**dr inż. Umesh Kalathiya**

*The field of **Natural sciences***
*The discipline of **Biological sciences***

Molecular basis for innovative strategies of vaccine or drug development by exploring structural features of different functional proteins

_____

**Summary of Professional Accomplishments**

International Centre for Cancer Vaccine Science, University of Gdańsk
ul. Kładki 24, 80-822 Gdańsk, Poland

**Gdansk 2022**

1. **Name:** Umesh Kalathiya

2. **Diplomas, degrees conferred in specific areas of science or arts, including the name of the institution which conferred the degree, year of degree conferment, title of the PhD dissertation**

2018    Ph.D. (dr inż.) in chemical sciences in the field of Biotechnology, obtained at Department of Pharmaceutical Technology and Biochemistry, Faculty of Chemistry, Gdańsk University of Technology, Poland. My doctoral dissertation is entitled: *Molecular properties of TRF1 and TRF2 proteins dimer forming TRFH domains and their interactions with TIN2 or Apollo peptides*, supervisor: prof. dr. hab. inż. Maciej Bagiński.

2012    Master of Science and Engineering (mgr inż.) degree in Biotechnology / Bioinformatics, Faculty of Chemistry, Wrocław University of Science and Technology, Wrocław, Poland. My master's thesis is entitled: *Design of the low molecular inhibitors of Cathepsin C*, supervisors: prof. dr hab. inż. Wacław Andrzej Sokalski and dr. inż. Michał Jewgiński.

3. **Information on employment in research institutes or faculties / departments or school of arts**

From 2020-present    Principal Investigator, of the grant funded by The National Science Centre (NCN), Poland. International Centre for Cancer Vaccine Science (ICCVS), University of Gdańsk, Poland

2019-2020    Adiunkt / Postdoctoral researcher, at International Centre for Cancer Vaccine Science (ICCVS), University of Gdańsk, Poland.

4. **Description of the achievements, set out in art. 219 para 1 point 2 of the Act**

A monothematic cycle of scientific achievement includes 7 scientific publications (6 scientific research papers and 1 review publication), and the points are received for them according to the article 219 paragraph 1 point 2 of the Act on Academic Degrees and Titles and on Degrees and Titles in Art. They are a series of 7 related publications conceptualizing the structural and functional aspects of proteins / enzymes providing novel insights in the field of innovative vaccine design or drug development. These works describing basic and application studies are the outcome of strong interdisciplinary collaboration with academic research universities, and they are published between 2019-2021 year.

I am the first or a senior author in all 7 publications constituting the scientific achievement, as well as I am the corresponding author. These papers constituted in the scientific achievement have a total value of Impact Factor (IF) of 35.744. The total number of MEiN (Ministerstwo Edukacji i Nauki, or (MNiSW) Ministry of Science and Higher Education) points awarded for

the achievement is 860, according to the announcement of the Ministry of Education and Science, Republic of Poland.

## a) The title of the scientific achievement

Molecular basis for innovative strategies of vaccine or drug development by exploring structural features of different functional proteins.

## b) The research publications belonging to the Scientific Achievement

**IF** – impact factor from the year of publication of the work, and if the IF was not present for the specific year, the IF of the previous year was used. **MEiN** – journal scoring according to the Ministry of Education and Science / Ministerstwo Edukacji i Nauk. Publications from previous **MEiN** scoring points, were scored according to the new **MEiN** scores.

**4.1.** Padariya, M., Fahraeus, R., Hupp, T., **Kalathiya, U.** (2021). Molecular determinants and specificity of mRNA with alternatively-spliced UPF1 isoforms, influenced by an insertion in the "regulatory loop". International Journal of Molecular Sciences, 22(23), 12744. https://doi.org/10.3390/ijms222312744; (**IF$_{2020}$ 5.92; MEiN$_{2021}$ = 140**).

**4.2.** Padariya, M., Kote, S., Mayordomo, M., Dapic, I., Alfaro, J., Hupp, T., Fahraeus, R., **Kalathiya, U.** (2021). Structural determinants of peptide-dependent TAP1-TAP2 transit passage targeted by viral proteins and altered by cancer-associated mutations. Computational and Structural Biotechnology Journal, 19, 5072–5091. https://doi.org/10.1016/j.csbj.2021.09.006; (**IF$_{2020}$ 7.27; MEiN$_{2021}$ = 100**).

**4.3. Kalathiya, U.**, Padariya, M., Pawlicka, K., Verma, C. S., Houston, D., Hupp, T. R., Alfaro, J. A. (2019). Insights into the effects of cancer associated mutations at the UPF2 and ATP-binding sites of NMD master regulator: UPF1. International Journal of Molecular Sciences, 20(22), 5644. https://doi.org/10.3390/ijms20225644; (**IF$_{2019}$ 4.556; MEiN$_{2019}$ = 140**).

**4.4. Kalathiya, U.,** Padariya, M., Baginski, M. (2019). Structural, functional, and stability change predictions in human telomerase upon specific point mutations. Scientific Reports, 9(1), 8707. https://doi.org/10.1038/s41598-019-45206-y; (**IF$_{2019}$ 3.998; MEiN$_{2019}$ = 140**).

**4.5. Kalathiya, U.,** Padariya, M., Mayordomo, M., Lisowska, M., Nicholson, J., Singh, A., Baginski, M., Fahraeus, R., Carragher, N., Ball, K., Haas, J., Daniels, A., Hupp, T. R., Alfaro, J. A. (2020). Highly conserved homotrimer cavity formed by the SARS-CoV-2 spike glycoprotein: A novel binding site. Journal of Clinical Medicine, 9(5), 1473. https://doi.org/10.3390/jcm9051473; (**IF$_{2020}$ 4.24; MEiN$_{2020}$ = 140**).

**4.6. Kalathiya, U.,** Padariya, M., Faktor, J., Coyaud, E., Alfaro, J. A., Fahraeus, R., Hupp, T. R., Goodlett, D. R. (2021). Interfaces with structure dynamics of the workhorses from cells revealed through cross-linking mass spectrometry (CLMS). Biomolecules, 11(3), 382. https://doi.org/10.3390/biom11030382; (**IF$_{2020}$ 4.88; MEiN$_{2021}$ = 100**).

**4.7. Kalathiya, U.,** Padariya, M., Fahraeus, R., Chakraborti, S., Hupp, T. R. (2021). Multivalent display of SARS-CoV-2 spike (RBD domain) of COVID-19 to nanomaterial, protein ferritin nanocages. Biomolecules, 11(2), 297. https://doi.org/10.3390/biom11020297; (**IF$_{2020}$ 4.88; MEiN$_{2021}$ = 100**).

The publications number **4.1**, **4.2**, and **4.6,** describe results obtained during the implementation of the research projects in which I am the **Principal Investigator** (title: Specificity in detection of PTCs in mRNA by NMD and its network, insights from cancer perspective and cross-linking). SONATINA, National Science Center, Poland (grant agreement no. 2020/36/C/NZ2/00108), 2020-2023 – project in progress.

**c) The discussion of the scientific objective of the above-mentioned work and the results achieved, together with a discussion of their application**

**Introduction:**

Cancer incidence increases worldwide mostly due to aging of the population in developed countries and improving healthcare with rising awareness in developing, as well as underdeveloped countries. Consequently, novel vaccine design strategies and finding new anticancer drug candidates need to be constantly developed. Cancer represents a significant burden on society and any new therapeutic drug that impacts patient prognosis is directly beneficial for society. Personalized medicine requires an increasing repertoire of drugs able to target more exclusive subpopulations of cancer patients, which improves the efficiency and safety of treatment for cancer patients. The activity spectrum of investigated structural properties of different proteins, and proposed molecules in my presented scientific achievements, provides contribution to treat malignancies and viral infection. Malignancies are the second cause of death in Poland alone, and due to the current global COVID-19 (coronavirus disease 2019) pandemic over 6 million deaths have been reported globally (as of 17 April 2022, over 500 million confirmed cases; www.who.int). About 26% of male and 23% of female deaths can be directly related to cancer in Poland (Wojciechowska et al., Cancer in Poland in 2010, National Cancer Registry). In 2016, the number of reported new cases of malignant tumors in Poland was 164140 (http://onkologia.org.pl). Moreover, SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) infected total cases in Poland as of 19 April 2022 were 5.99M and 116K deaths (https://ourworldindata.org/).

Majority of the gene mutations found in genetic disorders, including cancer, result in premature termination codons (PTC), and the rapid degradation of their mRNAs by nonsense-mediated mRNA decay (NMD) [1, 2]. Human genetic disorders are caused by diverse types of mutations including nonsense mutations, frameshift mutations, and mutations that cause alternative splicing events, resulting in PTC [2]. It has been estimated that up to 30% of all mutations resulting in human genetic disorders result in PTCs [1]. Many transcripts carrying a PTC are targeted for rapid degradation before they can be translated into protein through a multistep process termed as the nonsense-mediated mRNA decay. The molecular mechanism of NMD has not been fully delineated, though working models have been proposed. When the

translation complex pauses at a PTC that is upstream point of the multiprotein exon junction complex (EJC), and the eukaryotic release factors (eRFs) recruit the RNA helicase UPF1 (a vital component of the NMD mechanism) [3]. The phosphorylation and dephosphorylation of UPF1 are necessary steps before the degradation of the transcript by exonucleases [4]. Hence, the NMD process is termed as the quality-control checkpoint that detects and eliminates aberrant messenger RNAs (mRNAs) with premature termination codons [5]. The NMD machinery involves the UPFs (UPF1, UPF2, and UPF3a/b) proteins, and includes the suppressors with morphological effects on genitalia proteins (SMG1, SMG5, SMG6, SMG7, SMG8, and SMG9), these proteins are bound to mRNAs via the exon junction complex [6].

Understanding the interacting partners or the components from the NMD, EJC, and eRFs as well as their involvement in different biological pathways suggests that proteins from these complexes are involved in several critical and important functions of the human body. The components from these complexes are widely involved in the regulation of telomere maintenance, regulation of chromosome organization, and RNA transport / localization / catabolic process, in addition to the nonsense-mediated mRNA decay pathway. Moreover, analyzing the overall patient survival status relating to the components from NMD, EJC, and eRFs, suggest that there is substantial decrease in the survival for the patients from the altered group. Looking over the mutational landscape of NMD, EJC, and eRFs components from different cancer types suggest that these proteins are heavily mutated in several cancers. Therefore, this suggests that detailed understanding about the protein-protein and protein-RNA or RNA-RNA network is crucial. Investigating the mutational effect on the stability and binding of protein-protein / RNA and RNA-RNA can provide new insights into the molecular mechanisms of this complex process, as well as assist in development of novel therapeutic (vaccine or drug development) approaches. Targeting different NMD components with drug / small molecules, generating translation readthrough thus creating neoantigen by virtue of the amino acid, which is added at a stop codon, in a sense stimulates the tumour cell to make its own neoantigen vaccine. The recognition and detection of PTCs by different components needs detailed understanding since production of proteins with dominant negative functions results in severe genetic diseases and cancer.

Blocking the NMD pathway to make tumour cells to produce its own neoantigen vaccine, and these neoantigen peptides can go through the standard proposed antigen transport and presentation process of the peptide-loading complex (PLC). The cytotoxic T lymphocytes response to infected or transformed cells depends on the presentation of peptide antigens on major histocompatibility complex class I (MHC-I) molecules. The majority of MHC-I presented peptides are generated in the cytosol by the proteasome and transported into the endoplasmic reticulum (ER) lumen for further processing by the ERAP aminopeptidases, before loaded onto MHC-I molecules by the dynamic multi-component assembly peptide-loading complex [7]. PLC includes TAP1 and TAP2 (transporter 1 and 2) heterodimer complex, MHC-I, β2microglobulin (β2m), the chaperones tapasin, calreticulin, and ERp57 (ER protein 57) [8]. The TAP transporters constitute functionally key components of the antigen presentation pathway since they link the cytosolic pool of peptides with the PLC and the ER-resident MHC-I molecules [7].

Apart from the components of the NMD machinery that are widely involved in the regulation of telomere maintenance, the telomerase (a ribonucleoprotein; RNP) enzyme acts as a reverse transcriptase (hTERT) that synthesizes telomeric DNA repeats at the ends of chromosomes [9]. Telomerase is active in the early stages of life, maintaining the telomere length, and it becomes inactive in most somatic cells during adulthood [10]. The ability of a telomere to provide genomic stability decreases over time, owing to both the natural loss of telomeric structure with each cell division (the end replication problem) and the loss of telomerase activity. In addition, this process leads to aging [11]. However, in cancer cells, telomerase becomes reactivated, and it continuously maintains the short length of telomeres in rapidly dividing cells, leading to their immortality [12]. Most cancers have adapted mechanisms to protect lengths of telomeres, and this protection is achieved by telomerase activation in approximately 90% of human cancers. Therefore, the role of telomerase in cancer and aging makes it an important target for cancer therapies and age-associated disorders [13]. Analyzing the effect of the mutations over hTERT structure at the molecular level and provided valuable insights into the nature of potential structural changes because of mutations, especially at the functionally important regions or residues of the active site. These active sites could be used as a target for drug development pipelines, in the structure-based virtual screen approach (SBVS). Moreover, hTERT is highly expressed in ~90% of cancer cells, and thus, is a potential target for tumour vaccines [14]. Antigenic peptides derived from hTERT that were predicted to bind to MHC class II molecules were found to be able to induce primary human T-cell *responses in vitro* in breast cancer, melanoma, and leukemia [14].

Moreover, there is an urgent need for developing safe and effective therapeutics against SARS-CoV-2, and the most promising strategy to fight against this deadly virus is to develop an effective vaccine. At present there are no robust drugs for wide-spread dissemination available against coronaviruses including the SARS-CoV-2 virus. Therefore, I applied my optimized molecular dynamics and vaccine / drug development pipelines against the SARS-CoV-2 spike (S) glycoprotein, investigating structural properties of the protein. My scientific achievements provide a platform for drug discovery or vaccine design by elaborating the structure of S protein that will improve responsiveness in the next pandemic. In addition, to some novel compounds, several other known compounds were traced, which are already proposed for clinical trials including an mTOR (mammalian target of Rapamycin) pathway inhibitor, Sirolimus (Rapamycin; a macrolide) [14] and Ritonavir [15]. The S protein receptor-binding domain (RBD) plays an important role in virus transmission, and hence, it is one of the key molecular targets for vaccine development. Efforts were made to develop a spike RBD-based subunit vaccine utilizing a ferritin protein nanocage as a scaffold. Several fusion protein constructs were designed *in silico* by connecting the spike RBD via a synthetic linker (different sizes) to different ferritin subunits (H- and L-ferritin) or nanoparticles [16, 17].

Significant importance of my scientific achievements was to investigate structural properties of proteins **assisting in the development of novel therapeutic such as vaccine or drug development strategies.** The publication number 4.1, 4.2, and 4.3, highlight findings supporting the neoantigen development strategies, cancer-derived mutant peptide molecules presented over the MHC molecules. Identifying different structural dynamics of the UPF1 protein from the NMD pathway and its active-site residues or binding pockets are crucial steps

for the structural-based virtual screening approach. Using this information, a set of novel medicinal compounds can be screened, which could block the UPF1 activity, and case read-through of the PTC containing mRNA, resulting in the production of mutation peptides or protein. Findings from publication 4.2, contribute better understanding of the transport of such cancer-derived mutant peptides, and kinetics of peptide transport and presented over the MHC molecules. Publication 4.4, represents understanding the effect of cancer-mutants over the telomerase enzyme (hTERT) highly expressed in >90% of cancer cells, and thus, is a potential target for tumor vaccines. Different experimental and informatics techniques are known to investigate the structural properties of protein molecules, and they are highlighted in the publication no. 4.6. **In parallel, the publications no. 4.5 and 4.7, demonstrate important structural dynamics of the S glycoprotein** trimer cavity targeted by known medical compound, and the receptor-binding domain represented over the ferritin nanoparticles proposing its use in novel vaccine development strategy.

**Detailed description of the publications included in the scientific achievement:**

**(i)** Padariya, M., Fahraeus, R., Hupp, T., **Kalathiya, U.** (2021). Molecular determinants and specificity of mRNA with alternatively-spliced UPF1 isoforms, influenced by an insertion in the "regulatory loop". International Journal of Molecular Sciences, 22(23), 12744.

*Aim of the study:*

The UPF1 protein is a key component in the NMD pathway, also termed as the master regulator. Deletion of UPF1 in yeast not only stabilizes the nonsense-containing mRNAs but also enhances nonsense codon read-through as well and inhibits degradation of prematurely terminated polypeptides [18]. It has been shown that deletion of UPF1 induces the production of (approximately 6- to 8-fold increase) novel peptide read-through [19]. Interestingly, two alternatively spliced isoforms of UPF1 that differ only in length of the 'regulatory loop' exist in mammals [20]. UPF1 isoform_1 contains 11 aa (amino acids) insertion (353-GNEDLVIIWLR-363) in the domain 1B, which extends the 'regulatory loop' to 22 aa. Whereas the UPF1 isoform_2 (the more abundant UPF1 short loop isoform) is composed of only an 11 aa 'regulatory loop' which weakens the UPF1 affinity for mRNA. In addition, this insertion of 11 amino acids to the 'regulatory loop' can considerably increase the catalytic and/or ATPase activity and result in a higher affinity for mRNA.

Considering such influence of the 'regulatory loop' in the structural dynamics of UPF1 and its control over the catalytic and/or ATPase activity, **I investigated both isoforms of UPF1 in the presence or absence of the poly(U) mRNA** by applying the molecular dynamics simulation (MDS) technique. Additionally, it is known that UPF1 can bind with different mRNAs efficiently, but the molecular details are still not clear, and therefore, aim of this project was to **understand the UPF1 isoform_1 (having a longer 'regulatory loop') and isoform_2 binding with different mRNA motifs.** Moreover, significant structural changes upon inserting the most frequently occurring **cancer mutations in the UPF1 isoform_1 bound with mRNA motifs were investigated by me**. Furthermore, preferential binding affinity of UPF1 with mRNAs in a 3'UTR length-dependent manner has been reported [21]. UPF1-mRNA

interactions is known *in vivo* [22], that the UPF1 protein promiscuously interacts with mRNA before translation. In addition, the traced preferential binding with 3′UTRs for UPF1 at steady state originate from the selective displacement of UPF1 from coding regions by translating ribosomes [22]. The phosphorylated UPF1 mRNA footprint data and CLIP (cross-linking immunoprecipitation)-Seq of UPF1 *in vivo* identified UPF1 binding sites across transcriptome, describing UPF1 targets 3′ UTR GC-rich motifs [21]. Particularly, in 3′ UTR the GC-rich motifs (5′-CCUGGGG-3′, 5′-CCUGGGA-3′, 5′-CCUGGAA-3′, and 5′-CCUGAGA-3′) can be the target motifs in UPF1-dependent mRNA decay [23]. Due to such specificity of UPF1 towards a particular class of mRNA motifs, and to understand the UPF1-mRNA dynamics, I performed MD simulation of **UPF1 protein with different GC-rich RNA motifs, as well as made a comparative analysis with the AU-rich motifs** (5′-UUUUUUU-3′, 5′-UUAAUUU-3′, 5′-UUAGUUU-3′, and 5′-UUGAUUU-3′; Fig. 1a).
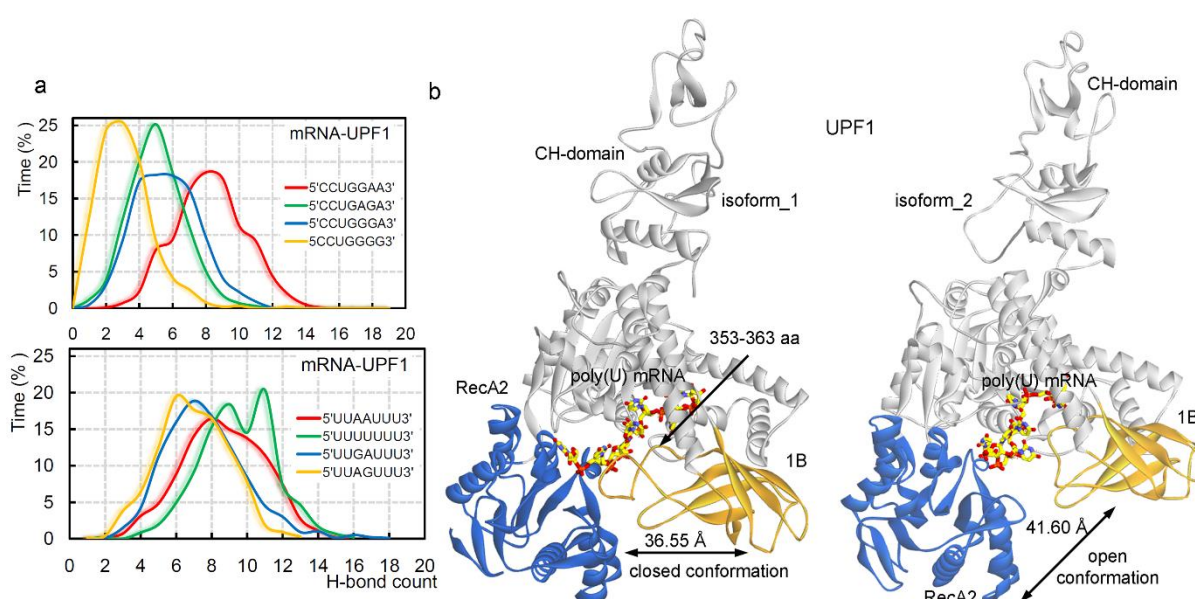


**Fig. 1.** Adopted binding patterns by UPF1 (isoform_1) with different mRNA motifs. **(a)** Frequency of the number of hydrogen bond (H-bond) interactions formed between UPF1 (isoform_1) and different GC-rich (5′-CCUGGGG-3′, 5′-CCUGGGA-3′, 5′-CCUGGAA-3′, and 5′-CCUGAGA-3′) or AU-rich (5′-UUUUUUU-3′, 5′-UUAAUUU-3′, 5′-UUAGUUU-3′, and 5′-UUGAUUU-3′) motifs. **(b)** The structural dynamics of the 'regulatory loop' in presence (isoform_1) and absence of (isoform_2) of the 11 aa (353-GNEDLVIIWLR-363) insertion.

***Description of results:***

11 aa (353-GNEDLVIIWLR-363) insertion / deletion in the 'regulatory loop' is responsible for the 'closed (isoform_1)' / 'open (isoform_2)', conformations between 1B and RecA2 domains, that could have impacted the catalytic activity of the UPF1 helicase **(analysis of the structural properties and data from MD simulations was performed by me)**. These two distinct conformations of UPF1 correspond to the important stacking pattern observed in the mRNA motifs, i.e., **the absence of a stacking formation in isoform_2 results in an 'open conformation'**. The amino acids E355, N354, and R363 from the insertion region for UPF1 isoform_1 induced interactions with the mRNA motif. Binding affinities of **GC- and AU-rich mRNA motifs with UPF1 highlighted that GC-rich mRNAs have better binding** with the protein in most studied motifs in our project (Fig. 1a). In addition, when complexed with the UPF1 protein, the GC-rich motifs are more stable compared to AU-rich motifs. MD simulation

data revealed that as the distance between 1B-RecA2 domain increases forming an 'open conformation', it has a decline in the UPF1-mRNA intermolecular interactions. e.g., UPF1 with 5′-CCUGGAA-3′ motifs had less distance between 1B-RecA2 domains (closed conformation) and higher mRNA–UPF1 interactions (Fig. 1b), whereas contrary behavior was observed for the 5′-CCUGGGG-3′ (open conformation) sequence. Among the AU-rich motifs, the poly(U) mRNA had the highest binding affinity with the UPF1 protein, and the least binding affinity was the 5′-UUAGUUU-3′ motif. **A few high occupancy interacting residues R363, D364, T627, and G862 from UPF1 isoform_1 were common binders among the GC-rich motifs, and such residues were R363, N535, and T627 among the AU-rich motifs**. Furthermore, investigating the cancer-associated mutations in UPF1 revealed that the mRNA motif with UPF1$_{P533T}$ lacks the stacking interactions, due to which the distance between 1B-RecA2 domains increased and the intermolecular protein–mRNA interactions showed a declining trend.

*The importance of the results:*

**Two alternatively spliced isoforms of UPF1 exist in mammals and they differ only in length of the 'regulatory loop' involved in binding with mRNA.** As this master regulator of NMD pathway is an RNA-dependent ATPase/helicase, any influence in structural properties could have direct impact on entire PTC detection in mRNA and degradation process. The data from MD simulation probe the binding specificity of UPF1 with different GC- and AU-rich mRNA motifs, indicating two distinct conformations between 1B and RecA2 domains of UPF1: 'open (isoform_2; without insertion)' and 'closed (isoform_1; with insertion; Fig. 1b)' conformation. These structural movements correspond to an important stacking pattern in mRNA motifs, i.e., absence of stack formation in mRNA, with UPF1 isoform_2 results in the 'open conformation'. Particularly, for UPF1 isoform_1, the increased distance between 1B and RecA2 domains has resulted in reducing the mRNA–UPF1 interactions. Such novel perspectives from mRNA–protein binding pairs identified in this work shall contribute to understanding the selectivity of respective partners, along with advancing NMD-associated structural dynamics and kinetics.

**(ii)** Padariya, M., Kote, S., Mayordomo, M., Dapic, I., Alfaro, J., Hupp, T., Fahraeus, R., **Kalathiya, U.** (2021). Structural determinants of peptide-dependent TAP1-TAP2 transit passage targeted by viral proteins and altered by cancer-associated mutations. Computational and Structural Biotechnology Journal, 19, 5072–5091.

*Aim of the study:*

Substantial effort has been made to understand the peptides that are translocated by the TAP1-TAP2 transporters and presented over the MHC molecules. Several biochemical studies have identified the nucleotide binding structural motifs, roughly mapped the peptide binding region, and provided the secondary structures responsible for the crosstalk between ATP hydrolysis in the TAP$_{NBDs}$ (nucleotide-binding domains) and conformational changes of the TM (transmembrane) helices [7, 24, 25]. However, the picture is still incomplete, as it has not been conclusively integrated into a robust trajectory of how the TAP dynamics and the substrates move to carry out a transport cycle. The precise conformational changes that occur because of substrate or nucleotide binding to TAP1 or TAP2 are also not characterized. However, it has

also been suggested that the transporters most likely do not undergo the same structural cycles and no common mechanism has been proposed. Impaired TAP function results in reduced surface expression of MHC-I, as the empty MHC-I molecules are unable to present antigens to the immune system and is a common target for viral as well as cancer immune evasion. **Considering these details, we investigated the TAP1 and TAP2 transporters at the atomic level with a particular focus on how viruses and cancer-derived mutations can inhibit TAP-mediated peptide translocation at a molecular level.** At first, along with dr inż. Monikaben Padariya (from University of Gdansk), I took part in constructing different viral protein structures and peptides structures, as well as did screening of viral proteins and peptides with the TAP1-TAP2 transporters.

Since we were interested to know if the transport of peptides through the TAP is affected by its length or its amino acid composition, we collected different peptide datasets derived from the immunopeptidome approach (Fig. 2a) and the ATLAS (Altered TCR Ligand Affinities and Structures) database. These peptides were *in silico* screened with the TAP1-TAP2 transporters, tracing **protein-peptide binding affinities and kinetics of peptide transport.** The cancer-associated mutations and viral proteins may share a common trend towards TAP transporters, which is attenuation of their activity. Defective TAP transporters in few tumor tissues have been identified by down-regulation of TAP mRNA or mutations. Therefore, **I determined the effects of point mutations on TAP1-TAP2 plasticity, tracing stability and binding affinity changes with respect to each other**. In addition, **I did proteomics analysis (bioinformatics approaches) for the INF-ɣ treated immunopeptidome datasets** (Fig. 2a), and identified possible peptides presented over the cell-surface MHC-I molecules. Moreover, I performed MD simulations inserting different cancer mutations in the TAP1 and TAP2 transporters (Fig. 2b) with different peptides, **scaling effects of somatic mutations on the transporter process.**
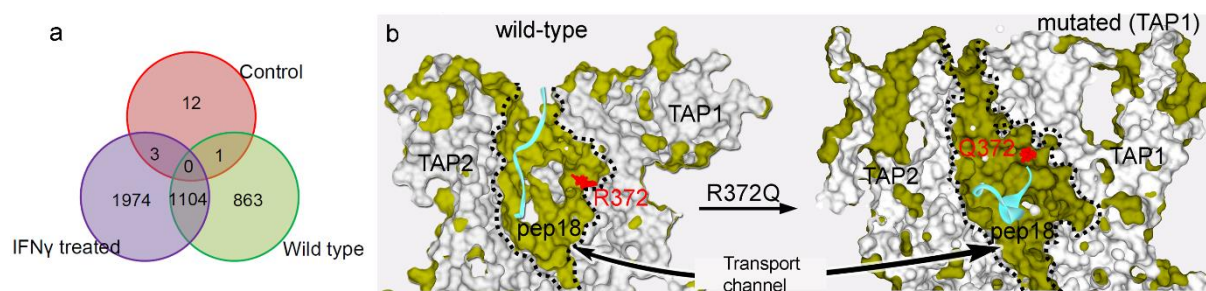


**Fig. 2.** Structural rearrangements of TAP1-TAP2 transporters examined in the presence of peptide or cancer-associated variants. **(a)** Venn diagram representing the common and unique peptides in all three/two respective IFN-γ (interferon-γ) samples. **(b)** Mutation in the TAP transport channel (shown in green) can block the peptide in the cavity, for example, TAP1$_{R372Q}$ hinders the peptide movement towards the ER (endoplasmic reticulum) lumen.

### *Description of results:*

To investigate the kinetics of the peptide transport process, different immunopeptidome datasets were generated from the melanoma A375 cells. Higher numbers of peptides were identified in the IFN-γ (interferon-γ) treated melanoma cells (3081 peptides) compared to that of the untreated cells (1968 peptides), suggesting an induced anti-cancer immunity pathway is being activated. The TAP1 or TAP2 amino acids binding with peptides in fingerprint analysis,

were also found mutated in different cancer types, which suggest that point mutation either **from TAP1 or TAP2 is sufficient to alter the protein-protein binding which may affect the peptide transport process**. Interestingly, the variants located at the peptide binding pocket of the TAP1-TAP2 complex have diverse conformations; the $TAP1_{R372Q}$ variant was found highly stable compared to that of the $TAP2_{R373H}$. The mutations in TAP1 may completely block both ER (endoplasmic reticulum) and cytosolic passage of the TAP transporter which may hinder the intake and evacuation of the peptides from cytosol-to-ER (**blocks the IF-to-OF conformation; inward- to outward-facing**), whereas mutation in the TAP2 protein mostly blocks the ER pore only. The mutated system $TAP1\text{-}SIINFEKL\text{-}TAP2_{R373H}$ had a higher number of protein-protein interactions, and the $TAP1_{R372Q}\text{-}pep18\text{-}TAP2$ complex is one of the lowest to form such binding (Fig. 2b). Mutation in the $TAP1_{R372Q}$ and $TAP2_{R373H}$ blocks the movement of the peptide upwards in the transport channel. The opening of transport cavity pore towards the ER lumen in the wild type system were traced, whereas the structural conformations of peptides in the presence of the **mutation systems suggest that the cancer mutation induces rigidity in the TAP1-TAP2 complex** and therefore, lacks the 'open' conformation towards the ER lumen.

### *The importance of the results:*

Single cancer-associated mutations can significantly affect the heterodimer formation of the TAP transporters, as well as the peptide transport kinetic process. Overall, considering the binding affinities of the peptide with transporters it could be proposed that smaller peptides (8, 9, or 10 mers) may transport comparatively quickly through the transport channel. Furthermore, by adversely interacting with the TAP transport passage, or affecting the $TAP_{NBD}$ tilt movement, viral proteins and cancer-derived mutations may induce allosteric effects or rigidity in the TAP complex blocking the IF-to-OF conformation of the tunnel. The areas from the TAP transporter towards ER-opening and the distance centre of mass between $TAP_{NBD}$ domains, highlights a difference in the transport of longer vs shorter peptides. However, the sequence composition of peptides can affect the rate of such conformations by the transporters. Findings from this project propose a model for **how different viral factors and cancer-derived mutations alter the peptide transport** process by directly targeting the TAP proteins, and how **cancer-derived mutations in TAP control the peptide transport that could have been involved in the MHC-I antigen presentation**. Moreover, these observations guide to understand the mechanism of how neoantigen peptides may be transported through TAP transporter.

**(iii) Kalathiya, U**., Padariya, M., Pawlicka, K., Verma, C. S., Houston, D., Hupp, T. R., Alfaro, J. A. (2019). Insights into the effects of cancer associated mutations at the UPF2 and ATP-binding sites of NMD master regulator: UPF1. International Journal of Molecular Sciences, 20(22), 5644.

### *Aim of the study:*

The conserved cysteine-histidine-rich domain (CH-domain) of UPF1 protein from the NMD pathway is displaced upon UPF2 binding, and domain 1B changes to a conformation where it does not clamp on the RNA 3' end, and as a result the RNA-unwinding activity is increased. This CH-domain in UPF1 is followed by superfamily 1 (SF1) helicase region that includes two

RecA-like domains (RecA1 and RecA2), which comprise the motor of the helicase [26]. Biochemical and structural analysis reveal that the complex of UPF2 or UPF3 binds to the CH-domain and causes a large conformational change, activating the ATPase or helicase activity of the UPF1 protein [27]. It is well studied that single amino acid mutations can affect or alter the kinetics of protein folding or cause protein aggregation and destabilization. Therefore, this project was focused on understanding the mutational landscape of the UPF1 protein; by exploring the structural **impact of cancer-derived mutations on UPF1 regions including the CH-domain (important for UPF1-UPF2 binding) and the ATP-binding site.** Previous studies have shown that simultaneous binding of UPF1 with ATP and RNA molecules results in nucleotide-induced conformational changes in the ATP-binding and hydrolysis cycle, and the CH-domain regulates the extent of RNA binding by the helicase [27]. **I examined the mutational landscape over the UPF1 protein structure retrieved form the cBioPortal database, that reveals a broad spectrum of mutations occurring in endometrial and ovarian cancers**. These point mutations were studied by integrated computational methods that can predict the effects of variations over the protein structure. Computing the binding energy using MM-PBSA (the molecular mechanics Poisson–Boltzmann surface area) approach, excluding the entropy contribution on the data retrieved from MD simulations of 1000 ns simulations (Fig. 3), I traced **conformational dynamics of the UPF1-AMPPNP (tri-phosphate moiety) and UPF1-UPF2 complexes upon inserting mutation** within these interfaces.
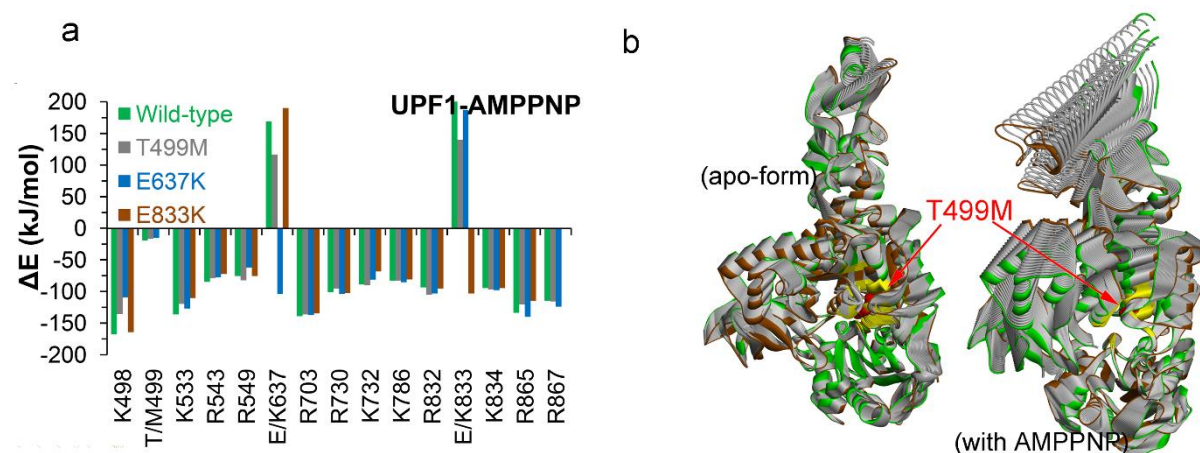


**Fig. 3.** Structural analysis of T499M, E637K, and E833K mutants within the UPF1 protein structure. **(a)** Individual residues of UPF1 contributing to the binding energy for AMPPNP, traced by MM-PBSA (the molecular mechanics Poisson–Boltzmann surface area). **(b)** The principal motion projected along the first eigenvector defined based on the combined trajectories (green marks the beginning and brown is the end of MDS; yellow are residues within 5 Å of the mutated residues).

*Description of results:*

The cancer genomic sequencing data suggests that residues in the ATP-binding site and CH-domain (i.e., UPF2 interacting interface) of the UPF1 protein are frequently mutated. As an initial screen, mutations emerging in the functional regions of UPF1 were investigated against ATP or UPF2 protein. The individual UPF1 variants resulting in **significant structural changes in the binding interface were K164R and R253W (with UPF2); and T499M, E637K, and E833K (with AMPPNP; Fig. 3a)**. Energy contribution of each UPF1 residue to

the binding with UPF2 / AMPPNP and the total binding energy for UPF1-UPF2 / AMPPNP, were calculated using the MM-PBSA technique (Fig. 3a). I traced **most promising key residues of UPF1** that were in common (wild type as well as mutant systems) forming hydrogen bond interactions with UPF2 or AMPPNP, and they were S152, S203, S205, and Q230 or R703, respectively. In addition, residues G495, T496, G497, K498, T499, K533, and G831 formed interactions with AMPPNP in wild type and in at least one mutant system. Based on these findings, we suggest that **phosphorylation of the Serine residues at position 152, 203, and 205 in UPF1 may modulate/control UPF1-UPF2 interactions / binding**. The CH-domain K164R and R253W mutant systems present well-defined conformational clusters when complexed with UPF2 compared to the apo-state and have increased UPF1-UPF2 interactions compared to the wild-type system. Moreover, the E637K and E833K alleles in the UPF1 substrate binding site improved the affinity of AMPPNP for the UPF1 protein, whereas the T499M (Fig. 3b) mutants hindered the triphosphate moiety of AMPPNP forming optimal interactions with UPF1.

*The importance of the results:*

Altogether, analysis of the structural changes in different UPF1 systems suggests the flexible nature of the substrate binding region in UPF1 with a predictable impact by certain cancer-associated mutations. These *in silico* site-directed mutagenesis findings can be useful in understanding the molecular mechanisms of important NMD components (i.e., UPF1 and UPF2) as well as to guide future experimental studies that aim to develop chemical tools to impact on UPF1-dependent processes. High occupancy interacting residues of UPF1 (traced from MD simulation) involved in the interactions with ATP analogues or with other proteins can provide **new insights into the molecular mechanisms of the UPF1-UPF2 / AMPPNP complex, as well as assist in the drug discovery strategies.**

**(iv) Kalathiya, U.,** Padariya, M., Baginski, M. (2019). Structural, functional, and stability change predictions in human telomerase upon specific point mutations. Scientific Reports, 9(1), 8707.

*Aim of the study:*

Telomerase is active in the early stages of life, maintaining the telomere length, and it becomes inactive in most somatic cells during adulthood. Most cancers have adapted mechanisms to protect lengths of telomeres, and this protection is achieved by telomerase activation in approximately 90% of human cancers. This study provided valuable insights into the nature of potential structural changes because of mutations, especially at the functionally important regions or residues of the active site. **I studied different potential mutations of telomerase enzymes and their effects when binding to various ligands** (known as potential inhibitors; C_9i [28], C_9k [28], 16A [29], and NSC749234 [30]; Fig. 4a). To study potential mutations of residues in active-site of human telomerase, two different approaches were followed; (i) the Blast tool implemented in the UniProtKB / Swiss-Prot database was used to search natural mutants, and (ii) for a particular residue from the active sites against all possible point mutations were searched to determine mutational consequences on the structure of the telomerase enzyme. Mutations that have shown similar behavior with all four ligands as well as those obtained the highest and lowest energies compared to wild type telomerase were

studied by molecular docking to reveal potential differences in ligand-binding properties. **I performed calculations identifying interaction energies and binding of the C_9i, C_9k, 16A, and NSC749234 inhibitors** in the active site of human telomerase upon inserting different point mutations (Fig. 4b). Considering the MDS data, I generated the binding pocket / important residues from the telomerase with ligands, along with identifying conformational dynamics from beginning-to-end of the MD simulations.
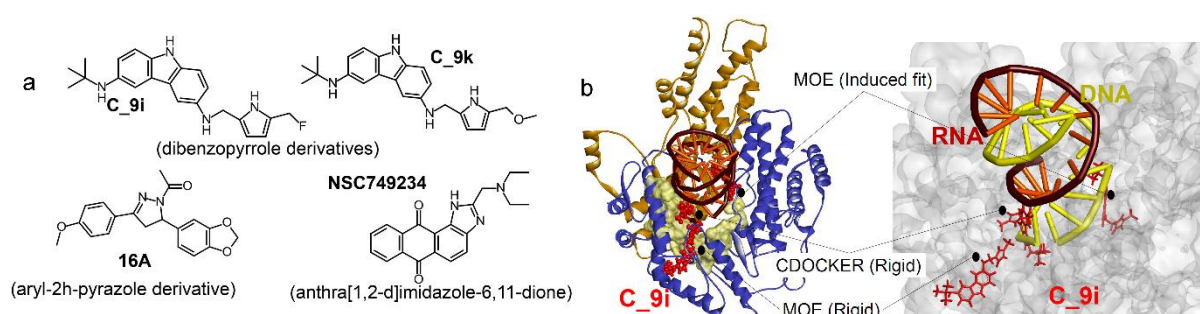


**Fig. 4.** The hTERT targeted by a set of few molecules. **(a)** The structure of ligand molecules, namely, C_9i, C_9k, 16A, and NSC749234, were studied with the human telomerase enzyme. **(b)** Binding modes of compound C_9i according to CDOCKER, MOE induced fit and rigid docking, and this C_9i ligand was found binding close to the DNA binding region within the telomerase enzyme.

### *Description of results:*

The human telomerase model structure was screened for the prediction of point mutations, to obtain a large-scale structural analysis of the mutations with different ligands (C_9i, C_9k, 16A, and NSC749234). Particularly, the C_9k inhibitor preferentially binds to wild type human telomerase compared to the other compounds. **Site-directed mutation (point mutations) analysis performed by me highlighted 16 residues** that may influence ligand affinity towards human telomerase (K626, R631, K710, D712, A716, Y717, D718, Q833, T839, V867, D868, D869, K902, P929, D945, and S947). Compared to the ligands 16A and NSC749234, the dibenzopyrrole derivatives (C_9i and C_9k) had improved their interactions with human telomerase in most of the studied mutant models. The residues **R631 and Y717 of wild type telomerase formed interactions with all four studied ligands and were also found frequently in most of the mutant models**. D868 formed stable H-bond interactions with the C_9i ligand in all simulated models, and the T567 and G834 residues interacting with C_9i were the same in the wild type and Y717H mutated models. Moreover, an interaction between Q569 and C_9i was the same in the Y717R and wild type telomerase models. The residues **Q833 and T839 interacting with the C_9k ligand were the same in both mutated models (Y717H and Y717R)**, and D868 formed interactions with the C_9k ligand in the wild type and Y717R systems.

### *The importance of the results:*

To the best of our knowledge, there have been limited theoretical studies on telomerase that analyze the effect of the mutations at the molecular level. Though hTERT is highly expressed in ~90% of cancer cells and is a potential target for tumour vaccines. Analyzing effect of the mutations over hTERT at the molecular level, provided **valuable insights into the nature of potential structural changes because of mutations, especially at the functionally**

**important regions or residues of the active site.** The detailed large-scale structural analysis of mutations or variants in human telomerase significantly enhances the predictive power of existing approaches that can be utilized in the **development of new / improved potent inhibitors against telomerase** by structure-based virtual screen approach.

**(v) Kalathiya, U.,** Padariya, M., Mayordomo, M., Lisowska, M., Nicholson, J., Singh, A., Baginski, M., Fahraeus, R., Carragher, N., Ball, K., Haas, J., Daniels, A., Hupp, T. R., Alfaro, J. A. (2020). Highly conserved homotrimer cavity formed by the SARS-CoV-2 spike glycoprotein: A novel binding site. Journal of Clinical Medicine, 9(5), 1473.

*Aim of the study:*

The SARS-CoV-2 spike protein is a homotrimer composed of three monomers that consists of an N-terminal ACE2 (angiotensin converting enzyme-2) binding domain (RBD), a central helix/heptad repeat, and a C-terminal region that interacts with the plasma membrane [31]. Homotrimer S protein assembly from monomeric forms can be rate limiting in cells, suggesting a possible space for intervention on the viral life cycle [32]. This project focuses on **understanding the trimer S glycoprotein and identifying the changes in the molecular properties due to conformational flexibility (Fig. 5a).** Findings from this project suggests that residues in the S2 subunit are less variable and more stable compared to the other regions, and the RBD domain obtained substantial flexibility which may be an obstacle in finding active hits. These details were used to identify potentially novel drug pockets or the active site regions specifically in the oligomeric SARS-CoV-2 spike glycoprotein. Specifically, **I targeted different pockets of the SARS-CoV-2 spike protein and found novel ligand molecules (from FDA approved chemical libraries; Fig. 5a and 5b) interacting with the homotrimer cavity**, and these molecules have gained high binding affinity compared to when docked other pockets of the S protein.
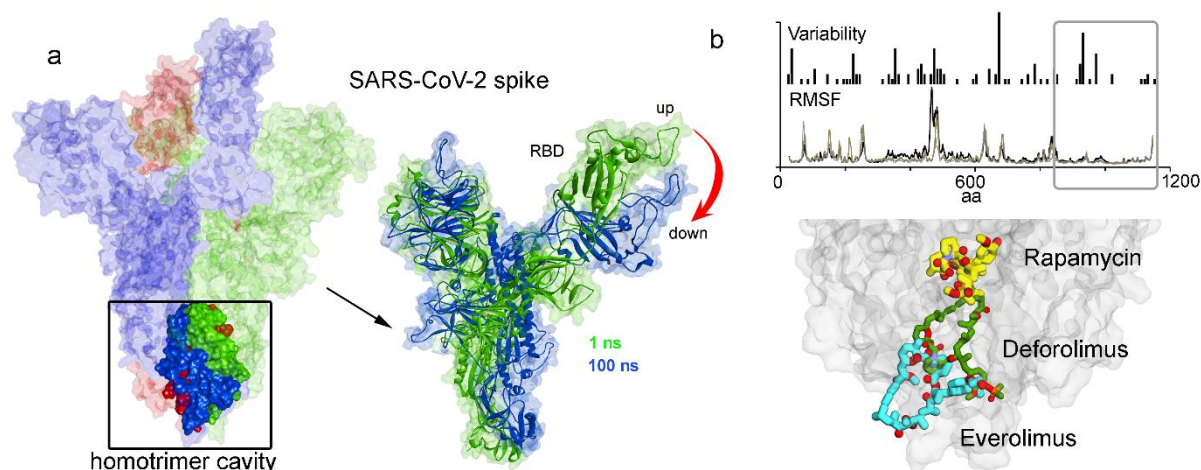


**Fig. 5.** Conformational dynamics of the SARS-CoV-2 spike glycoprotein. **(a)** Novel trimer cavity highlighted as shiny surface, and the right figures represents the "up" and "down" state conformation traced during the MD simulations of the monomeric form. **(b)** The root-mean-square fluctuation (RMSF) of the monomeric and trimeric forms, compared with the variability, and below figure shows a set of macrolide compounds traced during *in silico* screening.

*Description of results:*

Implementing the MD simulation on the SARS-CoV-2 S protein, **an "up" active state and a "down" inactive state conformation** of the spike protein in its monomer form were identified. The trimer cavity within this structural orientation suggests that it could work as **"bouncing spring" or "sarrus linkage" when the S protein interacts with the host cell receptor (ACE2; Fig. 5a).** These data suggest a physiological relevance due to heterogeneous protein conformational dynamics. For example, asymmetric conformational flexibility might have a functional role, perhaps in evading the exposure of B-cell epitopes (only one RBD domain is in the "up" conformation) and/or optimized interaction with the ACE2 receptor depending on virus strain. In addition, because of the "bouncing spring" mechanism (communication between the trimer pocket and the RBD conformation), it is possible that these different spike protein conformational isoforms provide another avenue to develop drug discovery programs that exploit and/or circumvent these dynamics. **I targeted the trimer pocket by structure-based virtual screen approach and identified a new functional class of drugs against S protein (Fig. 5b), with the hypothesis that such a ligand might perturb the predicted "bouncing spring" movement and the homotrimer formation**. Protein–ligand docking in the novel trimer cavity identified several hits that have already been published or proposed to inhibit the SARS-CoV-2 virus in cell systems. e.g., this project suggests **an action mechanism for molecules such as Chitosan and macrolide types** (e.g., Rapamycin). This mTOR inhibitor (Rapamycin) was found previously to disrupt LARP1 and mTORC1 binding and has been shown to reduce MERS infection by ~60% *in vitro* [14]. The postulated geroprotectors, such as Sirolimus (Rapamycin) and its close derivative, the Everolimus (RAD001), decreased infection rates in a small sample of elderly patients [15].

*The importance of the results*
The findings of this study can help to understand the **function of the highly conserved spike protein trimer cavity in the SARS-CoV-2 viral life cycle**, as well as provide a novel approach to target this class of infectious disease by the examination of spike protein trimer stability / assembly. A mechanism whereby Chitosan (and possibly its derivatives), as well as macrolide type molecules, might **bind to a pocket formed by the spike protein trimer and provide a novel domain to focus**. Several hits were identified when screening against the S trimer cavity and that have already been validated or suggested to inhibit the SARS-CoV-2 virus in published cell models. These findings identify a novel small molecule binding-site formed by the spike protein oligomer, that **can assist in future drug discovery programs aimed at targeting the coronavirus family of viruses**.

**(vi)** **Kalathiya, U.,** Padariya, M., Faktor, J., Coyaud, E., Alfaro, J. A., Fahraeus, R., Hupp, T. R., Goodlett, D. R. (2021). Interfaces with structure dynamics of the workhorses from cells revealed through cross-linking mass spectrometry (CLMS). Biomolecules, 11(3), 382.

Rapidly growing technologies to map out interactions between protein–protein / RNA / DNA is critically important since the biological function of these molecules is extremely influenced by their structures, complex formation, locations, and regulatory networks. This paper describes **a suite of methods that are valuable for detecting such interactions** between biomolecules of the cell. **I have given a special emphasis on the combination of cross-**

**linking with native mass-spectrometry technique (CLMS)** to yield mutual benefits while characterizing protein–protein / RNA / DNA interactions, though with a slightly modified approach. The CLMS strategies allow capturing and identifying not just stable, but also transient, dynamic, and weakly bonding molecules and thus, it emerged as the most striking example of multidisciplinary success among hybrid or integrative structural biology methods. In addition, CLMS provides inter-residue distances that can be integrated into the molecular modeling and MD simulation techniques, to achieve physiologically realistic quaternary structures. Different means to combine the knowledge of chemistry with MS to analyze essential biological systems of interest were scrutinized.

Several different cross-linkers are accessible with varied chemistries for the CLMS techniques, and therefore, one should consider the protein sequence of interest to determine what sort of combination of cross-linker and cleavage enzymes yield positive results. To start with one of the **most used cross-linker types, like amine reactive cross-linking** (NH2-NH2: BS3 and DSS, spacer length 11.4 Å; and BS2G, spacer length 7.7 Å) **and carboxyl-amine cross-linking** (COOH-NH2: EDC or sulfo-NHS zero spacer length), is advisable. **I highlighted three important (hydrogen / deuterium exchange, CLMS, and covalent labeling) mass spectrometry approaches** used in the structural proteomics analysis of binary complexes. I retrieved literature data about cross-linking reagents anatomy, chemistry, and evolution of cross-linkers. Together with dr inż. Monikaben Padariya (from University of Gdansk), **I was involved in preparing the schematic outline of the CLMS technique, to probe protein-DNA / RNA interactions**, and workflow for mass spectrometry analysis of UV-induced cross-links.

Though not previously carried out, in this review it has been speculated that **CLMS approach could be merged with novel interactomics techniques (proximity-dependent labeling by BioID or TurboID)**. Merging these two methods may bring spatial resolution to CLMS at a sub organelle level, because the BioID radius is estimated ~10 nm, and since CLMS mainly focuses on the proteins in the neighborhood of a given bait. These CLMS and proximal interactomics methods can be iteratively performed in living cells, i.e., the generated samples are the proximal species of a protein of interest labeled with biotin first and then cross-linked with interactors. Subsequently, perform streptavidin-biotin capture, followed by the digestion and cross-linked peptides identification of proximal interactors. Functional groups in a linker must be tested and characterized thoroughly, as well as demonstrated precisely within the CLMS framework, to avoid failure to detect cross-links or only a few cross-links.

Protocols that can be useful for **digestion in CLMS methodologies, such as 'in-gel', 'out-gel', 'in solution', and FASP** (filter aided sample preparation)**, with their successful usage were elaborated**. Alongside the CLMS application for identifying the protein–protein interactions, this paper covers or describes the **cross-linking methods for studying protein–RNA as well as protein–DNA interactions, which apparently arises as fertile ground for future CLMS utilization**. The typical CLMS workflow allows for the measurement of the proximity in three-dimensional space of amino acids, identifying proteins in direct contact with DNA or RNA, and it provides information on the folds of proteins as well as their topology in the complexes.

**(vii) Kalathiya, U.,** Padariya, M., Fahraeus, R., Chakraborti, S., Hupp, T. R. (2021). Multivalent display of SARS-CoV-2 spike (RBD domain) of COVID-19 to nanomaterial, protein ferritin nanocages. Biomolecules, 11(2), 297.

*Aim of the study:*

Ferritin-based nanocages have emerged as an attractive platform for vaccine generation [33]. In addition, ferritin is capable of a multivalent display of antigen molecules, and a multivalent presentation of antigens generally elicits a relatively potent immune response [34]. The influenza and SARS-CoV-2 viruses are related in terms of their structure and infectivity (i.e., both affecting the upper respiratory system). As the ferritin-based vaccine platform already shows a promising immune response against influenza [35], this project **proposes that the ferritin-based nanoparticle platforms would be equally effective in S protein RBD display**. In fact, recent studies verified that when the RBD domain (from different proteins) and ferritin were stitched together using variable linkers (genetic fusion), the synthetic chimeras displaying the RBD domain were effective in eliciting moderate to high immune responses in experimental animal models [36]. However, **no high-resolution structures of these synthetic chimeras are available, and there lacks a detail on the dynamics of these highly versatile systems**. Focusing on the SARS-CoV-2 virus, an open question exists as to how novel chimeric RBD nanocages would interact with the ACE2 receptor, and how different linkers fusing the RBD domain with the cage would alter such protein–protein interactions. To bring insight into these nanocage systems, **I designed (replicated; Fig. 6a) several spike RBD–ferritin synthetic proteins (utilizing two different ferritin systems H-ferritin and L-ferritin) with variable linkers** (a No-Linker, a small(S)-Linker with 5 amino acids (GGGGS), and a large(L)-Linker with 13 amino acids (GGGSGGGGSGGGS)) and performed extensive MD simulation (Fig. 6b). To confirm the characterization and validation for our synthetic SARS-CoV-2 spike RBD-ferritin chimeric nanocages**, I designed the M2e-H_ferritin** (influenza A virus matrix protein 2 ectodomain; M2e) construct as the template structure that was already experimentally validated [35].
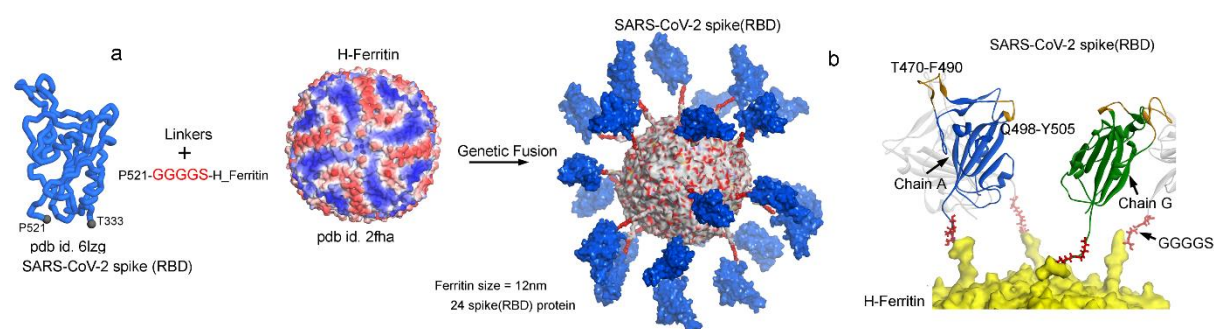


**Fig. 6.** The construction of a SARS-CoV-2 spike receptor-binding domain (RBD) protein over the surface of a ferritin nanocage. **(a)** Spike RBD–ferritin nanocage systems (5 aa S-Linker chimeric construct). The spike RBD domain (24 monomers) is presented over the H-ferritin. **(b)** The secondary structures of chain A and chain G in the spike RBD monomers suggests that both monomers interact with each other over the ferritin nanocage by the end of the MD simulation.

*Description of results:*

Nanomaterials, especially protein nanocages (e.g., ferritin), can be extremely useful for vaccine development against antigens such as the spike protein target derived from the SARS-CoV-2 coronavirus. Though smaller in size (H-ferritin cage size has a 12 nm outer diameter) when compared to the COVID-19 virion (a single virion size is ~60–140nm), these ferritin nanocages can mimic the SARS-CoV-2 coronavirus very effectively if a proper surface protein is incorporated in their system. In this project, **I constructed replicas of several spike (RBD)-H/L_ferritin constructs with variable linkers** (No-Linker, 5 aa S-Linker, and 13 aa L-Linker) and performed extensive molecular dynamics simulations. The secondary structure stability for each monomer of spike RBD protein over the H-ferritin nanocage suggests that the **No-Linker and 13aa L-Linker systems were more stable (with the minimum and maximum RMSD** (root-mean-square deviation of atomic positions) **difference at ~1 Å) and had less flexibility compared to the 5 aa S-Linker system** (~2 Å RMSD difference). In addition, the conformational dynamics of each simulated system suggest that **in the 13 aa L-Linker system, the spike RBD monomers were interacting specifically with each other, and most of them lacked the "up" active conformations**. In contrast to 13 aa L-Linker, in the 5 aa S-Linker system, a majority of the spike RBDs maintained an optimal distance to each other, resulting in more available free spike RBD for the ACE2 receptor interactions. Additionally, in the case of **the 5 aa S-Linker system, a greater number of the spike RBD domains maintained "up" active conformational state (receptor accessible; Fig. 6b)**, suggesting an optimal intermediate length of the linker. Like the H-ferritin, the L-ferritin nanocage showed a well-defined presentation of the spike RBD with only a 5 aa S-Linker. Experimentally validated M2e-H_ferritin [35] and investigated using our approach suggests that the M2e proteins maintained sufficient / optimal distance between each other over the H-ferritin nanocage, similar as in 5 aa S-Linker RBD system. Overall, our findings suggest that a **maximal level of "up" active conformations can be obtained from a proper combination of linker rigidity (flexibility) and steric support**, due to the inter-spike RBD interactions.

*The importance of the results:*

Due to its important role in virus transmission, the spike RBD is one of the key molecular targets for vaccine development. The findings from this study suggest that **synthetic SARS-CoV-2 spike RBD–ferritin nanocages are highly dynamic in nature and determine the optimum length of the linker** that is necessary for holding the antigen molecule safely. In addition, the behavior of the spike RBD regions (470-TEIYQAGSTPCNGVEGFNCYF-490 and 498-QPTNGVGY-505) responsible for interacting with the host ACE2 receptor was explicitly evaluated. These data propose an **effective multivalent synthetic nanocage, which might form the basis for new vaccine therapeutics designed against viruses such as SARS-CoV-2**. Based on our comprehensive computational study showing the dynamics of a spike RBD–ferritin nanocage, further experimental validation would determine whether such a chimera could impact future vaccine development against SARS-CoV-2 and related viruses. To the best of our knowledge, this is the **first comprehensive computational study showing the dynamics of ferritin–RBD constructs at molecular level**.

**Summary of the most important achievements:**

*1. Development of strategies to compare molecular properties of GC- and AU-rich mRNA motifs with the UPF1 protein, which suggest a GC-rich mRNAs have better binding affinity.*

The 'regulatory loop' in UPF1 has a control over the catalytic and/or ATPase activity of the protein. I investigated both isoforms of UPF1 in the presence or absence of the poly(U) mRNA, and GC- / AU-rich mRNA motifs by applying the MD simulation technique. I performed *in silico* docking screen of different GC- and AU-rich mRNA motifs with UPF1 protein, tracing changes in the interacting residues. 11 aa insertion (353-GNEDLVIIWLR-363) in the 'regulatory loop' is responsible for 'closed (isoform_1)' conformations between 1B and RecA2 domains, that could have impacted on the catalytic activity of the UPF1 helicase. This effect is directly linked to the absence of a stacking formation in isoform_2 results in an 'open conformation', as well as correlating with $UPF1_{P533T}$ cancer-derived mutation. Residues R363, D364, T627, and G862 defining active site pocket for GC-rich motifs, and R363, N535, and T627 amino acids for the AU-rich motifs were identified. Such novel perspectives from mRNA–protein binding pairs shall guide future experiments, as well as contribute to drug-development strategies for NMD pathway to produce neoantigens in cancer cells.

*2. Exploring the effects of cancer derived mutations within the transporters, the $TAP1_{R372Q}$ blocks the IF-to-OF conformation of peptide transport or kinetics towards the ER from cytosol.*

The variants located at the peptide binding pocket of the TAP1-TAP2 complex have diverse conformations; the $TAP1_{R372Q}$ variant was found highly stable compared to that of the $TAP2_{R373H}$. These mutations in TAP1 were found blocking both ER-lumen and cytosolic passage of the TAP transporter, which can hinder the intake and evacuation of the peptides from cytosol-to-ER (blocks the IF-to-OF conformation). Whereas mutations in the TAP2 protein were found mostly blocking only the ER-lumen pore of the TAP transporter. These cancer-derived mutations in TAP can control the peptide transport that could have been involved in the MHC-I antigen presentation.

*3. Optimizing the proteomics pipelines to trace proteins / peptides upon IFN-γ treated melanoma cells, and in silico screen these peptides against the TAP transporters.*

I performed proteomics analysis (implementing different bioinformatics approaches) for the INF-γ treated immunopeptidome datasets, and optimized pipelines to trace peptides from different samples / conditions. Identified peptides have a higher possibility of being presented over the cell-surface MHC-I molecules, as well as transported through the TAP transporter. Therefore, these peptides were *in silico* screened back with the TAP1-TAP2 heterodimer, tracing protein-peptide binding affinities and kinetics of peptide transport. Based on the peptide binding affinities with TAP transporters and the higher number of 9-10 mers (peptide length) were identified in the immunopeptidome datasets (Fig. 7). From these data it could be proposed that peptides having weak binding affinity with the TAP1-TAP2 transporters, could exhibit faster transport from the cytosol to the ER.
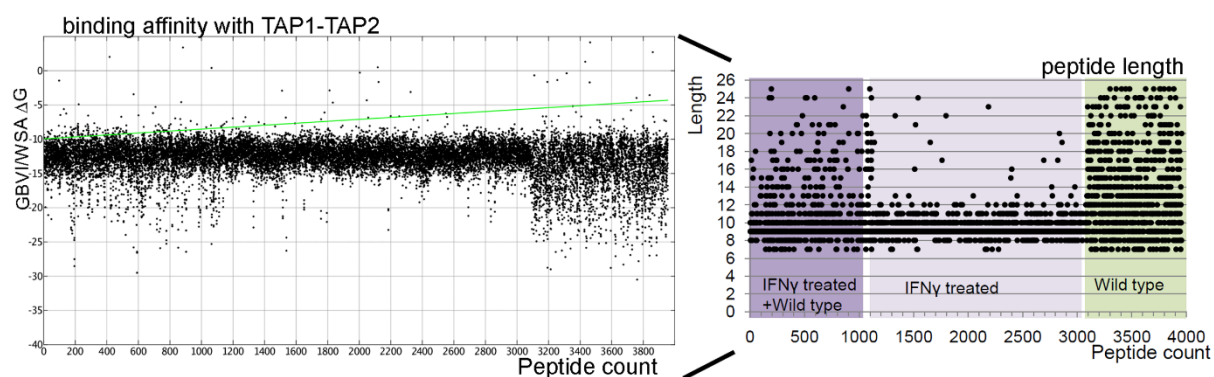
**Fig. 7.** Identical 3957 peptides from three samples (control, IFN-γ treated, and non-treated) that were screened against the TAP1-TAP2 transporters. Right panel represents the peptide length (differentiated according to the samples), and the left panel represents peptide binding affinity with the TAP transporters.

*4. Determination of frequency / probability of amino acids at residue mutation sites (mutation expression) over the hTERT enzyme, along with optimizing the in silico mutational landscape pipeline.*

I studied different potential mutations of telomerase enzymes, and their effects when binding to potential inhibitors. Important active site residues were subjected to trace all possible point mutations, to determine mutational consequences on the telomerase. This enzyme is found active in approximately 90% of human cancers, and hTERT derived antigenic peptides were previously reported, hence investigating the mutational landscape is crucial. The goal of this method is to modulate physical protein properties such as stability or affinity. Analyzing the effect of the mutations over hTERT at the molecular level, I provided valuable insights into the nature of potential structural changes because of mutations, especially at the functionally important regions or residues of the active site. These data can be utilized in the development of new / improved potent inhibitors against telomerase.

*5. Designing the multivalent display of SARS-CoV-2 spike RBD domain structures over different ferritin nanocages.*

Ferritin-based nanocages have emerged as an attractive platform for vaccine generation. Implementing the molecular modeling approach, I designed such nanomaterials to present the S protein RBD domain that is responsible to interact with the host cell receptor. Since there are no high-resolution structures of these synthetic chimeras available, such molecular dynamics investigation provided molecular level details on the dynamics of these highly versatile systems. I *in silico* designed (Fig. 6a and 8) several spike RBD–ferritin synthetic proteins (utilizing two different ferritin systems H-ferritin and L-ferritin) with variable linkers. In the 5 aa S-Linker system, a greater number of spike RBD domains maintained "up" active conformational state (receptor accessible). These data propose an effective multivalent synthetic nanocage, which might form the basis for new vaccine therapeutics designed against viruses such as SARS-CoV-2.
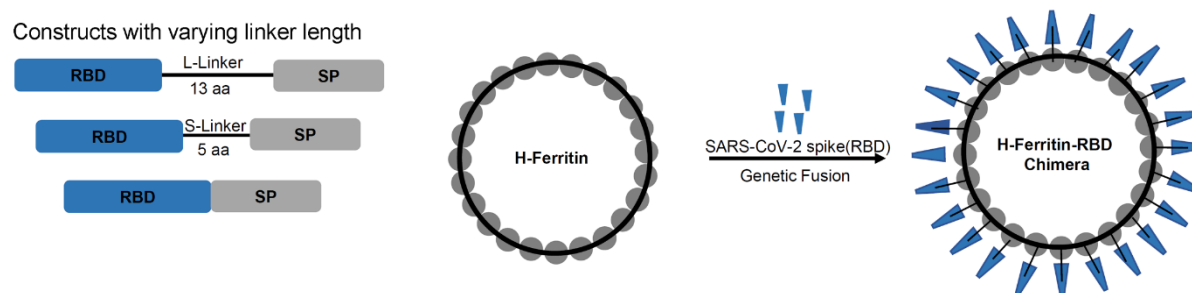
**Fig. 8.** A cartoon representation of different chimeric constructs with varying lengths of linkers (No-Linker; 5 aa S-Linker, GGGGS; and 13 aa L-Linker, GGGSGGGGSGGGS), displaying the spike RBD monomers over two different ferritin nanocages (H-ferritin).

*6. Identified a novel SARS-CoV-2 spike homotrimer cavity and targeted this pocket by medicinal compounds blocking conformational flexibility of the protein.*

The SARS-CoV-2 S protein is a homotrimer composed of three monomers, I performed molecular dynamics calculations understanding the trimer and identifying the changes in the molecular properties due to conformational flexibility. An "up" active state and a "down" inactive state conformation state of the spike protein in its monomer form were identified, as well as a "bouncing spring" or "sarrus linkage" conformation was analyzed for the S protein when it interacts with the host cell receptor (Fig. 5a). I targeted the trimer pocket using the SBVS approach and identified a new functional class of drugs against S protein, with the hypothesis that such medicinal compounds might perturb the predicted "bouncing spring" movement and the homotrimer formation. Moreover, an action mechanism for molecules such as Chitosan and macrolide types (e.g., Rapamycin) identified showing high affinity binding with S trimer cavity was developed (Fig. 5b). These structural properties and understanding of the S protein might assist in future drug discovery programs aimed at targeting the coronavirus family of viruses.

**Scientific plans:**

In 2020, I received funds from the National Science Center under the SONATINA program for the project entitled; Specificity in detection of PTCs in mRNA by NMD and its network, insights from cancer perspective and cross-linking (XL-MS). The main object of this project is to understand the mechanism with dynamic and structural properties of proteins from the NMD, EJC, and eRFs complexes that are involved in the recognition or binding to mRNA or the PTCs. The binding network between the protein-RNA / protein or RNA-RNA components using different computational methods (molecular modeling and molecular dynamics), as well as by applying the mass spectrometry (MS)-based structural techniques (cross-linking; XL-MS). During realization of this project, I have traced the GC- and AU-rich mRNA binding pattern with the UPF1 gene (publication no. 4.1). Moreover, blocking the UPF1 protein by novel medicinal compounds would produce a neoantigen that could have been transported to MHC molecules through TAP transporter (publication no. 4.2). A broad spectrum of cross-linking (XL-MS / CLMS) technique and protocol design were demonstrated (publication no. 4.6), and I will be using a particular DSS cross-linker to identify novel protein-protein networks from different cell lines using CLMS technique. In addition, as future tasks, I will be investigating

the conformational dynamics and structural properties of different proteins interacting with the UPF1 protein / mRNAs, as well as components form the NMD, EJC, and eRFs complexes.

Data produced for the publications 4.1 and 4.3, provided a support for the funds received from National Science Center (in 2021) under the OPUS program for the project entitled: "The impact of UPF1 ATP mimetics on the mutant immunopeptidome". I am a co-author of the grant proposal and took active part in designing the project with Prof. Theodore Hupp (from International Centre for Cancer Vaccine Science, University of Gdansk) and prof. Sławomir Makowiec (from the Gdańsk University of Technology). The project is focused on the core functions of UPF1 to find new small molecule ATP mimetics that can be used as chemical tools to define the function and druggability of the NMD pathway. Such novel compounds targeting UPF1 functions can increase stop-codon readthrough to enable the production of mutated full-length proteins / peptides, as well as these mutant peptides derived from stop codon readthrough can induce burden over the MHC class I. Inhibiting hUPF1 with a small molecule, generating translation readthrough thus creating neoantigen by virtue of the amino acid, which is added at a stop codon, in a sense stimulates the tumour cell to make its own neoantigen vaccine. I am performing a scientific leader task in this funded project with Prof. Ted Hupp, by coordinating the *in silico* data from structure based virtual screening, synthesized compounds from our chemist collaborators, and validating compound using optimized assay of UPF1. In addition, I am planning to continue my research direction relating to peptide loading complex and identifying the mutant peptides (neoantigens produced because of NMD inhibition) transported through TAP transporters as well as presented over the MHC molecules.

Novel medicinal compounds against the S protein trimer cavity (publication no. 4.5) from SARS-CoV-2 were identified, by applying my optimized *in silico* pipeline (screening of compound libraries). In the 2021 year, Prof. Neil Carragher (from University of Edinburgh) received funding from Medical Research Scotland for the project entitled: Examination of drug leads for binding to a novel homotrimer cavity formed by the SARS-CoV-2 spike glycoprotein, in which I took part as a co-investigator from Poland providing a set of *in silico* screened macrolide compounds. Furthermore, to continue my structural biology research relating to the S glycoprotein, in 2022 year, I received fund from UGrants-START program for the project entitled: Targeted inhibition of SARS-CoV-2 Spike Glycoprotein by novel medicinal compounds, reduced the replication of virus infected cells. This UGrants-START program from University of Gdansk supports research projects and publish them in high-rank JCR listed journals.

Furthermore, as a continuation of my interest in development of ferritin based novel vaccine, in 2021, with Prof. Simon Graham (from The Pirbright Institute, UK) and I as a leader from University of Gdansk applied for a grant application under the icard; international coordination of research on infectious animal diseases (application under consideration) competition. The project is entitled; PANCOVAC (application no. 65), Pan-coronavirus vaccination for broad cross-protection against present and future emerging coronaviruses in pigs. I will be designing recombinant ferritin nanocages expressing an array of spike protein receptor binding (RBD) and/or heptad-repeat (HR) domains from multiple porcine CoV genera.

5.    **Presentation of significant scientific or artistic activity carried out at more than one university, scientific or cultural institution, especially at foreign institutions**

**Description of scientific work:**

I began my Master of Science and Engineering (mgr inż.) studies in Biotechnology / Bioinformatics field in 2011 at Faculty of Chemistry, Wrocław University of Science and Technology, Wrocław, Poland. My master's thesis project was implemented in the group of prof. dr hab. inż. Wacław Andrzej Sokalski and dr. inż. Michał Jewginski entitled: Design of the Low Molecular Inhibitors of Cathepsin C. Findings this work resulted in my first scientific paper in the international recognized journal:

➢ **Kalathiya, U.,** Padariya, M., Jewginski, M., Baginski, M. (2014). Molecular docking studies towards development of novel Gly-Phe analogs for potential inhibition of Cathepsin C (dipeptidyl peptidase I). International Journal for Computational Biology, 3(1), 3-26.

After completing my master's degree from Wroclaw, I started my Doctoral Studies at the Department of Pharmaceutical Technology and Biochemistry, Faculty of Chemistry, Gdańsk University of Technology in the group of prof. dr. hab. inż. Maciej Bagiński. My doctoral dissertation was entitled: Molecular properties of TRF1 and TRF2 proteins dimer forming TRFH domains and their interactions with TIN2 or Apollo peptides. Applying the molecular dynamics simulation techniques over the shelterin proteins, I investigated the TRF1 and TRF2 proteins with different TIN2 or Apollo peptides. Moreover, the structural properties of the TRFH domain of TRF1 and TRF2 dimers with the TIN2 peptide were traced over MD simulation time course. The results obtained from the conducted scientific research were published in two research publications in high ranked JCR listed journals:

➢ **Kalathiya, U.,** Padariya, M., Baginski, M. (2018). The structurally similar TRFH domain of TRF1 and TRF2 dimers shows distinct behaviour towards TIN2. Archives of Biochemistry and Biophysics, 642, 52–62. https://doi.org/10.1016/j.abb.2018.02.005; (**IF$_{2018}$ 3.559; MEiN$_{2018}$ = 30,** according to the new MEiN scores **MEiN$_{2022}$ = 100**).

➢ **Kalathiya, U.,** Padariya, M., Baginski, M. (2017). Molecular basis and quantitative assessment of TRF1 and TRF2 protein interactions with TIN2 and Apollo peptides. European Biophysics Journal with Biophysics letters: EBJ, 46(2), 171–187. https://doi.org/10.1007/s00249-016-1157-7; (**IF$_{2017}$ 1.935; MEiN$_{2017}$ = 20,** according to the new MEiN scores **MEiN$_{2022}$ = 70**).

My Ph.D. defense (dr inż.) was held in November 2018 and obtained my doctoral degree in chemical sciences in the field of Biotechnology at Faculty of Chemistry, Gdańsk University of Technology, Poland. The reviewers of my doctoral defense were: prof. dr hab. Cezary Czaplewski from Faculty of Chemistry, University of Gdansk and prof. dr hab. Ewa Sikora from Nencki Institute of Experimental Biology, Polish Academy of Sciences. During my doctoral studies, along with the main research direction of Ph.D. defense, I was involved in other collaboration projects. In majority of these collaborated projects, I implemented *in silico*

drug design (bioinformatics) and molecular dynamics approach to investigate important structural properties of proteins. Moreover, using my knowledge in the programming languages, I contributed to database development projects. Overall, during my doctoral studies, I took active participation in 13 different collaborative publications, and they were:

➢ **Kalathiya, U.,** Padariya, M., Baginski, M. (2018). Extracting functional groups of ALLINI to design derivatives of FDA-approved drugs: Inhibition of HIV-1 integrase. Biotechnology and Applied Biochemistry, 65(4), 594–607. https://doi.org/10.1002/bab.1646; (**IF$_{2018}$ 1.559; MEiN$_{2018}$ = 20,** according to the new MEiN scores **MEiN$_{2022}$ = 40**).

➢ Padariya, M., **Kalathiya, U.,** Baginski, M. (2018). Structural and dynamic insights on the EmrE protein with TPP + and related substrates through molecular dynamics simulations. Chemistry and Physics of Lipids, 212, 1–11. https://doi.org/10.1016/j.chemphyslip.2017.12.004; (**IF$_{2018}$ 2.536; MEiN$_{2018}$ = 25,** according to the new MEiN scores **MEiN$_{2022}$ = 100**).

➢ Padariya, M., **Kalathiya, U.** (2017). Comparative molecular dynamics study of dimeric and monomeric forms of HIV-1 protease in ligand bound and unbound state. General Physiology and Biophysics, 36(2), 141–154. https://doi.org/10.4149/gpb_2016028; (**IF$_{2017}$ 1.479; MEiN$_{2017}$ = 15,** according to the new MEiN scores **MEiN$_{2022}$ = 40**).

➢ Padariya, M., **Kalathiya, U.,** Baginski, M. (2017). Molecular basis and potential activity of HIV-1 reverse transcriptase toward trimethylamine-based compounds. Biotechnology and Applied Biochemistry, 64(6), 810–826. https://doi.org/10.1002/bab.1543; (**IF$_{2017}$ 1.44; MEiN$_{2017}$ = 20,** according to the new MEiN scores **MEiN$_{2022}$ = 40**).

➢ **Kalathiya, U.,** Padariya, M., Baginski, M. (2016). Identification of 1H-indene-(1,3,5,6)-tetrol derivatives as potent pancreatic lipase inhibitors using molecular docking and molecular dynamics approach: Identification of 1H-Indene-(1,3,5,6)-tetrol Derivatives. Biotechnology and Applied Biochemistry, 63(6), 765–778. https://doi.org/10.1002/bab.1432; (**IF$_{2016}$ 1.413; MEiN$_{2016}$ = 20,** according to the new MEiN scores **MEiN$_{2022}$ = 40**).

➢ Padariya, M., **Kalathiya, U.** (2016). Structure-based design and evaluation of novel N-phenyl-1H-indol-2-amine derivatives for fat mass and obesity-associated (FTO) protein inhibition. Computational Biology and Chemistry, 64, 414–425. https://doi.org/10.1016/j.compbiolchem.2016.09.008; (**IF$_{2016}$ 1.331; MEiN$_{2016}$ = 20,** according to the new MEiN scores **MEiN$_{2022}$ = 70**).

➢ Padariya, M., **Kalathiya, U.,** Baginski, M. (2015). Structural and dynamic changes adopted by EmrE, multidrug transporter protein - Studies by molecular dynamics simulation. Biochimica Et Biophysica Acta (BBA) - Biomembranes, 1848(10 Pt A), 2065–2074. https://doi.org/10.1016/j.bbamem.2015.05.014; (**IF$_{2015}$ 3.687; MEiN$_{2015}$ = 35,** according to the new MEiN scores **MEiN$_{2022}$ = 100**).

➢ **Kalathiya, U.,** Padariya, M., Baginski, M., Padariya, C. (2015). SiMiSnoRNA: Collection of siRNA, miRNA, and snoRNA database for RNA interference / SiMiSnoRNA: RNA Interferansı için siRNA, miRNA ve snoRNA veritabanında depolanan siRNA, miRNA, and snoRNA koleksiyonları. Turkish Journal of

Biochemistry, 40(6). https://doi.org/10.1515/tjb-2015-0044; (**IF$_{2015}$ 0.211; MEiN$_{2015}$ = 15,** according to the new MEiN scores **MEiN$_{2022}$ = 20**).

➢ Padariya, M., **Kalathiya, U.,** John, J. G. (2015). Easy access tool for small interfering RNA (siRNA) data. In proceedings of 8th National Level Science Symposium. February 22, 2015, Rajkot, India, 2, 129-133. (ISBN: 9788192952116, Christ Publications, accessed from www.ss.christcollegerajkot.edu.in).

➢ **Kalathiya, U.,** Padariya, M., Baginski, M. (2014). Molecular modeling and evaluation of novel dibenzopyrrole derivatives as telomerase inhibitors and potential drug for cancer therapy. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 11(6), 1196–1207. https://doi.org/10.1109/TCBB.2014.2326860; (**IF$_{2014}$ 1.438; MEiN$_{2014}$ = 25,** according to the new MEiN scores **MEiN$_{2022}$ = 70**).

➢ **Kalathiya, U.,** Padariya, M., (2014). Inhibiting activity of HIV-1: Protease, reverse transcriptase and integrase all together by novel compounds using computational approaches. International Journal of Bioscience, Biochemistry, Bioinformatics (IJBBB), 448–457. https://doi.org/10.17706/ijbbb.2014.4.6.448-457.

➢ Padariya, M., **Kalathiya, U.,** Baginski, M. (2014). Docking simulations, molecular properties and ADMET studies of novel chromane-6,7-diol analogues as potential inhibitors of mushroom tyrosinase. Gene Therapy and Molecular Biology, 16, 201-217; (**IF$_{2013}$ 0.429; MEiN$_{2014}$ = 15,** according to the new MEiN scores **MEiN$_{2022}$ = 5**).

➢ Padariya, M., **Kalathiya, U.,** Berlicki, L., Baginski, M. (2014). Computer-aided design of organophosphorus inhibitors of urease. International Journal for Computational Biology, 3(1), 31-38.

I did my scientific internship during the period; June 2015 – September 2015 (3 months), at the laboratory of Prof B. Jayaram from The Supercomputing Facility for Bioinformatics and Computational Biology (SCFBio), Indian Institute of Technology (IIT), New Delhi, India (www.scfbio-iitd.res.in). This internship covered the project entitled: High-throughput screening approach to find new molecule (scaffold) inhibitors of tribolium castaneum telomerase and to design potential drug for cancer therapy. Internship extended my research to find potential telomerase inhibitors, by identification of potential lead compounds from libraries and evaluate their therapeutic potential using different computational methods optimized at Prof. Jayaram group. Preliminary data generated during this internship were further evaluated by molecular dynamics and biological experiments.

Additionally, during my doctoral degree I was involved in two funded projects, in which I took part as a co-investigator, and they were (i) New compounds with anticancer activity that disrupt telomere functions. The National Centre for Research and Development, Poland (TARGETTELO; 2017-2018), and (ii) New inhibitors of the catalytic subunit of telomerase. OPUS, National Science Center, Poland (2015-2016). In these projects, I applied a structure-based virtual screening approach to find novel molecules against the TRF1 and TRF2 protein from the shelterin, as well as against the telomerase enzyme. Particularly, in the TARGETTELO project I applied different approaches to trace novel medicinal compounds against the TRF proteins, e.g., fragment-based design protocol to generate derivatives of active molecules. In addition, collecting data of TRF1/2-TIN2 molecular dynamics data, a set of

peptide-like molecules were designed that could mimic the protein-peptide binding complex. These optimized pipelines were further implemented in the current projects (after doctoral degree) that I am working on related to NMD pathway and designing novel vaccine development strategies.

**Established collaborations:**

After receiving a PhD degree in November 2018, in 2019 I joined as a Adiunkt / Postdoctoral researcher at International Centre for Cancer Vaccine Science (ICCVS), University of Gdańsk, Poland. From 2020, I am a principal Investigator of the grant funded by The National Science Centre (NCN), Poland. During this time, I gave my contribution in the scientific field to probe novel approach in the vaccine design and drug development strategies, and I established several collaborations with well-known scientist in the field. In particular, the international collaborators were:

➢ prof. Neil Carragher, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland EH4 2XR, UK – drug discovery and its validation through experimental methods.

➢ prof. Chandra S. Verma, Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, 07-01 Matrix, Singapore 138671, Singapore – structural biology and protein investigation.

➢ dr. Soumyananda Chakraborti, National Institute of Malaria Research, Dwarka, New Delhi 110077, India – ferritin nanoparticles and its implications.

➢ prof. Juergen Haas, from Department of Infectious Disease, Edinburgh, Scotland EH4 2XR, UK – drug discovery against viral infection, e.g., SARS-CoV-2.

➢ Alison Daniels, Department of Infectious Disease, Edinburgh, Scotland EH4 2XR, UK – validation of drugs in SARS-CoV-2 infected cells.

➢ prof. Kathryn Ball, from Institute of Structural and Molecular Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK – Interferon signaling and its understanding.

➢ prof. Borek Vojtesek, from RECAMO, Masaryk Memorial Cancer Institute, Zlutykopec 7, 65653 Brno, Czech Republic – proteomics and mass spectrometry approach to trace proteins.

➢ Ashita Singh, from Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland EH4 2XR, United Kingdom – molecular biology and protein understanding.

> ➢ Kamila Pawlicka, from Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland EH4 2XR, United Kingdom – molecular biology and understanding of the NMD biological processes.

> ➢ dr. Minofar Babak, from Faculty of Science, University of South Bohemia, Branišovská 1760, 37005 Česk é Budějovice, Czech Republic – ionic liquids and structural biology.

After completion of my doctoral degree, I was involved in different projects and collaborated with researchers from different academic universities. In addition, to my main 7 scientific achievements, I took active participation in 7 other publications (after receiving doctoral degree), and they were:

> ➢ Padariya, M., Sznarkowska, A., Kote, S., Gómez-Herranz, M., Mikac, S., Pilch, M., Alfaro, J., Fahraeus, R., Hupp, T., **Kalathiya, U.** (2021). Functional interfaces, biological pathways, and regulations of interferon-related DNA damage resistance signature (IRDS) genes. Biomolecules, 11(5), 622. https://doi.org/10.3390/biom11050622; (**IF$_{2020}$ 4.88; MEiN$_{2021}$ = 100**).

> ➢ Padariya, M., **Kalathiya, U.,** Mikac, S., Dziubek, K., Tovar Fernandez, M. C., Sroka, E., Fahraeus, R., Sznarkowska, A. (2021). Viruses, cancer and non-self recognition. Open Biology, 11(3), 200348. https://doi.org/10.1098/rsob.200348; (**IF$_{2020}$ 6.41; MEiN$_{2021}$ = 100**).

> ➢ Alfaro, J. A., Bohländer, P., Dai, M., Filius, M., Howard, C. J., van Kooten, X. F., Ohayon, S., Pomorski, A., Schmid, S., Aksimentiev, A., Anslyn, E. V., Bedran, G., Cao, C., Chinappi, M., Coyaud, E., Dekker, C., Dittmar, G., Drachman, N., Eelkema, R., Goodlett, D., Hentz, S., **Kalathiya, U.,** Kelleher, N. L., Kelly, R. T., Kelman, Z., Kim, S. H., Kuster, B., Rodriguez-Larrea, D., Lindsay, S., Maglia, G., Marcotte, E. M., Marino, J.P., Masselon, C., Mayer, M., Samaras, P., Sarthak, K., Sepiashvili, L., Stein, D., Wanunu, M., Wilhelm, M., Yin, P., Meller, A., Joo, C. (2021). The emerging landscape of single-molecule protein sequencing technologies. Nature Methods, 18(6), 604–617. https://doi.org/10.1038/s41592-021-01143-1; (**IF$_{2020}$ 28.55; MEiN$_{2021}$ = 200**).

> ➢ Bibi, N., Farid, A., Gul, S., Ali, J., Amin, F., **Kalathiya, U.**, Hupp, T. (2021). Drug repositioning against COVID-19: a first line treatment. Journal of Biomolecular Structure & Dynamics, 1–15. https://doi.org/10.1080/07391102.2021.1977698; (**IF$_{2018}$ 3.31; MEiN$_{2021}$ = 70**).

> ➢ Uhrik, L., Hernychova, L., Muller, P., **Kalathiya, U.,** Lisowska, M. M., Kocikowski, M., Parys, M., Faktor, J., Nekulova, M., Nortcliffe, C., Zatloukalova, P., Ruetgen, B., Fahraeus, R., Ball, K. L., Argyle, D. J., Vojtesek, B., Hupp, T. R. (2021). Hydrogen deuterium exchange mass spectrometry identifies the dominant paratope in CD20 antigen binding to the NCD1.2 monoclonal antibody. The Biochemical Journal, 478(1), 99–120. https://doi.org/10.1042/BCJ20200674; (**IF$_{2020}$ 3.86; MEiN$_{2021}$ = 100**).

➢ Padariya, M., **Kalathiya, U.,** Houston, D. R., Alfaro, J. A. (2020). Recognition dynamics of cancer mutations on the ERp57-tapasin interface. Cancers, 12(3), 737. https://doi.org/10.3390/cancers12030737; (**IF$_{2020}$ 6.64; MEiN$_{2020}$ = 140**).

➢ Pawlicka, K., **Kalathiya, U.,** Alfaro, J. (2020). Nonsense-mediated mRNA decay: Pathologies and the potential for novel therapeutics. Cancers, 12(3), 765. https://doi.org/10.3390/cancers12030765; (**IF$_{2020}$ 6.64; MEiN$_{2020}$ = 140**).

**Summary of scientific achievement:**

In total my scientific research achievements include 30 articles published, from which 26 articles were published in the JCR listed journals, and 4 papers in international / national journals. Among which 7 articles constitute the achievements presented for evaluation, conceptualizing the structural and functional aspects of proteins / enzymes providing novel knowledge in the field of innovative vaccine design or drug development. In addition, I am the first corresponding or senior corresponding author in 18 articles of them. Summing impact factor (IF) of my all publication, the total IF is 116.432, and the citations of these papers is 195 (data retrieved from Scopus database and accessed on 19 April 2022). Moreover, the total score as per the list of Ministry of Education and Science (Ministerstwo Edukacji i Nauki; MEiN) is 2400 points. In this total 2400 points for my articles, it includes 9 publications for 100 points, 6 publications for 140, and 1 publication for 200 points.

I have participated in 6 different funded projects, among which 3 projects were funded by National Science Center, Poland, and 1 each project funded by The National Centre for Research and Development, Poland (TARGETTELO), UGrants-START, and Medical Research Scotland. I am the principal investigator for the project under SONATINA program (title: specificity in detection of PTCs in mRNA by NMD and its network, insights from cancer perspective and cross-linking) and UGrants-START program (title: targeted inhibition of SARS-CoV-2 Spike Glycoprotein by novel medicinal compounds, reduced the replication of virus infected cells). As well as a co-author of the project funded under OPUS program entitled: The impact of UPF1 ATP mimetics on the mutant immunopeptidome. I reviewed original works relating to my expertise for 10 articles published in the JCR listed journals. Moreover, I presented my research in 5 different conferences, among which 4 were poster presentations, and 1 oral presentation. I completed 3-month internship (title: high-throughput screening approach to find new molecule (scaffold) inhibitors of tribolium castaneum telomerase and to design potential drug for cancer therapy) in the laboratory of Prof B. Jayaram at The Supercomputing Facility for Bioinformatics and Computational Biology (SCFBio), Indian Institute of Technology (IIT).

➢ **Total number of publications: 30** (before doctoral degree: **16**, and after doctoral degree: **14**)
➢ **Scientific achievement: 7 papers** (first corresponding or senior corresponding author in 7 papers). **IF = 35.744,** MEiN = **860.**
➢ **The total number of publications excluding scientific achievements: 23** (first corresponding or senior corresponding author **11** papers). **IF = 80.688,** MEiN = **1540.**

- before doctoral degree: **16** papers. **IF = 20.398,** MEiN = **690.**
- after doctoral degree: **7** papers. **IF = 60.29,** MEiN = **850.**

➢ **The overall impact factor journals**, by year of publication: **116.432** (papers before doctoral degree: IF = **20.398,** papers after doctoral degree: IF = **96.034**).

➢ **The total number of MEiN points**: **2400** (papers before doctoral degree: MEiN = **690,** papers after doctoral degree: MEiN = **1710**).

➢ **Number of citations** based on databases accessed on 19 April 2022): **173** (Web of Science), **195** (Scopus), and **273** (Google Scholar).

➢ **Hirsch's index** according to Web of Science **H = 7**, Scopus **H = 7**, and Google Scholar **H = 8** (accessed on 19 April 2022).

➢ Number of **participations in funded projects: 6.**

➢ **Conference presentations:** 1 oral and 5 poster presentations.

## Supplementary Literatures:

1. M. Clerici, A. Mourão, I. Gutsche, N.H. Gehring, M.W. Hentze, A. Kulozik, J. Kadlec, M. Sattler, S. Cusack, Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2, *EMBO J.* **28** (2009) 2293–2306. https://doi.org/10.1038/emboj.2009.175.

2. L. Martin, A. Grigoryan, D. Wang, J. Wang, L. Breda, S. Rivella, T. Cardozo, L.B. Gardner, Identification and characterization of small molecules that inhibit nonsense-mediated RNA decay and suppress nonsense p53 mutations, *Cancer Res.* **74** (2014) 3104–3113. https://doi.org/10.1158/0008-5472.can-13-2235.

3. J. Lykke-Andersen, M.D. Shu, J.A. Steitz, Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon, *Cell.* **103** (2000) 1121–1131. https://doi.org/10.1016/s0092-8674(00)00214-2.

4. I. Kashima, A. Yamashita, N. Izumi, N. Kataoka, R. Morishita, S. Hoshino, M. Ohno, G. Dreyfuss, S. Ohno, Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay, *Genes Dev.* **20** (2006) 355–367. https://doi.org/10.1101/gad.1389006.

5. A.L. Karamyshev, Z.N. Karamysheva, Lost in translation: Ribosome-associated mRNA and protein quality controls, *Front. Genet.* **9** (2018) 431. https://doi.org/10.3389/fgene.2018.00431.

6. M.W. Popp, L.E. Maquat, Nonsense-mediated mRNA decay and cancer, *Curr. Opin. Genet. Dev.* **48** (2018) 44–50. https://doi.org/10.1016/j.gde.2017.10.007.

7. P. Praest, A.M. Liaci, F. Förster, E.J.H.J. Wiertz, New insights into the structure of the MHC class I peptide-loading complex and mechanisms of TAP inhibition by viral immune evasion proteins, *Mol. Immunol.* **113** (2019) 103–114. https://doi.org/10.1016/j.molimm.2018.03.020.

8. B. Sadasivan, P.J. Lehner, B. Ortmann, T. Spies, P. Cresswell, Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP, *Immunity.* **5** (1996) 103–114. https://doi.org/10.1016/s1074-7613(00)80487-2.

9. J.W. Shay, W.E. Wright, Role of telomeres and telomerase in cancer, *Semin. Cancer Biol.* **21** (2011) 349–353. https://doi.org/10.1016/j.semcancer.2011.10.001.

10. E.H. Blackburn, Telomeres: no end in sight, *Cell.* **77** (1994) 621–623. https://doi.org/10.1016/0092-8674(94)90046-9.

11. W.E. Wright, L. Hayflick, Nuclear control of cellular aging demonstrated by hybridization of anucleate and whole cultured normal human fibroblasts, *Exp. Cell Res.* **96** (1975) 113–121. https://doi.org/10.1016/s0014-4827(75)80043-7.

12. A.G. Bodnar, M. Ouellette, M. Frolkis, S.E. Holt, C.P. Chiu, G.B. Morin, C.B. Harley, J.W. Shay, S. Lichtsteiner, W.E. Wright, Extension of life-span by introduction of telomerase into normal human cells, *Science.* **279** (1998) 349–352. https://doi.org/10.1126/science.279.5349.349.

13. J.W. Shay, S. Bacchetti, A survey of telomerase activity in human cancer, *Eur. J. Cancer.* **33** (1997) 787–791. https://doi.org/10.1016/S0959-8049(97)00062-2.

14. J. Kindrachuk, B. Ork, B.J. Hart, S. Mazur, M.R. Holbrook, M.B. Frieman, D. Traynor, R.F. Johnson, J. Dyall, J.H. Kuhn, G.G. Olinger, L.E. Hensley, P.B. Jahrling, Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis, *Antimicrob. Agents Chemother.* **59** (2015) 1088–1099. https://doi.org/10.1128/AAC.03659-14.

15. A. Zhavoronkov, Geroprotective and senoremediative strategies to reduce the comorbidity, infection rates, severity, and lethality in gerophilic and gerolavic infections, *Aging (Albany NY).* **12** (2020) 6492–6510. https://doi.org/10.18632/aging.102988.

16. Y. Zhang, B.P. Orner, Self-assembly in the ferritin nano-cage protein superfamily, *Int. J. Mol. Sci.* **12** (2011) 5406–5421. https://doi.org/10.3390/ijms12085406.

17. Y. Jin, J. He, K. Fan, X. Yan, Ferritin variants: inspirations for rationally designing protein nanocarriers, *Nanoscale.* **11** (2019) 12449–12459. https://doi.org/10.1039/c9nr03823j.

18. F. He, A.H. Brown, A. Jacobson, Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway, *Mol. Cell. Biol.* **17** (1997) 1580–1594. https://doi.org/10.1128/mcb.17.3.1580.

19. C. Wu, B. Roy, F. He, A. Jacobson, Nonsense suppression position effect implicates poly(A)-binding protein in the regulation of translation termination, *SSRN Electron. J.* (2019). https://doi.org/10.2139/ssrn.3483671.

20. M. Gowravaram, F. Bonneau, J. Kanaan, V.D. Maciej, F. Fiorini, S. Raj, V. Croquette, H. Le Hir, S. Chakrabarti, A conserved structural element in the RNA helicase UPF1 regulates its catalytic activity in an isoform-specific manner, *Nucleic Acids Res.* **46** (2018) 2648–2659. https://doi.org/10.1093/nar/gky040.

21. J.A. Hurt, A.D. Robertson, C.B. Burge, Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay, *Genome Res.* **23** (2013) 1636–1650. https://doi.org/10.1101/gr.157354.113.

22. D. Zünd, A.R. Gruber, M. Zavolan, O. Mühlemann, Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3′ UTRs, *Nat. Struct. Mol. Biol.* **20** (2013) 936–943. https://doi.org/10.1038/nsmb.2635.

23. N. Imamachi, K.A. Salam, Y. Suzuki, N. Akimitsu, A GC-rich sequence feature in the 3' UTR directs UPF1-dependent mRNA decay in mammalian cells, *Genome Res.* **27** (2017) 407–418. https://doi.org/10.1101/gr.206060.116.

24. M.L. Oldham, R.K. Hite, A.M. Steffen, E. Damko, Z. Li, T. Walz, J. Chen, A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter, *Nature.* **529** (2016) 537–540. https://doi.org/10.1038/nature16506.

25. F. Seyffer, R. Tampé, ABC transporters in adaptive immunity, *Biochim. Biophys. Acta.* **1850** (2015) 449–460. https://doi.org/10.1016/j.bbagen.2014.05.022.

26. Z. Cheng, D. Muhlrad, M.K. Lim, R. Parker, H. Song, Structural and functional insights into the human Upf1 helicase core, *EMBO J.* **26** (2007) 253–264. https://doi.org/10.1038/sj.emboj.7601464.

27. S. Chakrabarti, U. Jayachandran, F. Bonneau, F. Fiorini, C. Basquin, S. Domcke, H. Le Hir, E. Conti, Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2, *Mol. Cell.* **41** (2011) 693–703. https://doi.org/10.1016/j.molcel.2011.02.010.

28. U. Kalathiya, M. Padariya, M. Baginski, Molecular modeling and evaluation of novel dibenzopyrrole derivatives as telomerase inhibitors and potential drug for cancer therapy, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 11 (2014) 1196–1207. https://doi.org/10.1109/TCBB.2014.2326860.

29. Y. Luo, S. Zhang, K.-M. Qiu, Z.-J. Liu, Y.-S. Yang, J. Fu, W.-Q. Zhong, H.-L. Zhu, Synthesis, biological evaluation, 3D-QSAR studies of novel aryl-2H-pyrazole derivatives as telomerase inhibitors, *Bioorg. Med. Chem. Lett.* **23** (2013) 1091–1095. https://doi.org/10.1016/j.bmcl.2012.12.010.

30. C.-L. Chen, D.-M. Chang, T.-C. Chen, C.-C. Lee, H.-H. Hsieh, F.-C. Huang, K.-F. Huang, J.-H. Guh, J.-J. Lin, H.-S. Huang, Structure-based design, synthesis and evaluation of novel anthra[1,2-d]imidazole-6,11-dione derivatives as telomerase inhibitors and potential for cancer polypharmacology, *Eur. J. Med. Chem.* **60** (2013) 29–41. https://doi.org/10.1016/j.ejmech.2012.11.032.

31. D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science.* **367** (2020) 1260–1263. https://doi.org/10.1126/science.abb2507.

32. B. Delmas, H. Laude, Assembly of coronavirus spike protein into trimers and its role in epitope expression, *J. Virol.* **64** (1990) 5367–5375. https://doi.org/10.1128/JVI.64.11.5367-5375.1990.

33. G. Jutz, P. van Rijn, B. Santos Miranda, A. Böker, Ferritin: a versatile building block for bionanotechnology, *Chem. Rev.* **115** (2015) 1653–1701. https://doi.org/10.1021/cr400011b.

34.  S. Chattopadhyay, J.-Y. Chen, H.-W. Chen, C.-M.J. Hu, Nanoparticle vaccines adopting virus-like features for enhanced immune potentiation, *Nanotheranostics.* **1** (2017) 244–260. https://doi.org/10.7150/ntno.19796.

35.  M. Qi, X.-E. Zhang, X. Sun, X. Zhang, Y. Yao, S. Liu, Z. Chen, W. Li, Z. Zhang, J. Chen, Z. Cui, Intranasal nanovaccine confers homo- and hetero-subtypic influenza protection, *Small.* **14** (2018) e1703207. https://doi.org/10.1002/smll.201703207.

36.  A.E. Powell, K. Zhang, M. Sanyal, S. Tang, P.A. Weidenbacher, S. Li, T.D. Pham, J.E. Pak, W. Chiu, P.S. Kim, A single immunization with spike-functionalized ferritin vaccines elicits neutralizing antibody responses against SARS-CoV-2 in mice, *ACS Cent. Sci.* **7** (2021) 183–199. https://doi.org/10.1021/acscentsci.0c01405.

6. **Presentation of teaching and organizational achievements as well as achievements in popularization of science or art**

   **Teaching achievements and contributions regarding dissemination of science:**

   ➢ Computer Science laboratories at Department of Physical Chemistry, Faculty of Chemistry, Gdansk University of Technology (Environmental Protection course in English) (2013-2016).

   ➢ Molecular Modeling laboratories at Faculty of Chemistry, Gdansk University of Technology (Bioinformatics course) (2014).

   **Scientific supervision of students:**

   ➢ As a co-advisor of master's students from School of Biological Sciences, The University of Edinburgh = 2 students.

7. **Apart from information set out in 1-6 above, the applicant may include other information about his/her professional career, which he/she deems important.**

   ➢ During my doctoral degree I was awarded funding by the Polish National Agency for Academic Exchange (NAWA), Poland under the program The Ignacy Łukasiewicz Scholarship, (2013-2016). In addition, I received the best Ph.D. student scholarship from Rector of Gdańsk University of Technology (2014-2015).

   ➢ Conferences or workshops attended:
      - Molecular biology and immunology of cancer - R&D perspectives: SCANBALT forum 2019, Gdansk, Poland (24-25 September 2019).
      - Veterinary Science, Computational Mass Spectrometry and Therapeutics, Gdansk, Poland (13 January 2020)

*A detailed list of all my scientific achievements can be found in the Appendix 4.*

…..…………..………..……………….

**(Applicant's signature)**