wide analysis is lacking. This necessitates our machine-trains-machine learning approach to derive a predictive model for fast iMTS-L sequence detection.

In many cases, potential iMTS-Ls might be hidden in the three-dimensional structure of proteins and therefore not accessible once proteins are folded. However, these sequences might be relevant during the folding reaction of proteins and provide binding sites for chaperones, co-chaperones as well as proteins with tetratricopeptide repeat (TPR) domains, such as Tom70. Alternatively, in organisms lacking Tom70 other mitochondrial receptors with TPR domains could recognize iMTS-L sequences. Om64 in plants (Nickel et al. 2019) or ATOM69 in the mitochondrial outer membrane of Kinetoplastids (Mani et al. 2015) have been proposed to perform a Tom70 like function.

In this description, we present a tool for a simplified iMTS-L prediction. Protocols for iMTS-L prediction had been published before (Backes et al. 2018; Boos et al. 2018), and the last edition of the TargetP application (TargetP 2.0) can also be enabled to calculate propensity scores technically. It has been shown, that TargetP2.0 shows an improved specificity to detect classical N-terminal MTS (Almagro Armenteros et al. 2019). However, due the divergent results predicting the iMTS-L profile of the well-studied Atp1, we believe that version 2.0 is not ideally

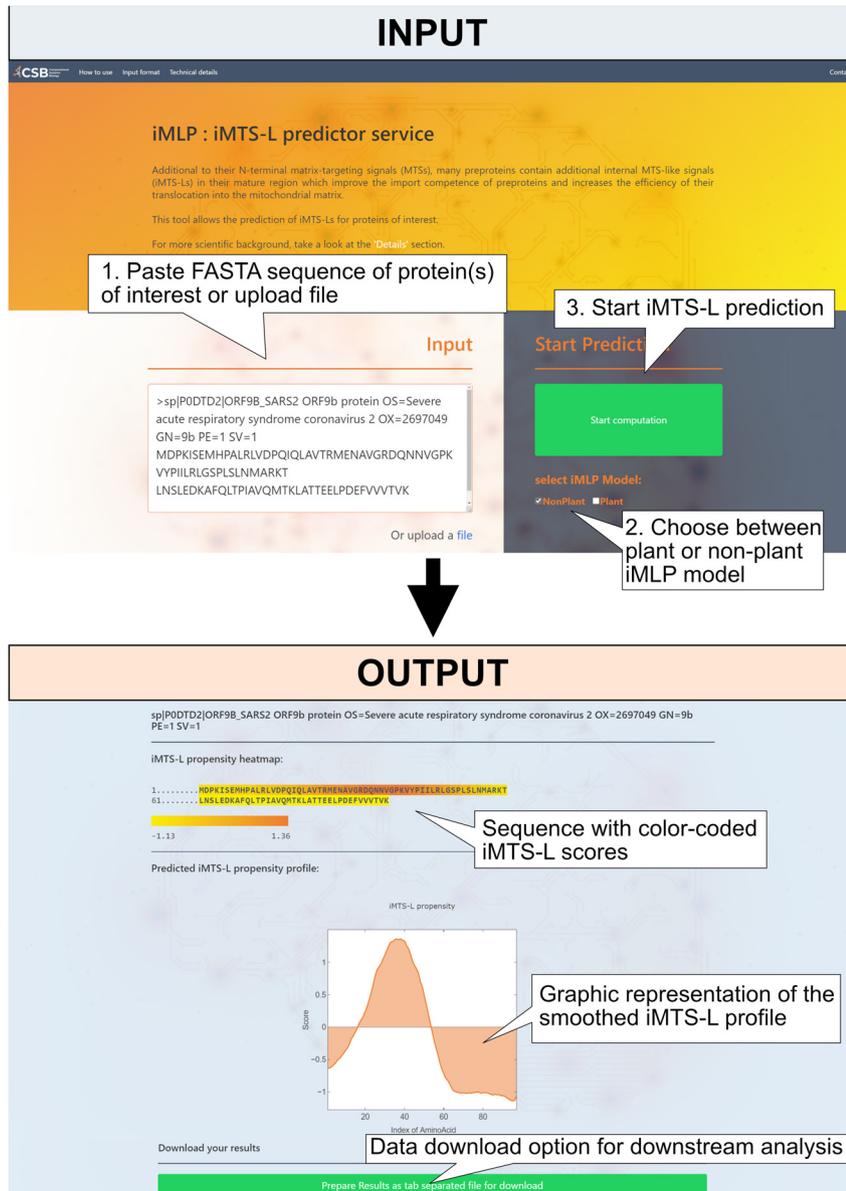


Figure 4: The iMLP window allows simple prediction of iMTS-L propensity scores. The figure shows a screen shot of the iMLP window. The FASTA sequence of the Orf9b protein of SARS-CoV2 was used for iMTS-L prediction. The output window indicates the iMTS-L probability scores as a heatmap-colored sequence as well as a profile. The Tom70 binding region of the Orf9b sequence (Jiang et al. 2020) is predicted with high probability score by iMLP.

suited to detect iMTS-Ls. This might be explained by slight differences in the physio-chemical properties between iMTS-Ls and classical MTS and the superior specificity of TargetP2.0 for the later.

In contrast, the iMLP predictor is easier to use, and makes tedious manual pre- and postprocessing obsolete. Moreover, the prediction speed is highly improved due to the thin network architecture, which makes the iMLP algorithm ideal for a web service-based prediction.

In vivo, the intracellular distribution of proteins is achieved by a number of biological reactions, including the binding of targeting factors, the insertion of proteins into translocation pores, the specific interactions with chaperones that drive protein translocation and the specific degradation of missorted proteins by proteases. Thus,

complex patterns presumably contribute to the intracellular distribution of proteins and the iMLP predictor will provide an exciting tool to study the ability of iMTS-L sequences to modulate intracellular protein targeting and to elucidate their prevalence in the proteomes of different organisms.

Materials and methods

Dataset and data processing

In order to prepare the test and training datasets, the first step is the generation of n suffixes s_i of the protein sequence, where n is the length of the original sequence, and

$$\{i \in \mathbb{Z} \mid 0 \leq i \leq n - 1\}$$

is the amount of amino acids truncated from the N-terminus of the original sequence to generate the suffix s_i . Subsequently, each suffix s_i is subject of TargetP MTS prediction (TargetP V1.1). The resulting prediction scores for each s_i are then concatenated to a positional score sequence vector S . Vector S is then smoothed using a Savitzky–Golay filter (Savitzky and Golay 1964) with the window size of the expected value of the distribution of known MTS lengths. This results in the iMTS-L probability profile $P_a^{\text{iMTS-L}}$, where a denotes the amino acid index in the sequence. For the detection of the actual iMTS-L region and to separate them from noise, a normalization step is necessary, which is realized by extracting iMTS-L profiles over ($n = 5000$) random sequences. The expected value $\mu^{\text{iMTS-L}}$ and the SD $\sigma^{\text{iMTS-L}}$ of this distribution is then used to normalize the iMTS-L probability profile and estimate the iMTS-L propensity score $P_a^{\text{iMTS-L}}$:

$$P_a^{\text{iMTS-L}} = \frac{P_a^{\text{iMTS-L}} - \mu^{\text{iMTS-L}}}{\sigma^{\text{iMTS-L}}}$$

With this normalization, the $P_a^{\text{iMTS-L}}$ score is 0 if the iMTS-L propensity at position a along the sequence is equal to that of a random sequence and e.g. 1 if it is one standard deviation more MTS-like. The described processing procedure could then be used to construct two learning data sets named in conjunction with the used targetP model “non plant” and “plant”. The “non plant” data set was based on the Uniprot reference proteomes of *Saccharomyces cerevisiae* (UP000002311), *Homo sapiens* (UP000005640) and *Mus musculus* (UP00000589) yielding a set of 48654 protein sequences with about 25 million iMTS-L propensity scores. The “plant” data set was constructed using the reference proteomes of *Arabidopsis thaliana* (UP000006548), *Zea mays* (UP000007305) and *Oryza sativa* subsp. *japonica* (UP000059680) with a total size of 110539 protein sequences yielding ~40 million iMTS-L propensity scores.

iMLP deep learning architecture

To setup a direct prediction of the iMTS-L propensity scores solely based on the protein sequence we employ state-of-the-art deep learning techniques using the Microsoft Cognitive Toolkit (CNTK, available at <https://github.com/Microsoft/CNTK>). In order to recover sequential dependencies between amino acids of the protein sequence, we chose a recurrent neural network architecture based on Long short-term memory (LSTM) cells. The model employs no constraints on the lengths of the protein sequences used as an input. Following an embedding layer (dimension = 24), transforming each amino acid symbol to a vector representation, we used two bidirectional recurrent layers. Each bidirectional layer consists of two spliced LSTM layers (dimension = 100) to implement both a negative and positive look-ahead. The output is realized as a single neuron dense layer.

Model training and evaluation

The model was optimized in a supervised fashion using the Adam optimizer (Kingma and Ba 2014) to minimize the squared error between the calculated $P_a^{\text{iMTS-L}}$ and the output of the final model layer as a loss function. Learning was carried out for 9000 epochs with a batch size of 2000, the learning rate was adjusted based on the

performance on the last epoch. The optimization was carried out on a randomly sampled subset consisting of 90% of the constructed learning data set, leaving the remainder of the learning data set to validate the model on previously unseen data (Figure 2A).

iMTS-L protein propensity estimation

To facilitate $P_a^{\text{iMTS-L}}$ the comparison across sequences, a iMTS-L protein propensity is calculated, that summarizes the iMTS-L profile information in a single score. This overall iMTS-L protein propensity score for a given sequence is calculated by averaging over all $P_i^{\text{iMTS-L}}$ that are higher than propensities of random sequences (meaning summing up non-negative normalized propensity scores of all amino acids of the sequence):

$$\text{iMTS-L} = \frac{s}{n} \text{ with } s = \sum_{a>\text{MTS}} \max(P_a^{\text{iMTS-L}}, 0) \text{ and } n = \sum_{a>\text{MTS}} [P_a^{\text{iMTS-L}} > 0]$$

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was funded by the TRR 175 (project D02) and the Landesforschungsschwerpunkt BioComp.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., Heijne, G., von, Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* 2: e201900429.
- Avendaño-Monsalve, M.C., Ponce-Rojas, J.C., and Funes, S. (2020). From cytosol to mitochondria: the beginning of a protein journey. *Biol. Chem.* 401: 645–661.
- Backes, S., Bykov, Y.S., Räschle, M., Zhou, J., Lenhard, S., Krämer, L., Mühlhaus, T., Bibi, C., Jann, C., Smith, J.D., et al. (2020). The mitochondrial surface receptor Tom70 protects the cytosol against mitoprotein-induced stress. *bioRxiv*, <https://doi.org/10.1101/2020.09.14.296194>.
- Backes, S. and Herrmann, J.M. (2017). Protein translocation into the intermembrane space and matrix of mitochondria: mechanisms and driving forces. *Front Mol Biosci* 4: 83.
- Backes, S., Hess, S., Boos, F., Woellhaf, M.W., Gödel, S., Jung, M., Mühlhaus, T., and Herrmann, J.M. (2018). Tom70 enhances mitochondrial preprotein import efficiency by binding to internal targeting sequences. *J. Cell Biol.* 217: 1369–1382.
- Boos, F., Mühlhaus, T., and Herrmann, J. (2018). Detection of internal matrix targeting signal-like sequences (iMTS-Ls) in mitochondrial precursor proteins using the TargetP prediction tool. *Bio-Protocol* 8: e2474.
- Brix, J., Rüdiger, S., Bukau, B., Schneider-Mergener, J., and Pfanner, N. (1999). Distribution of binding sequences for the mitochondrial import receptors Tom20, Tom22, and Tom70 in a presequence-

- carrying preprotein and a non-cleavable preprotein. *J. Biol. Chem.* 274: 16522–16530.
- Brix, J., Ziegler, G.A., Dietmeier, K., Schneider-Mergener, J., Schulz, G.E., and Pfanner, N. (2000). The mitochondrial import receptor Tom70: identification of a 25 kDa core domain with a specific binding site for preproteins. *J. Mol. Biol.* 303: 479–488.
- Calvo, S.E., Julien, O., Clauser, K.R., Shen, H., Kamer, K.J., Wells, J.A., and Mootha, V.K. (2017). Comparative analysis of mitochondrial N-termini from mouse, human, and yeast. *Mol Cell Proteom MCP* 16: 512–523.
- Chacinska, A., Guiard, B., Müller, J.M., Schulze-Specking, A., Gabriel, K., Kutik, S., and Pfanner, N. (2008). Mitochondrial biogenesis, switching the sorting pathway of the intermembrane space receptor Mia40. *J. Biol. Chem.* 283: 29723–29729.
- Emanuelsson, O., Brunak, S., Heijne, G. von, and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2: 953–971.
- Emanuelsson, O., Nielsen, H., Brunak, S., Heijne, G. von. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300: 1005–1016.
- Friedl, J., Knopp, M.R., Groh, C., Paz, E., Gould, S.B., Herrmann, J.M., and Boos, F. (2020). More than just a ticket canceller: the mitochondrial processing peptidase tailors complex precursor proteins at internal cleavage sites. *Mol. Biol. Cell* 31: 2657–2668.
- Gordon, D.E., Hiatt, J., Bouhaddou, M., Rezelj, V.V., Ulferts, S., Braberg, H., Jureka, A.S., Obernier, K., Guo, J.Z., Batra, J., et al. (2020). Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* 370: eabe9403.
- Grevel, A., Pfanner, N., and Becker, T. (2019). Coupling of import and assembly pathways in mitochondrial protein biogenesis. *Biol. Chem.* 401: 117–129.
- Heijne, G. von. (1986). Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J.* 5: 1335–1342.
- Jiang, H.-W., Zhang, H.-N., Meng, Q.-F., Xie, J., Li, Y., Chen, H., Zheng, Y.-X., Wang, X.-N., Qi, H., Zhang, J., et al. (2020). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell. Mol. Immunol.* 17: 998–1000.
- Kingma, D.P. and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv*, <http://arxiv.org/pdf/1412.6980v9>.
- Kreimendahl, S. and Rassow, J. (2020). The mitochondrial outer membrane protein tom70-mediator in protein traffic, membrane contact sites and innate immunity. *Int. J. Mol. Sci.* 21: 7262.
- Mani, J., Desy, S., Niemann, M., Chanfon, A., Oeljeklaus, S., Pusnik, M., Schmidt, O., Gerbeth, C., Meisinger, C., Warscheid, B., et al. (2015). Mitochondrial protein import receptors in Kinetoplastids reveal convergent evolution over large phylogenetic distances. *Nat. Commun.* 26: 6646.
- Mårtensson, C.U., Priesnitz, C., Song, J., Ellenrieder, L., Doan, K.N., Boos, F., Floerchinger, A., Zufall, N., Oeljeklaus, S., Warscheid, B., et al. (2019). Mitochondrial protein translocation-associated degradation. *Nature* 569: 679–683.
- Miserey-Lenkei, S., Trajkovic, K., D'Ambrosio, J.M., Patel, A.J., Čopič, A., Mathur, P., Schauer, K., Goud, b., Albanèse, V., Gautier, R., et al. (2021). A comprehensive library of fluorescent constructs of SARS-CoV-2 proteins and their initial characterization in different cell types. *Biol Cell*. <https://doi.org/10.1111/boc.202000158>.
- Nickel, C., Horneff, R., Heermann, R., Neumann, B., Jung, K., Soll, J., and Schwenkert, S. (2019). Phosphorylation of the outer membrane mitochondrial protein OM64 influences protein import into mitochondria. *Mitochondrion* 1: 93–102.
- Samavarchi-Tehrani, P., Abdouni, H., Knight, J.D., Astori, A., Samson, R., Lin, Z.-Y., Kim, D.-K., Knapp, J.J., St-Germain, J., Go, C.D., et al. (2020). A SARS-CoV-2 – host proximity interactome. *bioRxiv*, 2020.2009.2003.282103.
- Savitzky, A. and Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36: 1627–1639.
- Shi, C.-S., Qi, H.-Y., Boularan, C., Huang, N.-N., Abu-Asab, M., Shelhamer, J.H., and Kehrl, J.H. (2014). SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. *J. Immunol.* 193: 3080–3089.
- Vögtle, F.-N., Wortelkamp, S., Zahedi, R.P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N., et al. (2009). Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* 139: 428–439.
- Yamamoto, H., Fukui, K., Takahashi, H., Kitamura, S., Shiota, T., Terao, K., Uchida, M., Esaki, M., Nishikawa, S., Yoshihisa, T., et al. (2009). Roles of Tom70 in import of presequence-containing mitochondrial proteins. *J. Biol. Chem.* 284: 31635–31646.
- Young, J.C., Hoogenraad, N.J., and Hartl, F. (2003). Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. *Cell* 112: 41–50.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/hsz-2021-0185>).