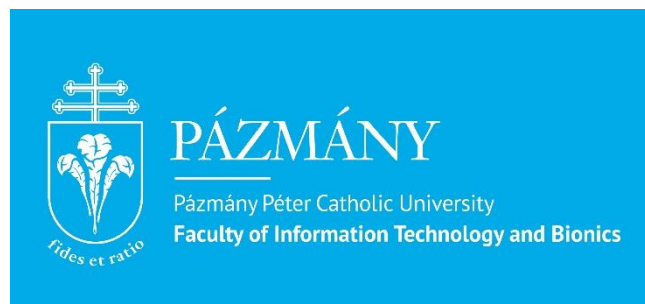




**Pázmány Péter Catholic University
Faculty of Information Technology
Students' Scientific Association
2021**

**NMR-based assignment and protein dynamics measurement of the
EVH1 domain of the postsynaptic Homer1**

Brigitta Alexandra Maruzs

**Supervisors:
Dr. Zoltán Gáspári
Dr. Bálint Ferenc Péterfia**

Bionics section

TABLE OF CONTENTS

1. Introduction	3
1.1 Scientific background	3
1.2 Previous results.....	3
1.3 Aims of the study.....	4
2. Literature survey.....	5
2.1 The postsynaptic density and its scaffold proteins	5
2.2 The Homer protein family	6
2.3 Splice variants and domain organisation of Homer1.....	7
2.4 The EVH1 domain family	8
2.5 Pathology of Homer1	10
2.6 Known interactions of Homer1	11
2.7 Nuclear magnetic resonance (NMR) spectroscopy	12
2.8 Resonance assignment.....	13
2.9 Protein dynamics	14
2.10 Structure calculation	15
3. Methods	16
3.1 Resonance assignment.....	16
3.2 Analysis of protein dynamics	18
3.3 Protein structure calculation	19
3.4 In vitro mutagenesis of the EVH1 domain	21
3.5 Protein production and expression	23
3.6 Protein isolation.....	25
3.7 Spectroscopic measurements for protein structure description	27
4. Results	28
4.1 Resonance assignment and structure calculation.....	28
4.2 Analysis of protein dynamics	29
4.3 In vitro mutagenesis of the EVH1 domain	31
4.4 Examination of known substitutions of the EVH1 domain	34
5. Summary.....	37
5.1 Discussion.....	37
5.2 Future plans	37
5.3 Acknowledgements	37
References	38

1. INTRODUCTION

1.1 Scientific background

The human brain is a widely investigated, complex structure, that is in the focus of several hundreds of research groups with interests ranging from molecular studies to stimulations of neural networks. Studies are trying to get us closer to understand brain functions, with which potential therapeutic methods can be determined. Postsynaptic density proteins are localised at excitatory synapses, and their alteration are involved in such disorders, like autism spectrum disorder (ASD) and schizophrenia. Our research group at Pázmány Péter Catholic University aims to describe PSD proteins, hereby get an insight to molecular processes of the brain and its disorders. Our focus is on the investigation of protein structure and dynamics based on NMR measurements, as well as the interactions between PSD proteins. Additionally, our current perspective is the characterisation of point mutants of a couple of PSD proteins, described in ASD. The structural frame of this protein network is provided by scaffolding proteins, particularly Homer and Shank proteins, from which I was focusing on Homer.

In this work I would like to give an insight to the organisation and importance of the postsynaptic density, especially on Homer proteins, their isoforms and the EVH1 domain family. I give the analysis of the domain's dynamics, structure, interactions and disorders associated to the domain. To this end, several bioinformatical tools were used, which I would like to introduce, then continue with the detailed description of protein dynamics and structure calculation, as well as the production and isolation of mutant proteins. Finally, I would like to analyse these results.

1.2 Previous results

In the literature we could only find X-ray structures revealed from the EVH1 domain, but no detailed NMR-based (dynamical and structural) investigation. However, two data sets are available on the assignment. One on the EVH1 domain of the human Homer3a variant by Victoria Higman and Peter Schmeider et al., from 2007, but only amide NH-shifts can be found in Biological Magnetic Resonance Data Bank (BMRB) and results were not published. This research had a preceding study by Linda J. Ball and Hartmut Oschkinat (Ball et al., 2002). The other assignment was performed from the first 189 amino acids of the Homer1 protein, that involves the EVH1 domain of my interest (Tanaka et al., 2000), but also a disordered region after the domain, that hardened the assignment and was declared by the authors to be removed. So far, this study was not further improved.

In laboratory, the cloning of the wild-type Homer1 EVH1 (Uniprot ID: Q9Z2Y3) construct along with the optimisation of protein production and expression were done by Csilla Homonnay and István Pap¹. For molecular cloning a Homer1a plasmid was used as template, that contains mouse cDNA

¹ Their Thesis connected to this research:
Csilla Homonnay – A szinaptikus jelátvitelben és plazmticitásban résztvevő Homer fehérje és komplexei kísérletes és bioinformatikai vizsgálata (2017)
István Pap – A Homer fehérjecsatlád bioinformatikai és kísérletes vizsgálata (2017)

and was kindly provided by Enora Mountin from the University of Geneva. The generated construct was named as V2-H118, after the expression vector to into which the insert was cloned (pEV2) and the 118 amino acids of the EVH1 domain. However, the encoded protein sequence is longer, due to the six-histidine long 6xHis tag, preceded by 2 amino acids and followed by a linker region of 17 amino acids, but with a TEV cleavage the majority (22 residues out of the 25) can be split off from the amino-terminal of the expressed protein.

I have been working in this laboratory for three years so far, therefore my previous studies serve as basis of this Thesis as well. During my bachelors, first I prepared a ^{15}N -labelled EVH1 domain for NMR. This study seemed to be promising, therefore I could continue with a double isotopic (^{13}C - ^{15}N) labelled sample, that served as basis of protein assignment. That sample exhibited good quality signal dispersion – based on NMR –, efficient labelling – based on mass spectrometry – and appropriate secondary structure – based on circular dichroism. However, it contained those previously mentioned 25 amino acids, including the 6xHis tag, that region can be confusing by structural-dynamical analysis and interaction measurements. Therefore, a new sample was needed, on which TEV cleavage had to be performed in order to go further to the molecular level analysis of the domain. TEV cleavage was the last step, that I could accomplish before my graduation, I also sent the samples for NMR, but the assignment could not be completed. Since I have started master course, my research focus was on the interaction measurement between Homer and Shank proteins, therefore the assignment and protein dynamics were postponed. This served as a start for the present study.

1.3 Aims of the study

My main task was to do the resonance assignment on the previously recorded two- and three-dimensional spectra, that served the basis of the determination of internal dynamics and three-dimensional structure, from which molecular function can be inferred. In order to fulfil these tasks, I had to get familiar with several computational tools.

For the examination of the effects of known substitutions of the EVH1 domain, the *in vitro* mutagenesis of the V2-H118 construct, protein expression from these plasmids and purification of the produced proteins were needed. When the protocols were set, I could prepare ^{15}N -labelled mutant and wild-type proteins and do several measurements with them, including NMR – from which a preliminary protein assignment could be done –, circular dichroism spectroscopy (CD), mass spectrometry (MS) and tryptophan fluorescence. After the analysis of the results, it seems to be beneficial in the future to examine interaction of wild-type and mutant EVH1 domains with Shank PRRs, as well as their internal dynamics.

2. LITERATURE SURVEY

2.1 The postsynaptic density and its scaffold proteins

Synapses are semi-membraneless, protein-dense compartments, that are highly unique and responsible for neural signal processing. Their proper formation and responsiveness are essential for the functioning of mammalian brains (Zeng et al., 2018). In chemical synapses the information flow is ensured by neurotransmitters. Neurotransmitters are stored in vesicles of presynaptic neurons, that are emptied via calcium ion signalling into the synaptic cleft, from which postsynaptic neurons can take up neurotransmitters and convey information further. (Szabad and Lippai, 2016). Based on the type of the neurotransmitter, chemical synapses can be either excitatory or inhibitory.

The postsynaptic density (PSD) is the electron dense thickening beneath the postsynaptic membrane of excitatory synapses, mostly situated at the tip of dendritic spines (Kaizuka and Takumi, 2018). This structure is densely packed with proteins, that form disc-shaped, 30-50 nm wide assemblies, so called nanodomains and undergo assembly/disassembly as a result of neuronal stimuli. By the course of synaptic

plasticity, the size, shape, structure and synaptic strength of PSD changes, leading to long term potentiation (LTP) and long-term depression (LTD) (Zeng et al., 2018). Based on proteomic studies, PSD contains several thousands of proteins, including transmembrane channels and receptors, adhesion molecules, scaffolding proteins, signalling enzymes and elements of the actin cytoskeleton (Kaizuka and Takumi, 2018). One special characteristic of this structure is that it is not continuously covered by membrane: from its one side, PSD is open to the cytoplasm, although connected to the postsynaptic membrane from the other. PSD is considered to be autonomously formed in living neurons via phase separation, when protein concentrations are above a certain threshold. This process enables molecules to remain in condensed state, promote synaptic development and maturation, regulate synaptic plasticity and also ensures colocalization, therefore regulation of interactions (Zeng et al., 2018). PSD formation has a high impact in the amount of those proteins that are needed to be produced in all neurons and synaptic processes, therefore it is not surprising, that mutations of numerous PSD proteins are associated with human neurological and psychiatric diseases, particularly autism spectrum disorder (ASD) and schizophrenia (Kaizuka and Takumi, 2018).

In mammalian organisms, PSD is most abundant in MAGUK (membrane-associated guanylate kinases) proteins, especially in PSD-95, that can be found still in the core layer (Dosemeci et al., 2016).

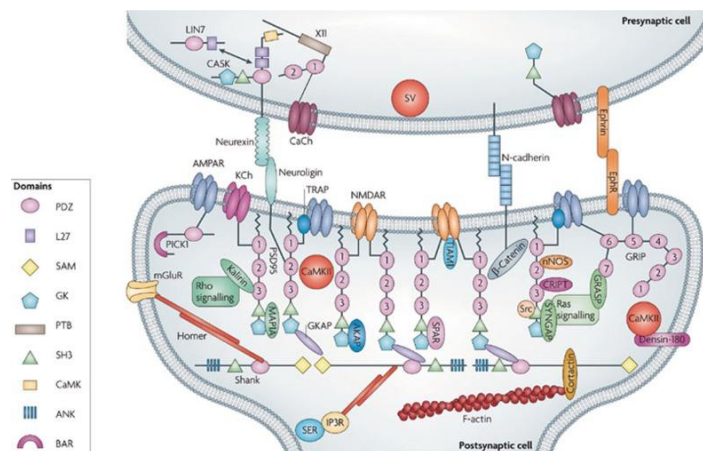


Figure 1. Dense protein network of the postsynaptic density (Feng et al., 2009, Figure 1)

This protein has key role in the organisation and stabilisation of PSD. It has three PDZ domains, followed by an SH3 (Src Homology 3) and a GK (guanylate kinase) domain (Blanpied et al., 2018).

PSD-95 establishes interaction with GKAP (guanylate-kinase associated proteins), that provides a stable platform for the pallium. This protein is highly disordered and present in half of the amount of PSD-95. Besides PSD-95, it establishes connection with Shank proteins as well (Dosemeci et al., 2016).

Shank and Homer proteins form an extended scaffold, that serves as structural framework of PSD. Shank proteins (SH3 and multiple ankyrin repeat domains protein) also named ProSAPs (Proline-rich synapse-associated protein) associate with each other through their SAM domains into sheet-like structures. This family has three members, Shank1, Shank2 and Shank3; and their mostly studied domains are SH3 and PDZ (Dosemeci et al., 2016). It establishes interaction with Homer, PSD-95 and GKAP proteins as well.

The Homer protein family contains three members: Homer1, Homer2 and Homer3 and owns EVH1 (Ena/VASP homology domain 1) domain on their amino-terminals and – apart from Homer1a isoform - coiled coil regions on their carboxyl-terminals. The latter structure ensures multimerization between Homer proteins. Homer proteins can interact with mGluRs and assemble them into signalling complexes (Tu et al., 1999), among several other partners.

2.2 The Homer protein family

Homer family members have upregulated expression following synaptic activity; therefore Homer genes are termed as immediate-early genes (IEG) based on their expression pattern. Three mammalian genes have been described, that encode for Homer1, 2 and 3 (Beneken et al., 2000). Homologs have been described only in animals (Shiraishi-Yamaguchi and Furuichi, 2007) with different tissue distributions (Hermey et al., 2017). In this study I was focusing on Homer1, because that protein is mostly expressed in brain regions, especially in the cortex and hippocampus (CA1 region). In the postsynaptic density they function as adaptor proteins, that can regulate synaptic localization or crosstalk between target proteins. Homer2 family members are mostly expressed in muscles. Highest expression of Homer3 family members is mainly observed in the spleen (Shiraishi-Yamaguchi and Furuichi, 2007).²

² <https://www.ncbi.nlm.nih.gov/gene/9456>, Date of download: 13. 09. 2021
<https://www.ncbi.nlm.nih.gov/gene/9455>, Date of download: 13. 09. 2021
<https://www.ncbi.nlm.nih.gov/gene/9454>, Date of download: 13. 09. 2021

2.3 Splice variants and domain organisation of Homer1

Alternative splicing can modulate the activity-dependent transcriptional signature, that cause changes in the molecular structure and composition of neurons. This fine-tuning synaptic strength is essential in the development of the nervous system, particularly in learning and memory (Denkena et al., 2020). All three Homer proteins have more splice variants. Homer1 gene comprises twelve exons and nine isoforms, of which Homer1a and Ania3 lack exon 2. These two splice variants are induced upon synaptic activity, whereas Homer1b and Homer1c are continuously expressed in the hippocampus.

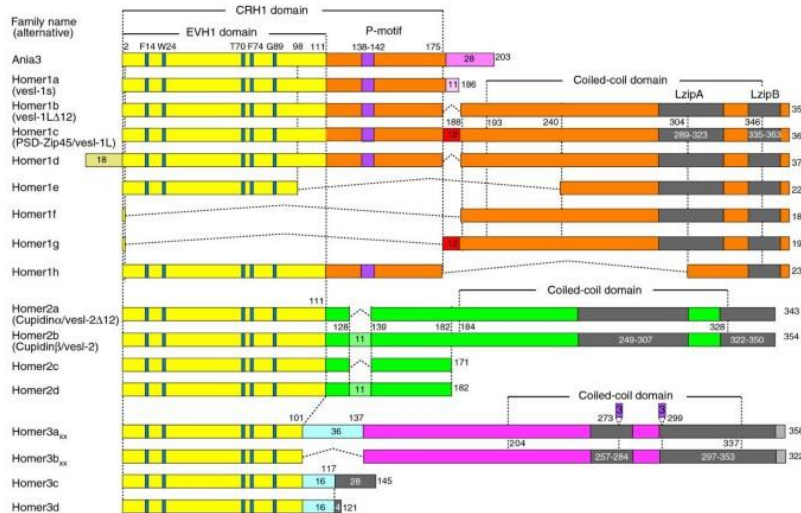


Figure 2. Splice variants and domain organisation of Homer family proteins (Shiraishi-Yamaguchi and Furuichi, 2007, Figure 1)

(Hermey et al., 2017). Long Homer variants are able to form tetramers through their coiled coils and act as molecular hubs in PSD, that are essential of the structural and functional integration of dendritic spines (Hayashi et al., 2009). Activity-regulated expression of the short Homer1a variant, that lacks the coiled coil, disrupts these multimers by competitively binding target proteins with its EVH1 domain. Consequently, the Homer1 truncating alternative splicing result in dominant negative functional block of the long, constantly expressed variants, which phenomenon is used in psychoactive stimulants (Hermey et al., 2017).

Homer proteins can regulate cell-surface targeting and clustering of mGluR1α/5, but the different Homer isoforms in opposed ways: long variants inhibit, while Homer1a increases this process. Also, their effect on spine maturation is different: the coexpression of Homer1b with Shank induces dendritic spine maturation and enlarge spine heads, leading to increased synaptic currents, whereas Homer1a induces a negative feedback loop on the structural regulation and function of synapses (Shiraishi-Yamaguchi and Furuichi, 2007).

All Homer1 transcripts share the same 5' untranslated region (UTR) and the first 450 amino acids of their open reading frame, that code for the EVH1 domain and proline motif. All long Homer1 variants except for Homer1a and Ania3 contain coiled coil segments. In those two, truncation of the coding sequence results in shorter proteins (short variants)

The amino-terminal of Homer proteins show homology to EVH1/VH1 domains of VASP (Ena/Vasodilator-stimulated phosphoprotein) protein family and WASP (Wiskott-Aldrich Syndrome protein) homologues. Homer EVH1 domains are approximately 110 amino acids long – that is only 11 residues shorter than Homer1a –, possess highly conserved structure and bind proline rich motifs of mGluRs, Shank and Drebrin proteins, IP₃Rs (inositol-triphosphate receptors) and RyRs (ryanodine receptors). The consensus proline-rich sequence of the Homer ligand is PPxxF, where x refers to an arbitrary amino acid, but substitution of any other position disrupts the interaction. This motif differs between the interaction partners (Beneken et al., 2000). Interestingly, not just the EVH1, but the first 175 amino acids of Homer1 are highly conserved and called the conserved region of Homer1 (CRH1). After the EVH1 domain, the CRH1 includes a proline-containing motif (P-motif) in mammalian Homer1 proteins. This is a five-residue long motif (138-S-P-L-T-P-142 *mus musculus*), that can establish intermolecular binding with an EVH1 domain at a site, that partly overlaps the binding site of proline-rich motifs. Consequently, the EVH1 domain can either bind the P-motif or the target protein's PPxxF motif, which arrest (former) or induce (latter) multimerization and PSD complex formation. Accordingly, the P-motif has a negative effect on PSD protein interactions, just like Homer1a. Both can remodel PSD signalling complexes in an activity-dependent, reversible way. The previously mentioned coiled coils of long Homer isoforms is placed on their carboxy-terminal and followed by two leucine zippers. This structure mediates homophilic or heterophilic interactions between different Homer family members with the aid of leucine zipper motifs, therefore, regulates signal transduction or crosstalk between target proteins. Synaptic localization of long variants is highly dynamic as a result of synaptic activity (Shiraishi-Yamaguchi and Furuichi, 2007)

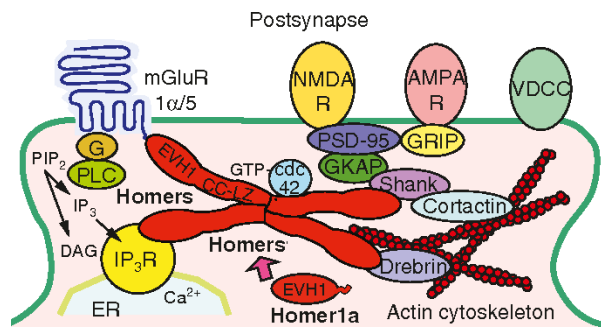


Figure 3. Disruption of the protein cluster by the dominant negative functional block of Homer1a (Shiraishi-Yamaguchi and Furuichi, 2007, Figure 4c)

2.4 The EVH1 domain family

The EVH1 domain (Ena/VASP homology domain 1) belongs to the superfamily of PH (Pleckstrin homology) domains that mostly appear in a large variety of signalling proteins.³ It is believed, that EVH1 domains have evolved from versatile PH/PTB folding scaffold and diverged in order to fulfil their functions in the cell and maintain their binding abilities (Ball et al., 2002). PH, PTB (phosphotyrosine-binding), RanBD (Ran-binding domains) and EVH1/WH1 domains together compose the superfamily of PH domains, and all possess a partly opened beta-barrel that is capped by an alpha

helix.³ Despite of structural similarity, their sequence homology is quite low and their binding sites and partners also differ (Perhoda et al., 1999).

EVH1 domains are mostly observed in multi-domain proteins. These proteins usually have function in signal transduction, cytoskeletal organisation or nuclear transport. Representatives of EVH1 domains are present in organisms from yeast to mammals.⁴ They consist of two antiparallel β -sheets, that form a β -sandwich that is capped by a C-terminal α -helix at one end. On the outside of each sheet, two concave surfaces are formed, one of them – together with an interstrand loop - is the base of the peptide-binding groove (Prehoda et al., 1999). EVH1 domains are believed to have highly conserved residues, that are essential for hydrophobic core packing, consequently for the appropriate and stable folding. Interestingly, a cluster of three, surface-exposed aromatic sidechains (Tyr16, Trp23 and Phe79, VASP numbering), are highly conserved amongst all EVH1 domains, except for the Homer-Vesl family, where Tyr16 is replaced by an aliphatic isoleucine. This cluster forms the recognition site for their ligands.

EVH1 domains can be classified based on the consensus sequences of the target PRS. Class I EVH1 domains have the consensus binding motif of FPx ϕ P (where x is arbitrary, ϕ is hydrophobic residue), therefore can bind focal adhesion proteins, like zyxin and vinculin amongst numerous mammalian proteins. Class II EVH1 domains have been described in members of the Homer-Vesl family, that recognise their PRS with a core PPxxF motif, found in mGluR1, IP₃Rs, RyRs and Shank proteins (Ball et al., 2002). Interestingly, the consensus binding sequence of Homer-Vesl EVH1 domains (PPxxF) is just the reverse of the binding sequence of the Ena/VASP (FPPPP). The distinctive mode of targeting reduces the possibility of cross-reaction with other PRSs (Beneken et al., 2000).

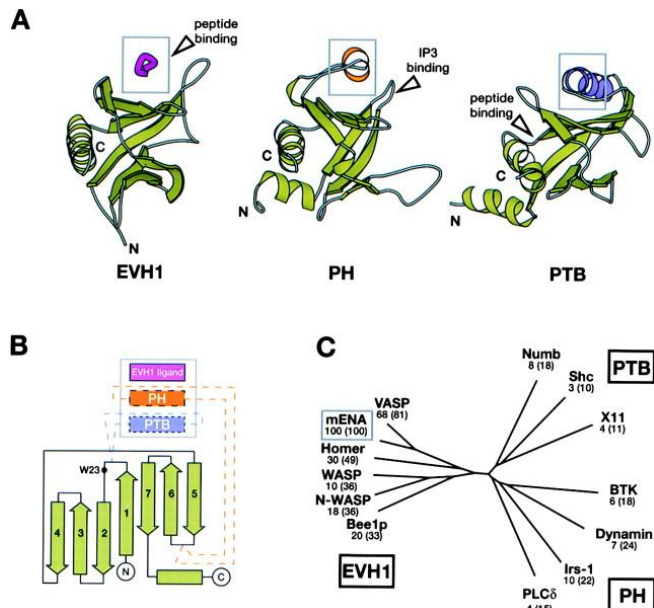


Figure 4. PH domain superfamily and their structural similarity (Prehoda et al., 1999, Figure 5)

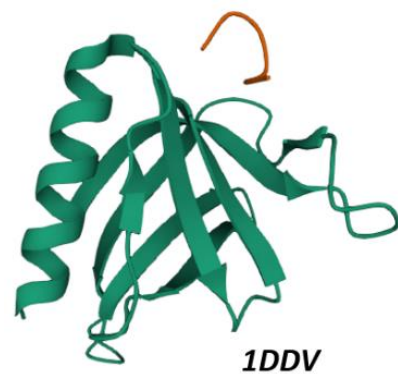


Figure 5. X-ray crystallographic structure of the Homer EVH1 domain with mGluR peptide (PDB ID: 1DDV)

³ <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011993/>, Date of download 16. 09. 2021

⁴ <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000697/>, Date of download 16. 09. 2021

2.5 Pathology of Homer1

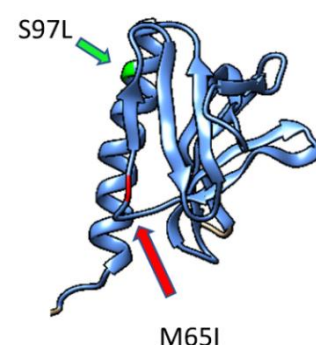
Alterations in Homer1 induce changes in PSD formation, that leads to malformation of dendritic spines (Lu et al, 2017). Dendritic spine shape and density are in close relation with learning and memory, their abnormality or reduction can lead to mental disorders. Homer proteins are heavily involved in brain diseases. Genes encoding those Homer1 proteins that are involved in glutamatergic signalling through mGluR1/5 directly or NMDAR indirectly are candidates for developing neuropsychiatric phenotypes and diseases like epilepsy and schizophrenia. Seven single-nucleotide polymorphisms (SNPs) have been identified, of which three are exonic. Based on knockout mice studies, Homer1 is also involved in behavioural abnormalities - that are consistent with schizophrenia – and addiction (Shiraishi-Yamaguchi and Furuichi, 2007). Changes in the ratio of long and short Homer variants in the central nervous system may inhibit or activate signalling processes, that cause neurological disorders (Lu et al., 2017). This ratio change is observable by sleeping as well, when the expression of Homer1a elevates (promoted by adenosine), therefore disrupts the polymers established by the long variants, which leads to the disintegration of mGluR signalling network and AMPA removal (Diering et al, 2017).

Different studies have described the impact of Homer1 in Alzheimer's disease, which develops as a consequence of the accumulation of the amyloid β protein ($A\beta$), originating from the mutated amyloid precursor protein. In early stages learning and memory impairments can be observed as well as cognition disorders, which is due to the changes in synaptic structure, function and plasticity for sure. Based on *in vitro* studies, $A\beta$ proteins cause the disruption of complexes of long Homer variants including the reduction of glutamate receptor number (Luo et al., 2012), whilst by mouse transcranial magnetic stimulation, Homer1a was a curing factor, independent from glutamate receptors (Wang et al., 2015).

DNA	protein
195G>T	M65I
209C>T	S97L
425C>T	P142L
968G>A	R323H
1080C>T	3'UTR

Table1. Rare, potentially deleterious variants in Homer1 in autism (Kelleher et al., 2012)

Figure 6. Autism related mutations in the Homer1 EVH1 domain (Private picture, done with Chimera)



Autism spectrum disorders (ASD) define a heterogeneous group of disorders (Banerjee et al., 2016), that are characterised by impaired communication and social interaction, as well as restricted and repetitive behaviour. No single brain region can be straightforward associated with a heterogeneous disorder; accordingly, its pathology is also hard to be investigated. Known genetic factors are believed to contribute to less than 20% of cases, but still the background genetics is unexplained. Both single-gene mutations and *de novo* copy number variants (CNVs) have been reported in relation with ASD. In case of the former, genes in charge of neurogenetic syndromes, like genes of the protein translation machinery, FMR1 – fragile X syndrome is the most inherited cause of autism – and genes within

microdeletions, that are in connection with non-syndromic (idiopathic) autism, like SHANK3, NLGN3 and NLGN4 have to be mentioned (Kelleher et al., 2012). The reported mutations in candidate genes encode cell adhesion molecules (NRXN), scaffolding proteins (SHANK, HOMER1), regulators of protein synthesis (FMR1) and degradation (TSC1/2) and chromatin regulators (MECP2). These genes converge on shared pathways, that meet at synapses (Guang et al., 2018). The importance of mGluR signalling pathways have been already reported together with some rare SNVs (single nucleotide variants) in candidate genes (Kelleher et al., 2012), that can also be associated with intellectual disability and epilepsy apart from autism. Based on the common genes (TSC1 and TSC2), it is suggested that non-syndromic and syndromic autisms share pathophysiological processes. HOMER1 has also been reported as an autism risk gene and enables the existence of autism-specific HOMER1 missense variants, that affect functionally significant motifs and regulatory sequences and therefore provides connection between autism-associated gene products (mGluR and Shank). Five missense point mutations were reported in HOMER1 gene associated to autism (Table1), from which two (M65I and S97L) can be found in the EVH1 domain (Figure 6., Kelleher et al., 2012).

2.6 Known interactions of Homer1

The central PSD scaffold is maintained by Shank and Homer proteins, therefore, their interaction is crucial for the appropriate protein network. Both are heavily involved in the maturation and regulation of dendritic spines and synapses. Shank is an extremely long protein of the PSD, with several domains. Its main interaction domain is the PDZ, which is essential in the synaptic localisation of Shank and serves as major component of the NMDA/PSD-95/GKAP/Shank complex (Naisbitt et al., 1999). However, in the Homer-Shank interaction the long (about thousand residues), disordered, proline-rich region of Shank participates, that binds to the EVH1 domain (Kursula et al., 2019). In Shank a single Homer binding site has been reported, which can be found in the disordered region. The consensus PPxxF sequence varies between the different Shank isoforms: PLPPPLEFSN (residues 1563–1572 numbering in *Homo sapiens* host organism) in Shank1 and LVPPPEEFAN (residues 1307–1316 numbering in *Homo sapiens* host organism) in Shank3. This interaction is also a potential biochemical link between Homer and Shank associated receptors, that is essential in the generation of excitatory synapses (Tu et al., 1999).

Homer1 also interacts with mGluR1 α /5 on their carboxy-terminals and is able to regulate their postsynaptic localisation, translation, intracellular signalling and trafficking to the plasma membrane. The impact on surface targeting and clustering of mGluR1 α /5 depends on the Homer splice variant. Mostly long variants enhance cell-surface expression, stimulate intracellular calcium ion release, as well as facilitate the activation of PI₃K (phosphatidyl-inositol-3-kinase) in response to mGluR signalling (Luo et al., 2012).

Transient receptor potential cation channels (TRPC) are components of store-operated calcium ion channels and composed of seven different subtypes (TRPC1-7). Their C-terminus contains a proline-

rich sequence, that possess the Homer-binding site. Interestingly, TRPC1 has the canonical sequence on its N-terminus as well. As one of the key regulators of intracellular Ca^{2+} , IP_3Rs mediates gating of TRPCs. Together with Homer1b/c, they form a complex, which under resting condition prevents TRPC activation. Disruption of the TRCP- IP_3R linkage by a mutation or Homer1a evokes spontaneous activation of TRCP (Luo et al., 2012). IP_3Rs also possess the canonical Homer-binding motif in their cytosatically disposed N-terminals, which enables them to bind to another EVH1 domain in the TRCP-Homer1b/c- IP_3R complex or in an mGluR-Homer tetramer interaction. In brain, IP_3Rs can be found in the spine apparatus, which structure approaches that margin of PSD, where mGluRs are densely packed. mGluRs release Ca^{2+} from intracellular IP_3R -gated stores. The two types of receptors can be connected via Homer1 multimers, which linkage may be disrupted by Homer1a (Tu et al., 1999)

Drebrin is involved in actin building and essential for synaptic spine development and plasticity (Li et al., 2019). Its carboxy-terminus lacks recognisable domains and shows low sequence conservation except for two Homer binding motifs (HBMs), that can bind the EVH1 domain. In cellular context, Homer

tetramers enhance filopodia

elongation and the actin building activity of Drebrin, which is antagonised by Homer1a. This interaction has a key role in the modulation of actin cytoskeleton in synapses (Li et al., 2019).

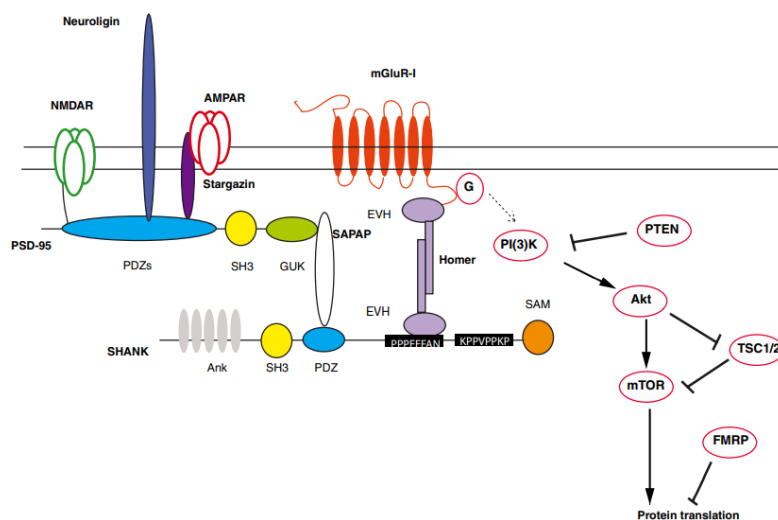


Figure 7. Schematic representation of interacting proteins in the network of the postsynaptic density (O'Connor et al., 2014, Figure 1)

2.7 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy has become a priceless tool for the study and determination of three-dimensional structures of macromolecules and their complexes, especially with respect to intrinsically disordered proteins and other, non-crystallisable systems. Today, three main kinds of structure determination methods are used with which atomic-level resolution can be obtained: NMR spectroscopy, X-ray crystallography and cryo-electron microscopy, from which X-ray is the most widely used and accepted technique. However, only NMR has the capability to ensure investigation on internal dynamics in solution over wide timescales and to characterise transient states at low population levels, that have an importance in macromolecular recognition, assembly, allostery and signal transduction (Bax, et al., 2019).

At its initial application, solution structure determination was based on two-dimensional ^1H - ^1H NMR experiments, where two-dimensional COSY (correlation spectroscopy) maps demonstrate cross-

peaks between pairs of (chemically non-equivalent) geminal and vicinal protons. In the case of TOCSY (total correlation spectroscopy) experiments magnetisation is spread over the total spin system of each amino acid, therefore due to interaction of all protons of a spin system, additional signals to the COSY spectra appear, that confines TOCSY patterns within a certain type of amino acid. Consequently, the type of amino acid can be determined. Here, type does not refer to the 20 chemical structures as the so-called J- and U-type residues contain multiple specific amino acids. Additionally, by NOESY (Nuclear Overhauser Effect spectroscopy) spectra, cross-relaxation of spins are used for the assignment of specific amino acids to the protein sequence. Sequential assignment is the process where we assign the spin systems to a specific residue within the sequence using the information on amino acid type and the characteristic NOE pattern between consecutive residues. Evaluation of cross-peak intensities aids the estimation of proton-proton distances. Provided we have a sufficient number of such interatomic distances, the full 3D structure of a protein can be determined with suitable calculation methods (see below). However, ^1H - ^1H NMR experiments have limitations on protein size, namely that the small range of chemical shifts of protons (few ppm) when combined with the high number of proton pairs cause overlaps of cross-peaks and therefore ambiguities in assignment. To solve this resolution problem, a third frequency axis was introduced, that corresponds to the NMR frequencies of the ^{15}N spins of a ^{15}N isotopic labelled sample. Accordingly, the size threshold of structurally determinable proteins could be increased, but for proteins larger than 18-20 kDa preparation of double isotopically labelled (^{15}N or ^{13}C - ^{15}N) samples are needed to complete the resonance assignment step. Due to the correlation of three different nuclei these experiments are also named as “triple-resonance” and have the advantage on the reduction of spectral overlapping (Bertini & Parigi, 2012 B).

2.8 Resonance assignment

Several types of triple-resonance experiments are defined, which are named after all the nuclei that are involved in magnetisation transfer in the order of their use. Names of those nuclei, whose frequencies could not be observed in the spectra but are involved in magnetisation transfer are in brackets (Bertini & Parigi, 2012). Resonance assignment is the process of assigning chemical shifts in the spectra to the certain atom of the certain amino acid of the molecule measured.

Two-dimensional spectra of only ^{15}N -labelled sample:

- ^1H - ^{15}N HSQC: the covalently linked protons and nitrogens show correlation. Since – apart from some amino acids, like arginine, lysine and histidine that has more N-H bond, and from proline and the first residue of the sequence, that lack it - the only N-H linkage is the amide bond, the number of peaks in the spectra is at first approximation equal to the number of amino acids. The position of the amide cross-peaks is highly sensitive on the local environment thus, the HSQC is often unique for the protein and the conditions used. Therefore, this spectrum is also called the fingerprint of a protein. By assignment, peaks of the HSQC should be identified by the residues.

- HNHA and HNHB: Identify the α (or β) hydrogen ($H\alpha$ or $H\beta$) – linked to the α (or β) carbon atom - of the current residue. $H\alpha$ values are corrected with so-called random coil shifts – often termed secondary shifts - are characteristic for local conformations like α -helix or β -sheet.

Three-dimensional triple-resonance spectra:

- HNCO: Correlates the carbonyl ^{13}C carbon with the amide N-H group, therefore one can assign the carbon atom of the carbonyl group of the amino acid. This means, if there is a peak in the ^{13}C dimension of the HNCO at the position of an HSQC peak, then that peak is a certain backbone N-H of a residue, although side chain amides of Asn and Gln residues can also produce a similar pattern.
- HN(CO)CA and HNCA: Enable the identification of the $C\alpha$ of the previous/actual residue in relation to the N-H group. This is the basis of ordering peaks in the spectra as a function of the amino acid sequence.
- HN(CO)CACAB and HNCACB: Identification of $C\beta$ of the previous/actual residue. Complements and confirms sequential assignment after $C\alpha$ identification.
- HCCH-TOCSY: Identification of strips at each carbon frequency in the side chain of one residue, where all side chain hydrogens of the actual and neighbouring carbons are seen, therefore serves as basis of side chain assignment.
- TOCSY-HSQC: Used for characterisation of residue types, based on the number and position of side chain hydrogens.
- NOESY-HSQC: Determines those hydrogen atoms, that are in close vicinity.⁵ Its drawback relative to the homonuclear NOESY is that only distances were one of the amide N-H can be observed. The advantage is the lesser signal overlap discussed above.

2.9 Protein dynamics

The largest advantage of NMR spectroscopy lies in the investigation of internal dynamics. Chemical and conformational exchange caused by the magnetic environment induced modulation of chemical shifts and linewidths are directly incorporated in NMR spectra and relaxation. Therefore, information can be acquired on kinetic processes in microsecond-second timescale, which is characteristic for protein association/dissociation (Bertini & Parigi, 2012, A).

Relaxation is the common name of the processes when previously perturbed spins return to their equilibrium state. Nuclear spin relaxation is mainly due to time-dependent stochastic modulations of interacting spins, that lead to their reorientation with respect to each other and the external magnetic field. Relaxation processes of protonated heteroatoms (^{15}N , ^{13}C) are mainly determined by their interaction with the ^1H and the anisotropy of the chemical shift, therefore providing site-specific site-

⁵ <https://www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/> Date of download: 02. 11. 2021

specific dynamic information. Parameters of relaxation that can be experimentally determined are R_1 (spin-lattice relaxation rate; inverse of spin-lattice relaxation time T_1), R_2 (spin-spin relaxation rate; inverse of spin-spin relaxation time T_2) and heteronuclear NOE (Nuclear Overhauser Effect between nuclei). Ratio of R_2/R_1 (equal to T_1/T_2) can be used to estimate the rotational correlation time (τ_c), that is determined by the rotational diffusion tensor – depending on molecular tumbling – and the shape of the molecule (Wang et al., 2009). The correlation time of intrinsic local motions is labelled by τ_e . The amplitude of internal motions of individual amide N-H groups (or methyl side chains) is determined by S^2 , the generalised order parameter. This is generally derived from the above-mentioned relaxation measurements and in this case it is characteristic for motions on the picosecond-nanosecond time scale. Its value varies between zero - that means free, isotropic state – and one - rigid state. R_2 relaxation rate can be influenced by the chemical exchange constant R_{ex} , that describes the transition between two conformation states with different chemical environments (Bertini & Parigi, 2012, A). Dynamical parameters (S^2 , τ_e , R_{ex}) can be used to characterise each and every amino acid, that can be fitted by several models, taking their different combinations into account. Increasingly complex models are only accepted if they provide significant improvement compared to the simpler ones with fewer parameters.

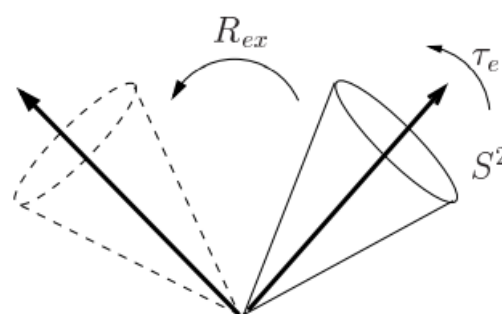


Figure 8. Possible interpretation of parameters of protein dynamics: The thick arrow of NH vector moves along the envelope of a cone, that is characterised by its angular field (order parameter, S^2) and characteristic time (τ_e). In case of bigger conformational reorganisation, the NH vector jumps into other position, but has similar movement (characterised by the exchange constant, R_{ex}).

2.10 Structure calculation

For protein structure calculation, several conformational restraints are used, that mainly include hydrogen-hydrogen distance limits and/or dihedral angle restraints, that additionally can be supplemented with residual couplings (RDCs). In NMR structure calculation usually a molecular dynamics-based approach is used, where a random structure is used as a start and during the calculations the structure is modified according to the restraints, then the relative weight of geometric parameters increased, keeping as much restraints as possible.

The output of NMR structure calculation is typically a set of several conformers that were calculated by the same conditions and the best ones in terms of restraint correspondence and geometry are usually selected. We also expect that these conformers are highly similar, representing the same structural state.

3. METHODS

3.1 Resonance assignment

NMR based determination of protein 3D structure relies on sequence-specific resonance assignment, which is a time-consuming and hard task in case of more than one hundred residues with all their signals. First of all, spectra have to be checked, in the way, that the number of N-H crosspeaks (not counting the signals of side chains of basic amino acids) should approximately correspond to the number of residues in the protein sequence, there should not be much overlapping. The main point of backbone assignment is to find correspondence between the peaks of the HSQC spectra and the residues of the particular protein sequence. This task can be done with various software (XEASY, Sparky, CCPNMR), here backbone assignment was done with the use of CARA (Computer Aided Resonance Assignment), which is sufficient to proceed to protein dynamical calculations (Keller, 2015). Side chain assignment – required for structure calculation - needs additional spectra and tools with additional spectrum display capabilities, therefore it was performed with CCPNMR (Collaborative Computational Project for Nuclear Magnetic Resonance Assignment)⁶. During my bachelors, I was also dealing with this protein, producing a double labelled sample for NMR measurement, which contained a 6xHis tag in its amino terminal. In case of interaction measurements and structure calculation, the presence of a polyhistidine tag is better avoided. Therefore, the 6xHis tag was cleaved off from the double labelled sample and sent to NMR, but the evaluation of new spectra was not completed until now. In the upcoming paragraphs, I would like to describe the process of protein assignment, but it has to be mentioned, that after a point the expertise of a specialist was needed. For the sake of completeness, in this Thesis I describe the main aspects of side chain assignment and structure calculation, but these steps were done with considerable help from Dr. András Czajlik, who performed the major part of these tasks along with teaching me the necessary know-how.

3.1.1 Backbone assignment

NMR spectra and the protein sequence were loaded into CARA. First an automated peak picking was used, where peaks in the HSQC spectra can be labelled above a user-defined threshold. Then, the potential peaks were saved as a peak list, that can be loaded into SynchroScope. SynchroScope is a window of CARA, with which spectra can be projected onto the HSQC, therefore one can navigate in three dimensions. Peaks were defined in the HSQC where a corresponding peak could be found in the HNCO, representing the carbon atom of the carbonyl group. Then the HN(CO)CA – identifying the C α of the previous amino acid (C α -1 label) -, the HNCA – C α of the current amino acid -, HN(CO)CACB - C β of the previous amino acid (C β -1 label)-, and HNCACB - C β of the current amino acid – were projected onto the HSQC in series. In each spectrum the maximum intensity of the peak identifies the chemical shift value of the corresponding atom. Based on these assignments, one can have a preliminary

⁶ <https://www.ccpn.ac.uk/>, Date of download: 11. 11. 2021

thought on the amino acid type and the order of peaks in the HSQC. Amino acids, like glycine, serine, threonine and alanine can be pillars in assignment, due to their characteristic $C\alpha$ and $C\beta$ shifts. Glycine has a $C\alpha$ about 45 ppm and no $C\beta$, alanine has high $C\beta$ about 30 ppm, serine and threonine have $C\alpha$ about 40 ppm and low $C\beta$ at about 60 ppm. Glycine and alanine have characteristic side chain protons as well. Further types of amino acids can be characterised based on ^{15}N HSQC-TOCSY spectra (Figure 10.), where hydrogen shifts can be visualised for each residue. CARA has another helpful window, called StripScope, where the residues can be placed next to each other as strips based on their $C\alpha$ and $C\beta$ shifts. This means, that if a $C\alpha$ and $C\beta$ of an amino acid corresponds to the $C\alpha-1$ or $C\beta-1$ of another residue, then they are probably neighbours in sequence. The problem is that there is usually more than one match, therefore one has to select the best match, then align the selected one to the sequence. In practice, usually longer matching stretches are needed to unambiguously define the assignment of a given region.

Presence of good quality ^{15}N HSQC-NOESY and ^1H - ^1H NOESY spectra are essential as control for correct backbone assignment, that help to find spatially close hydrogens. Besides ^{15}N - ^1H spectra, ^{13}C - ^1H HSQC can be used for example for the identification prolines in the spectrum – these residues give no signal in ^{15}N HSQC due to their missing amide N-H as well – as a start for side chain assignment, by the description of $C\alpha$ - $\text{H}\alpha$ and $C\beta$ - $\text{H}\beta$ correlations. The amino-terminal region, which is also missing from the traditional ^{15}N HSQC, can be determined with the usage of HACAN spectra.

3.1.2 Side chain assignment

For side chain assignment, ^{15}N HSQC-TOCSY spectra sometimes do not provide the required signal quality. It can be supplemented by HNHA and HNHB, which are practical for finding α and β hydrogens of each residue (β not necessarily exists) and the previously mentioned ^{13}C HSQC. After assignment of $\text{H}\alpha$ and $\text{H}\beta$, HCCH-TOCSY is the key spectrum for the identification of side chain chemical shifts, and also serves as a control for checking spin system types. This spectrum was loaded into CCPNMR in two rounds, once, when x and y axes represent carbon and z hydrogen chemical shifts (description of side chain carbon atom along y), and once when the z axis is a carbon (previously defined), while x and y represent a carbon and a hydrogen respectively (determination of the hydrogen of the carbon of the z -axis along y). In the easiest case, this would mean the following: first x is $C\alpha$, y is $C\beta$, and z is $\text{H}\alpha$, if we know $C\alpha$ and its $\text{H}\alpha$, then $C\beta$ can be identified; second x is $C\alpha$ again but now y is $\text{H}\beta$ and z is $C\beta$ if we know the two carbons, then $\text{H}\beta$ can be determined. Given, that as a result backbone assignment, we have a strong hypothesis about the types of spin systems in the spectra, it is practical to start side chain assignment with residues with “easier”, shorter, non-noisy and non-overlapping side chain. Most difficult cases are lysine and arginine, that are basic amino acids and have long side chains with often overlapping chemical shifts. Side chain N-H groups might can be identified with the use of 2D NOESY spectrum and 3D ^{15}N HSQC-NOESY projection. Another difficult task is the assignment of aromatic side chains (tryptophan, tyrosine and phenylalanine), for which CBCGCDHD and

CBCGCDCEHE spectra were used, that are named after the involved carbons in order (β , γ , δ and ϵ) and the visible hydrogen of the last carbon. The most problematic thing is to assign lysine side chains, partly due to the difficulty in differentiating between resolved and overlapping geminal proton signals of $H\beta$, $H\gamma$, $H\delta$, and $H\epsilon$. Here the combination of nearly all the previously mentioned spectra is needed, but it has to be mentioned, that full assignment can be hardly ever reached.

The outcome of protein assignment is a list of chemical shifts of the identified atoms of each residue. These chemical shifts have to be properly referenced to be exactly comparable between measurements. The NMR spectra were recorded by Professor Dr. Gyula Batta at the University of Debrecen with a Bruker Avance 700 spectrometer at 298K and DSS (sodium trimethylsilylpropanesulfonate, $C_6H_{16}O_3Si$) as reference material. Therefore, the chemical shifts of all carbons have to be corrected with 2,64 ppm and of hydrogens with 0,084 ppm.

3.2 Analysis of protein dynamics

In practice, evaluation of protein dynamics starts with processing NMR spectra. On one hand, two separate heteronuclear NOE (hetNOE) spectra, on the other T_1 and T_2 relaxation data are needed. The Bruker pulse sequence library used by Prof. Batta uses a measurement with interleaved recording of the hetNOE spectra with and without irradiation and pseudo-3D spectra for T_1 and T_2 relaxation where the third dimension is the relaxation delay applied. These spectra were processed with Bruker's TopSpin software to extract the individual spectra and then peak analysis was performed with NMRFAM-SPARKY⁷ (Lee et al., 2015). First, longitudinal (R_1) and transversal (R_2) relaxation rates together with NOEs have to be determined. For this, peak intensity measurement and assignment had to be done and then exponential function had to be fitted to the integrals. This step was done with NMRFAM-SPARKY. Then, a model has to be fitted based on the four files, for which a model-free analysis software, Tensor2 was used. Finally, τ_c can be estimated from an existing protein structure, for which our preliminary structure on the EVH1 domain was used.

3.2.1 The NMRFAM-SPARKY program

NMRFAM-SPARKY is a software for NMR data analysis, that enables picking, assigning, and integrating peaks with the use of a graphical interface. It was developed for aiding in structure determination and dynamical characterisation of proteins and nucleic acids. The program can handle any number of two- or three-dimensional spectra at the same time (Lee et al., 2015).

In this study NMRFAM-SPARKY was used in the evaluation of protein dynamics. It received an input list of N-H chemical shifts, that can be generated from the assignment. Based on them, peaks were read for the UCSF converted files of T_1 and T_2 relaxation spectra separately, then relaxation peak heights were calculated, from which T_1 and T_2 relaxation rates could be extracted. NOE spectra also had

⁷ <https://www.cgl.ucsf.edu/home/sparky/manual/index.html>, Date of download: 11. 11. 2021

to be converted into UCSF files on which the N-H chemical shift list were applied and peak height calculated.

3.2.2 The MODELLER program

MODELLER is a program for homology and comparative modelling of three-dimensional structures of proteins, based on an input sequence alignment and a known, related structure, to which the sequence has to be modelled. A model, that contains all non-hydrogen atoms is calculated, then a comparative protein structure modelled based on spatial restraints. The program is also able to perform *de novo* modelling of loops, model optimisation, multiple alignment and comparison of protein structures and clustering, searching sequence databases.⁸

The program can be run with Python, through the modeller module. First the EVH1 sequence of the H118 was aligned to the template structure of 1DDW.pdb, that is performed through a gap penalty function, that reduces alignment errors. Then a 3D model of the target is calculated from the target-template alignment with MODELLER's AutoModel class. As a result, five models with coordinates are constructed in PDB format. From the five models, the "best" had to be selected based on DOPE score and then evaluated based on dope potential.

3.2.3 The HYDROPRO program

HYDROPRO calculates basic hydrodynamic properties from coordinates of the input PDB structure for macromolecules like proteins and small nucleic acids. This involves the calculation of translational diffusion coefficient, sedimentation coefficient, intrinsic viscosity, relaxations times, radius of gyration and optionally some solution and scattering related properties. As an input PDB file, our preliminary structure on the EVH1 domain was given (Ortega et al., 2011).

3.2.4 The TENSOR2 program

For model fitting TENSOR2, a model-free software was used. It gets a PDB structure and, the SPARKY output relaxation data as input, then (with the exclusion of some residues with extreme τ_e and R_{ex} values) isotropic and anisotropic models can be fitted. From this, isotropic correlation time (τ_c) is calculated. Isotropic models can be fitted for globular proteins with approximately spherical shape, like the EVH1 domain. Finally, internal mobility parameters are calculated, that give an output list of all five dynamical parameters for each residue. Depending on the motional model providing the statistically best fit, some of these parameters can have a value of 0.

3.3 Protein structure calculation

Manual resonance assignment is preferred to automatised since the latter might distort the calculation. In order to determine a protein structure, the complete NMR assignment – backbone and

⁸ <https://salilab.org/modeller/> Date of download: 25. 10. 2021

side chains – are needed, from which a peak list is generated, that has to be imported into 2D NOESY, 3D ¹⁵N NOESY and 3D ¹³C HSQC-NOESY spectra separately. 2D NOESY has good resolution but high spectral overlapping, while the other two has opposite characteristics. Based on NOESY spectra the distances are known, but the identity of spots, between which these distances are measured, are not, therefore structure calculation is usually linked to a map, on which cities should be localised based on distances between them. Structure calculation is expected to be reasonable, when the number of NOE is sufficiently high, around 13-20 per amino acid residues.

The longest part of calculation are the iterative steps (here seven were used according to the standard AtnosCandid protocol applied), when restraints are defined based on the spectra and then adjusted based on the structures calculated with the given restraint set. In addition to the NOE restraints, the TalosN program was used to predict backbone assignment torsion angles from the chemical shift list and these angles were also added as restraints to help define the secondary structure of the protein. This prediction is done for every amino acid, that are characterised based on their reliability. Angle values with confidence classification as strong or generous should only be used. The exact structure calculation was done with AtnosCandid using the Cyana program, to which chemical shift lists of the above mentioned three spectra, the TalosN output, and the protein sequence had to be given as inputs. This protocol uses a kind of “black box” method, it is fully automated. In the output files several conditions have to be checked: the agreement, which refers to the appropriety of the input peak list and has to be above 90%; the average target function (below 100 counts good); and by the seventh cycle the average value (below 1 great), and average RMSD (root mean squared deviation), that calculates 20 structures (below 1,2 preferred). In the separate folder of the last cycle, Ramachandran statistics and violations (distance angle, torsion angle and van der Waals) have to be checked additionally. By the former, no disallowed geometry should appear, and considering violations, all should be zero. If there are restraint violations present, the assignment has to be rechecked, especially for aromatic rings and for glutamine, asparagine, lysine and arginine side chains. The fully automatic nature of this method is also its disadvantage as there are few modifiable parameters and the inclusion and exclusion of chemical shifts is a rather crude and often not fully justifiable way of improving the calculations. On the other hand, manual NOE assignment and modification of the restraint list is a cumbersome work and it can take very long to obtain an acceptable structure with non-automated methods.

3.4 In vitro mutagenesis of the EVH1 domain

3.4.1 The pEV2 expression vector

In our laboratory, we use *Escherichia coli* bacteria for protein expression. These bacteria can take up plasmids – circular DNA molecules replicated independently from chromosomal DNA -, containing genetic elements that help the expression of the DNA of interest, therefore we also call these plasmids expression vectors. The insert DNA coding the wild-type Homer EVH1 domain was ligated into a pEV2 vector (a modified pET-15b plasmid that bears TEV cleavage site instead of Thrombin site) between the NdeI and the BamHI restriction cleavage sites. This pEV2 expression vector contains a promoter and a terminator region before and after the insert, an ampicillin resistance gene (ensuring the reproducibility of only those genes that took up the plasmid) and a replication origin. A 6xHis tag and a TEV protease cleavage site can be found upstream the insert, so that the 6xHis tag can be cleaved off by TEV protease treatment.

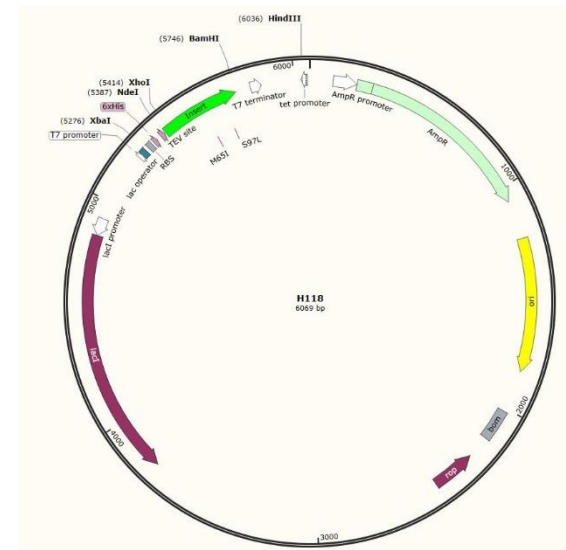


Figure 9. The pEV2 expression vector with the V2-H118 insert (green arrow), 6xHis tag and TEV cleavage site (light purple), ampicillin resistance gene (light green), replication origin (yellow), lacI repressor (purple). Mutation and restriction cleavage sites are labelled.

3.4.2 Molecular cloning

Primers were designed for the Homer EVH1 M65I and S97L mutants separately. The 5' ends of these outward pointing primers were positioned back-to-back in order to amplify the entire plasmid. In addition, one of them contained the mutation in its centre or closer to its 5' end. These primers were used with the wild-type V2-H118 plasmid template, which got its name from the hosting plasmid (pEV2) and the encoded protein construct (118 amino acids long Homer EVH1 domain). Using these primers on the V2-H118 template, one can generate linear DNA fragments of the template, containing either the M65I or the S97L mutation respectively.

The following PCR reaction mixture was used for the amplification: 4 µl 5xHF buffer, 0,4 µl 10 mM dNTP, 0,2 µl Fusion High Fidelity Polymerase, 2 µl template (5 ng/µl), 0,5-0,5 µl forward and reverse primers (10-10 µM) and water up to 20 µl total volume per each reaction. PCR protocol consisted of a 98° C, 1 minute initial denaturation, followed by 35 cycles of denaturation (98°C – 10 s), annealing (Ta – 30 s) and extension (72°C – 3,5 min), and finally 72°C – 5 minutes end termination. In the course of denaturation, the double strands of the DNA were divided, while primers were paired with the template through base complementary during annealing. In the third phase, the extension, the desired

DNA fragment was synthesized through dNTP incorporation. The synthesised PCR products were purified after the PCR with gel extraction.

The generated linear PCR products have to be phosphorylated and ligated in order to circularise them. This means the use of two enzymes: PNK (polynucleotide kinase) and T4 DNA Ligase. At first, DNA sequences were denatured, with 10 minutes of 70°C heat shock followed by rapid cool down on ice. Then they were phosphorylated with PNK treatment, in the following mixture: 10-50 ng DNA, 2 µl 10x T4 Ligase Buffer, 1 µl 50% polyethylene-glycol 4000, 0,75 µl PNK and water up to 15 µl. PNK reactions were incubated at 37°C for 30 minutes, then 10 minutes 75°C heat treatment was executed to inactivate the PNK enzyme. The heat sensitive ATP provides the phosphate group for ligation, therefore additional 0,5 µl of ATP was added per reaction after the heat treatment. Then the already phosphorylated PCR products could be circularised with 0,5 µl of T4 Ligase enzyme per reaction. This process was followed by additional three cycles of 22°C, 10 minutes and 17°C, 20 minutes with the lowest ramp rate. Due to the similar mobility of circular plasmid and linear PCR product, it is not possible to get rid of all the wild-type (circular) plasmid, that had been used for template, thus they can form colonies as well. In order to ensure successful ligation, preparation of negative control samples is necessary. Since two enzymes were used, two negative controls can be prepared: one that lacks the PNK enzyme and one that lacks the ligase. In both cases only the originally circularised plasmid templates can form colonies. Molecular cloning was regarded successful, if the number of colonies were much higher (3-4 times) on sample plates than those on the negative controls.

3.4.3 Transformation and plasmid isolation

Escherichia Coli DH5a cells were used in the laboratory for plasmid DNA propagation. The process of transformation is when we make bacterial cells to take up plasmids. It has low efficacy (one bacterium usually takes up only one plasmid), but enables us to isolate some of the ligation products. Ligation products (2,5 µl from each) were transformed into 25 µl of competent cells per reaction, stored on ice, then 44°C, 1,5 minutes heat shock was given to them, in order to make the bacterial cell membrane fluid and enable DNA molecules to get inside their cytoplasm. Thereafter cells were placed on ice to cool them, then 100 µl LB media were given to each of them, followed by 30 minutes incubation at 37°C. 100 µl were plated from each onto ampicillin containing plates, that ensures the colonisation of only those bacterial cells that took up a plasmid. Plates were stored at 37°C overnight. If the transformation is successful, one can see colonies on the plates next day and some of these colonies can be picked up and inoculated in LB-Amp media for mini liquid culture.

For plasmid isolation Thermo Scientific GeneJet Miniprep Kit was used separately for every cell culture. At first, cells of the mini liquid cultures were centrifuged with 4500 rpm for 5 minutes, then each pellet was suspended in resuspension solution. Thereafter, lysis buffer was added for the decomposition of membrane lipid structure and the denaturation of DNA and protein contents. This was followed by the usage of neutralisation buffer, that precipitates resolved proteins, membrane

components and genomic DNA, but not the plasmid DNA. Therefore, after a centrifugation, plasmid DNA was found in the supernatant, which was placed onto a silica mini column, that bounded it. With a double washing step, everything could be removed from the column, but the plasmid DNA. The pure plasmid DNA was brought down with Elution buffer. After this step, plasmid concentration was measured with Nanodrop2000 photometer and plasmids were stored at -20°C.

Due to the needs of other research studies of our laboratory, molecular cloning of mutant EVH1 constructs were done by Dr. Bálint Péterfia.

3.4.4 Sanger sequencing

Sanger sequencing – also named as chain determination method - was developed by Frederick Sanger in 1977 and defines the order and identity of bases in a DNA sequence. This is a highly accurate and widely used technique with long-read capabilities. Sanger sequencing makes use of dideoxynucleotides (ddNTPs), which are the chemical analogues of the DNA monomer deoxyribonucleotides (dNTPs). ddNTPs do not possess 3' hydroxyl group, therefore the bond with the next dNTP cannot be formed and the DNA cannot be elongated. By mixing ddNTPs that are labelled with fluorescent dyes with different colour for each base, one primer, unlabelled dNTPs and the template DNA into a DNA extension reaction, strands of each possible length will be produced due to the random incorporation of ddNTPs, that terminate DNA synthesis. The identity and order of bases can be detected with capillary electrophoresis (Heather and Chain, 2016). Sanger sequencing of mutant EVH1 sequences was done by Eurofins Genomics.

3.5 Protein production and expression

Escherichia coli bacteria are widely used, due to their quick and cheap growth (they consume nearly all kind of nutrients) and fast replication (division in half an hour). For protein expression these bacteria are used in our laboratory as well, but before, plasmids have to be transported into the appropriate competent cells. In order to produce protein in a higher amount from plasmids with T7 promoter, BL21(DE3) *Escherichia Coli* bacteria were used. This means, that T7 RNA polymerase gene was integrated into the bacterial BL21 genome, with DE3 bacteriophage. Such expression vectors, like pEV2 (or pET25b as well), that carry a lac repressor gene, a lac operator and a T7 and a natural promoter, when used in DE3 lysogens, the lac repressor represses transcription of the T7 RNA and the T7 promoter blocks transcription of target gene as long as no lactose analogue is given.

Transformation of plasmids into BL21(DE3) cells is similar to transformation into DH5α cells. In our case, competent BL21(DE3) cells provided a much smaller number of colonies as usual, therefore 5 µl (100 ng) plasmids were transformed into 50 µl of them. This was followed by heat shock (1,5 minutes, 44°C), 5 minutes incubation on ice, addition of 250 µl LB media and incubation at 37°C, 220 rpm for 30 minutes. 100 µl were plated on ampicillin containing LB plates per construct.

Due to the fact, that there was no protein expression protocol at hand for the mutant EVH1 constructs, we assumed that they should be treated as the wild-type one. I had to repeat transformation and plating of the M65I and S97L mutants several times, due to the low competency of BL21 cells. Precultures were prepared by inoculating colonies grown on the plates into 7 ml LB-Amp media, which were incubated overnight thereafter at 37°C, 190 rpm. Since the precultures were dense enough next day, they could be poured into 500 ml of sterile LB media and grown at 37°C, 130 rpm until they reached the appropriate density. This density is measured in McFarland Units (MFU), which is an artificially determined unit, based on that in biotechnology, turbidity correlates with the troubling of the media – the more turbidity, the higher amount bacteria. Protein expression should be induced at 4-5 MFU with the addition of 1 mM isopropyl- β -D-thiogalactoside (IPTG), a lactose analogue, that can bind to the lac repressor protein, and dissociate together from the operator region and therefore enable transcription. This makes us able to induce protein expression the time when it is suitable for us. IPTG induction was followed by incubation at 20°C, 140 rpm overnight. Unfortunately, transformation of S97L mutants was successful only for the third time, therefore their induction happened on a Friday afternoon, which meant, that their incubation lasted at 20°C for one day, and at 10°C for additional 1,5 days. Upcoming workday, cells were centrifuged in two rounds (20 minutes and additional 10 minutes at 4°C, 4500 rpm). Then supernatants were poured off and pellets were stored at -20°C.

In order to investigate the effects of point mutations on protein structure and dynamics, we decided to compare their so-called fingerprints, the HSQC spectra of the wild-type and mutant EVH1 domains. Accordingly, ^{15}N stable isotope labelled protein had to be prepared from each construct for NMR measurement. The production of these proteins only differs from the above-mentioned protocol, that M9 minimal media was used for cell cultivation instead of the LB. When the colonies were observable on the plates, 50 ml M9 minimal media was prepared for pre-culture per construct, in which heavy nitrogen isotope (^{15}N) was given in the form of ammonium-chloride as the only nitrogen source. Then 3-4 colonies were inoculated into that solution, except for S97L mutant, that gave only one colony. These liquid precultures were incubated overnight at 37°C, 130 rpm. They got denser next day, therefore additional 450 ml M9 minimal media was added for each one. The bacteria were grown at 37°C, 130 rpm, until they reached 4-5 MFU, where IPTG induction was performed. Protein expression was carried out at 20°C, 130 rpm for 16-20 hours, then cells were harvested by a 15 min, 4500 rpm centrifugation at 4°C. Cell pellets were stored at -20°C until protein purification.

Unfortunately, protein production of isotope labelled samples altered from the previously described in the way, that the pre-cultures contained 6% of ^{13}C out of the total carbon source of bacteria. Based on the fact, that natural carbon isotopes are preferred to heavy ones and in the precultures the nutrient uptake of essential proteins is prior to the recombinant's, the ^{13}C content of the final recombinant protein was expected to be 2-3%. Therefore, this failure has no effect on the purity of the sample, but it had to be considered in NMR relaxation and mass spectrometry measurements.

3.6 Protein isolation

For protein isolation the same protocol was used for both unlabelled and ^{15}N -labelled proteins. Wild-type and mutant constructs were simultaneously purified under the same conditions (buffer, pH, temperature etc.), in order to be certain, that experienced differences were not due to technical issues. Since the produced proteins are within bacterial cells in the previously frozen pellet, they had to be dissolved in Lysis Buffer (300 mM NaCl, 50 mM NaPi, pH 7,4) - 1,5 g pellet in 15 ml - and then bacterial cell wall had to be broken, for which an Ultrasonic Homogeniser was used (BioLogics). This process was done on ice, to avoid protein denaturation, and in 10-15 minutes blocks, followed by 20 minutes centrifugation at 4°C , 11000rpm.

The main goal of chromatographic protein purification is to separate the recombinant protein of interest from every other molecule. First, immobilised metal affinity chromatography (IMAC) was performed. For this a NUVIA nickel column was used, that possess such nickel beads that are able to bind polyhistidine-tags with high affinity, like the 6xHis tag on the N-terminal of the Homer protein. The NUVIA column had to be regenerated before usage and connected to the FPLC (Fast protein liquid chromatography) instrument. After equilibration of the column, the sample was injected and the column bounded the 6xHis tag, therefore the whole protein. Isocratic purification was applied, with the usage of Binding buffer (300 mM NaCl, 50 mM NaPi, pH 7,4), when many impurities were removed from the column. It was followed by a washing step, using Binding buffer supplemented with 50 mM imidazole, with which one can get rid of endogenous proteins binded slightly on the nickel beads, but not the 6xHis tagged proteins. Then 250 mM imidazole was applied, that eluates the recombinant protein from the column. The volume of elution fractions were 13 ml per construct, therefore 13 μl DTT (1M) and 300 μl TEV protease ($\sim 0,3$ mg) - in order to remove the 6xHis tag and the linker region before the TEV cleavage site - and 13 μl NaNs (sodium-azide) - to reduce the possibility of contamination - were added to each.

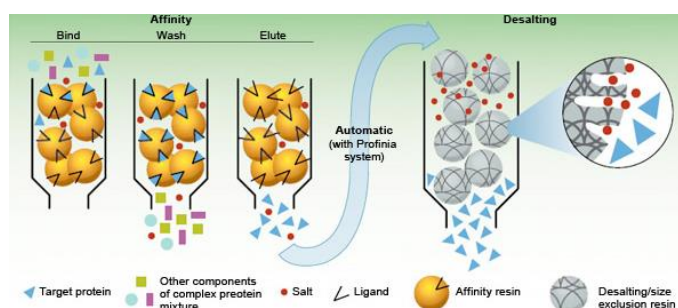


Figure 10. Schematic picture of the processes of immobilized metal affinity chromatography and buffer exchange⁹

⁹ Bio-Rad 2021: <https://www.bio-rad.com/en-hu/applications-technologies/affinity-chromatography?ID=LUSMJIDN>, Date of download: 10. 10. 2021

Ion exchange chromatography (IEC) enables the separation of charged molecules. Ions can be either positively or negatively charged, that are able to reversibly bind to the protein of interest. This process was executed with the usage of a positively charged 5 ml Bio-Scale Mini Macro-Prep S Cartridge, which is able to bind the negatively charged Homer protein at pH 7,4. Since IEC is based on charge difference, the starting buffer has low salt content (20 mM NaCl, 50 mM NaPi, 7,4pH), the end buffer has high (1000mM NaCl, 50mM Napi, pH 7,4). Accordingly, between IMAC and IEC desalting step had to be performed, when the sample was taken into low salt buffer. Desalting allows only the injection of 10 ml from the elution fraction of IMAC, otherwise the elution fraction of buffer exchange would be contaminated with high salt content. IEC starts with equilibration with the starting buffer.

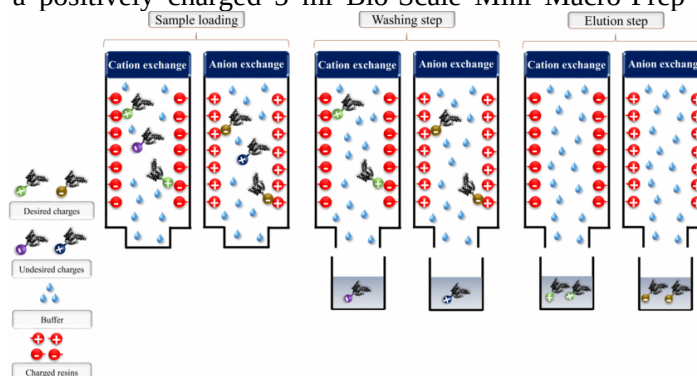


Figure 11. Schematic picture of the process of ion exchange chromatography (Gomari et al., 2020)

Then, sample can be loaded and immobilized on the column. Then a washing step comes, when one can get rid of the unnecessary, unbounded proteins. The process ends with gradient elution with the application of an increasing ratio of high salt buffer. In this step the binding between the charged groups of the column and the protein of interest is broken, therefore the protein comes down in the elution fraction. The column was washed with 10 column volume of high salt buffer between the IEC runs of different Homer constructs.

In case of isotopically labeled proteins, an additional chromatographic purification, a size exclusion chromatography (SEC) is needed, that ensures high purity for NMR measurement. Before SEC, protein samples have to be concentrated into 1 ml, because that is the maximal volume, that can be injected onto the SEC column (SEC70 analytical gelchromatographic column, Bio-Rad). Sample concentration was done by an ultrafiltration column in several rounds of centrifugation at 4°C, 4500 rpm, until samples reached a 10mg/ml concentration. After equilibrating the SEC column with the low salt buffer, samples were injected. In order to avoid cross contamination, the whole FPLC system was washed with buffer between the purification of the different constructs. Elution fractions came down 13 ml after injection, in 3 ml, to which 3 µl of sodium azide was given. Since a highly concentrated sample is needed for NMR measurements, additional centrifugations were performed on the samples with the ultrafilter unit. During this process, concentration of the samples were monitored with NanoDrop2000 in order to avoid protein precipitation. Sample purity and molecular weight were checked with SDS-PAGE (Sodium-dodecyl-sulfate polyacrylamide gelelectrophoresis).

3.7 Spectroscopic measurements for protein structure description

3.7.1 Circular dichroism spectroscopy (CD)

Circular dichroism spectroscopy is used for the identification of protein secondary structure, including the description of changes of structural elements as a consequence of mutation. On the other hand, CD has low resolution, therefore – in contrast to NMR and X-ray -, one cannot get information at amino acid level. The method relies on the differential absorption of the two components of plane polarised light, that are rotating in opposite direction – left (L) and right (R). Absorption of chiral chromophores in asymmetric environment is different, by left or right polarisation, that results in radiation with elliptical polarisation, which is measured as molar ellipticity (θ). As a function of wavelength HT voltage [V], absorbance (used for normalisation if needed) and CD [mdeg] values are plotted after measurement, the latter curve represents the secondary structure: if α -helix is dominating, then a negative band can be observed in the range of 210-230 nm, while by high β -strand ratio a positive (around 195-200 nm) and a negative (at 216 nm) peaks are characteristic (Buday et al., 2018, A). Measurements were done between 195 and 250 nm, with cuvette of 1 mm pathlength and 10 μ M concentration samples at temperatures increased by 10°C, between 20°C and 60°C in order to check temperature dependence as well.

CD measurements were done at the Separation Science Research and Education Laboratory of Eötvös Lóránd University on a Jasco 1500 CD spectrometer (A040661638) with the help of Viktor Dr. Farkas.

3.7.2 Tryptophan fluorescence

Residues with aromatic side chains (tryptophan, tyrosine and phenylalanine) are fluorescent, from which tryptophan is the most sensitive to chemical environment. These amino acids are hydrophobic, therefore their emission spectra are shifted into the direction of longer wavelengths as they get into more hydrophilic environment. As a consequence of temperature increment, water molecules are going to get nearer to these residues in parallel with conformational instability and partial unfolding. Urea is a denaturant, that promotes protein unfolding. The measurement was performed on a TECAN Spark 20M instrument, into which a UV-STAR 96-well plate was placed. Measurement parameters were calculated adapted from Bucher et al., 2021. Samples were prepared in 0,5 mg/ml concentration in 50 μ l from each construct in low salt phosphate buffer with three different urea concentrations: 0M, 1M and 2M urea. Excitation was done at 295 nm, and emission intensities were detected between 310 nm and 400 nm in 10 nm increments. Fluorescence emission shift was visualized by calculating 350 nm/320 nm emission ratios. Due to the capability of the machine, temperature dependency was measured between 18°C and 42°C.

4. RESULTS

4.1 Resonance assignment and structure calculation

Resonance assignment was done as it was discussed in Methods section. First, the spectra of the double labelled, TEV cleaved C-N-H118 sample and the exact protein sequence of these 118 amino acids (without the first methionine, that is lost due to bacterial protein expression) were loaded into CARA. Then a peak list was defined in SynchroScope, onto which different spectra were projected in order to identify the carbon of the carbonyl group (HNCO), the C α of previous (HN(CO)CA) and current (HNCA) residues and the C β of previous (HN(CO)CACB) and current (HNCACB) residues. Then a preliminary order of spin systems in the spectra were set, that was further supported with NOESY, HNHA and TOCSY projections based on spatial vicinity and side chain hydrogen shifts. The correctness

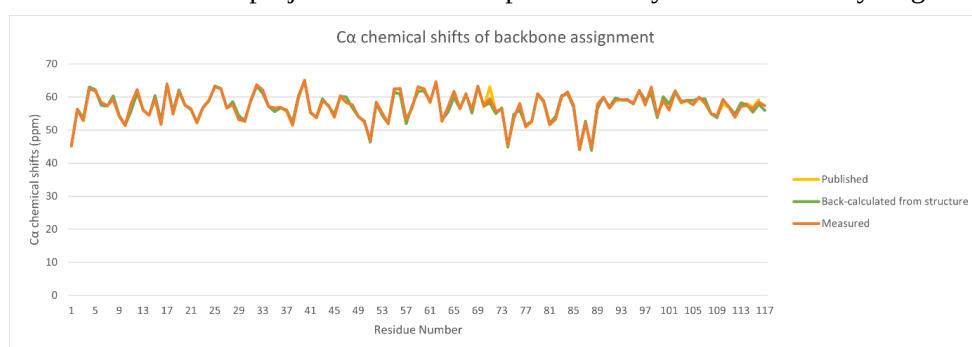


Figure 13. C β chemical shifts correlate to published data

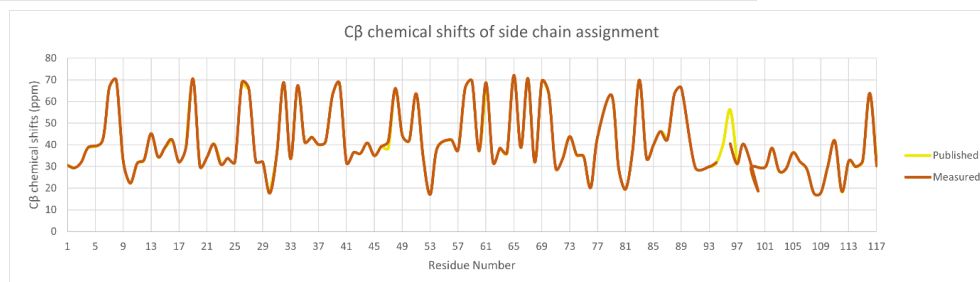


Figure 12. C α chemical shifts in agreement between published, measured, and structural data

of this order could be visualised in StripScope. Fortunately, the signal dispersion was of so high quality, that apart from a few residues (involving prolines) the whole backbone could be assigned. This was further improved, when spectra, including ^{13}C HSQC and HACAN became also available from additional measurements. These chemical shifts were corrected for reference and then C α chemical shifts were compared to the preliminary research on this domain from 2000 (Tanaka et al., 2000) and also to the estimated shifts calculated based on our preliminary solution NMR structure. As a result, we could see high correspondence between the chemical shifts (Figure 12).

Side chain assignment was started in CARA and was completed in CCPNMR. Based on protein structure calculations, it seems that the backbone shifts (C α values) were determined correctly (Figure 12), as well as the side chains C β values (Figure 13), however in case of residues that are harder to be assigned (described in Methods section), continuous fine-tuning is needed.

	Cα		Cβ			Cα		Cβ	
	Measured	Published	Measured	Published		Measured	Published	Measured	Published
2 Gly					60 Thr				
3 Glu					61 Ile				
4 Gln					62 Thr				
5 Pro					63 Pro				
6 Ile					64 Asn				
7 Phe					65 Met				
8 Ser					66 Thr				
9 Thr					67 Phe				
10 Arg					68 Thr				
11 Ala					69 Lys				
12 His					70 Thr				
13 Val					71 Ser				
14 Phe					72 Gln				
15 Gln					73 Lys				
16 Ile					74 Phe				
17 Asp					75 Gly				
18 Pro					76 Gln				
19 Asn					77 Trp				
20 Thr					78 Ala				
21 Lys					79 Asp				
22 Lys					80 Ser				
23 Asn					81 Arg				
24 Trp					82 Ala				
25 Val					83 Asn				
26 Pro					84 Thr				
27 Thr					85 Val				
28 Ser					86 Tyr				
29 Lys					87 Gly				
30 His					88 Leu				
31 Ala					89 Gly				
32 Val					90 Phe				
33 Thr					91 Ser				
34 Val					92 Ser				
35 Ser					93 Glu				
36 Tyr					94 His				
37 Phe					95 His				
38 Tyr					96 Leu				
39 Asp					97 Ser				
40 Ser					98 Lys				
41 Thr					99 Phe				
42 Arg					100 Ala				
43 Asn					101 Glu				
44 Val					102 Lys				
45 Tyr					103 Phe				
46 Arg					104 Gln				
47 Ile					105 Glu				
48 Ile					106 Phe				
49 Ser					107 Lys				
50 Leu					108 Glu				
51 Asp					109 Ala				
52 Gly					110 Ala				
53 Ser					111 Arg				
54 Lys					112 Leu				
55 Ala					113 Ala				
56 Ile					114 Lys				
57 Ile					115 Glu				
58 Asn					116 Lys				
59 Ser					117 Ser				
					118 Gln				

Figure 14. Comparison of the assignment to published results based on Cα and Cβ chemical shifts. Green: successful assignment from our side, orange: successful assignment in publication, black: could not be assigned due to lack of Cβ in glycines. Overlap between assigned residues are between G2 and Q118

As a result, we could assign all Cα and Cβ chemical shifts except for S97, that clearly exceeds the completeness of the previously published data set.

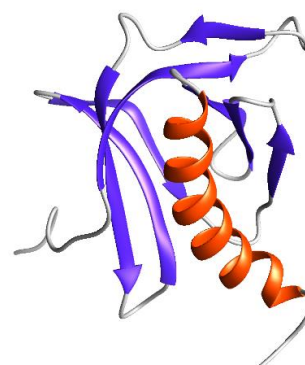


Figure 15. Preliminary structure of Homer EVH1 domain with solution NMR.

Structure calculation is still in progress, however the overall fold is very well defined (Figure 15). The problem is that not all restraints are satisfied and since AtnosCandid-Cyana is largely automated, we cannot be certain about the exact origin of the restraint violations observed.

4.2 Analysis of protein dynamics

T₁ and T₂ relaxation data were evaluated in NMRFAM-SPARKY, to which exponential curves were fitted. By default, these curves should decline exponentially, however, we could observe this decay from the third timepoint in case of both ¹⁵N and T₁ relaxation measurements.

The program TENSOR2 needed an input data file – that was converted from the output files of T₁ and T₂ relaxation peak heights and NOEs – and an input PDB file – which is usually the best-scored output PDB file from MODELLER (used to obtain models corresponding to our exact construct boundaries), but at this stage our preliminary structure on the EVH1 domain could already be used. For model fitting, residues with outlier τ_e or R_{ex} values were excluded. Isotropic and anisotropic models were fitted, which both gave the same correlation time (τ_c) of 7,8 ns, and a nice plot

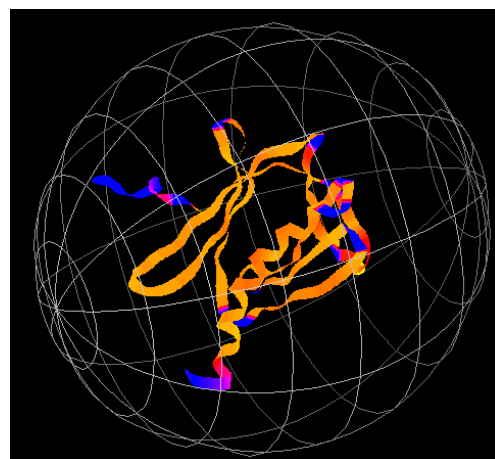
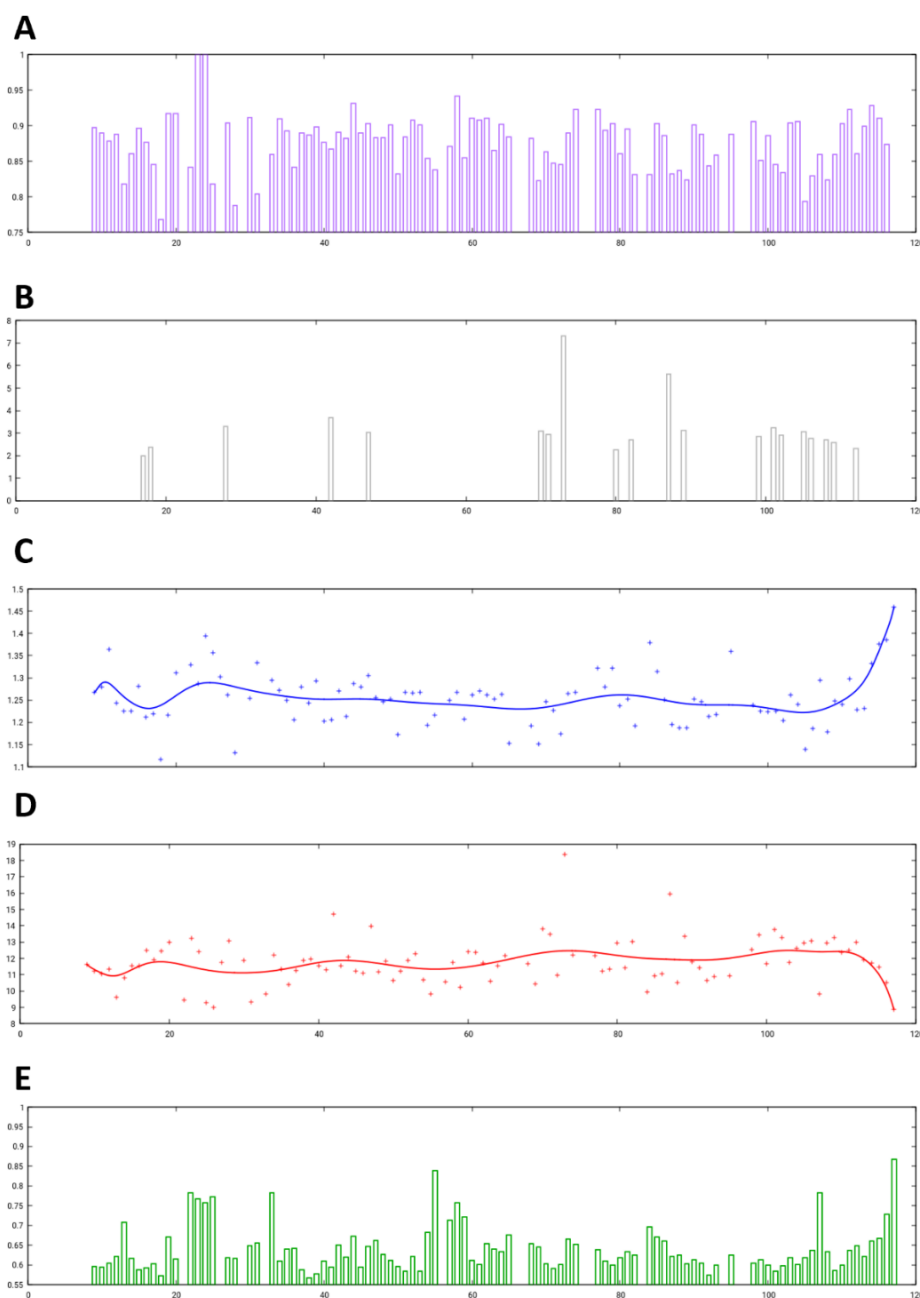


Figure 16. Tensor2 model fitted for the EVH1 domain. Coloured by S² values from blue (0) to yellow (1)

for the diffusion tensor components (Figure 16). I note that the correlation time estimated with the program HYDROPRO based on the structure only is 10,7 ns, a value considerably higher than the value obtained from the relaxation data. This discrepancy is due to the fact, that terminal residues were excluded from the Tensor2 fit, based on R_{ex} and τ_e values. If these residues were also excluded from the calculation of HYDROPRO, than the estimated correlation time came out to be 8,76 ns, therefore the difference between the Tensor2 and HYDROPRO outcome is within 1 ns, that is acceptable. Calculation of internal mobility values resulted in high ($>0,85$) S^2 values for most residues, consistent with a well-defined, globular structure. R_{ex} values were observed mainly for those residues that we already excluded based on high T_1/T_2 values.

Figure 17. Evaluation of protein dynamics: A: S^2 , B: R_{ex} , C: R_1 , D: R_2 , E: NOE



4.3 In vitro mutagenesis of the EVH1 domain

In my previous studies I have produced the wild-type Homer EVH1 domain for NMR measurement. This time my aim was to construct the mutant variants of this protein that have been described in connection with autism (Kelleher et al., 2012). Protocols were used as described in methods section.

Primer pairs were designed separately for the different mutants. In case of both pairs, one of the two primers had to contain the desired mutation in its centre or closer to its 5' end. Primers were evaluated based on AlleleID scoring.

H118- M65I_F	ATCACACCAAACAT T ACATTACTAA
H118- M65I_R	GGTGCTATTTATTATTGCCTTTGAG
H118- S97L_F	AGCATCATCTTT T AAAATTTCGCAGA
H118- S97L_R	CAGAGGAGAATCCCAGTCCATAAAC

Table 2. Primer sequences, mutation sites are labelled with red

After PCR and purification, PCR products were checked with agarose gelelectrophoresis.

Phosphorylation and ligation were successful based on transformation results. Higher number of colonies were observed on sample plates, than on negative controls.

Mini liquid cultures were prepared from some colonies separately, that was followed by plasmid isolation. Absorbance of plasmid samples were highest at 260 nm, that is characteristic for DNA molecules, which means, that plasmid isolation was successful (data not presented).

Plasmids were sent to Sanger sequencing in order to be sure about the appropriacy of their sequences. Based on the results, they have the exact same sequence, that we wanted, therefore mutant proteins can be produced from them.

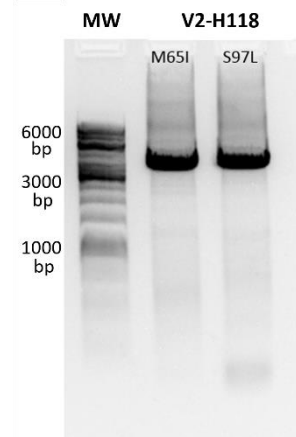


Figure 18. PCR products of mutant plasmids before PNK treatment and ligation. MW: GeneRuler

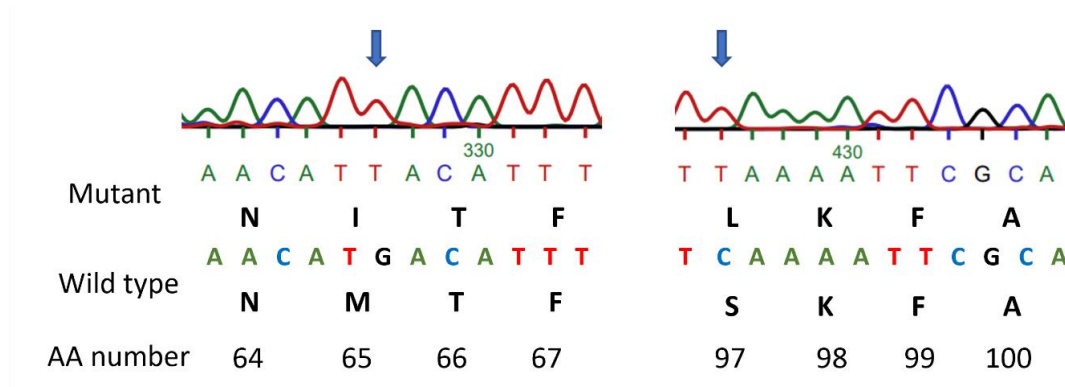


Figure 19. Sequencing results. Places of point mutations are indicated with blue arrows

Protein production had to be tried out in unlabelled form, for which protocols mentioned in Methods section were used. The critical point of protein production was transformation into BL21 cells. These competent cells are generated by a well-regulated protocol, and not always all perform well. Interestingly, only low number of cells took up mutant plasmids, but they were enough for protein production. Unlabelled proteins were expressed in LB media, then induced and centrifugated. Thereafter, cells were broken up with ultrasound homogenisation and proteins were chromatographically purified. Figure 20 shows the SDS-PAGE run of elution fractions of the unlabelled wild-type and mutant Homer118. IMAC elution fractions containing the proteins with 6xHis tags (lanes 2, 5, and 8), their weight is about 16,6 kDa. That samples are followed by the TEV cleaved ones respectively, which were at the size of 14 kDa, that corresponds to the weight of the H118 with removed His tag. IEC elution fractions are at the same size, but have better purity (lanes 4, 7 and 10). All in all, we could conclude that all proteins were in the appropriate size and purity based on the PAGE run, therefore the protocols seemed to be work, and we could move on to ¹⁵N-labelled samples.

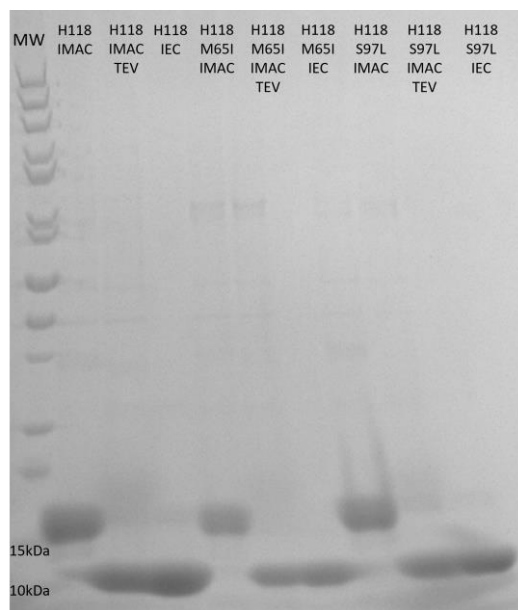


Figure 20. PAGE picture of unlabelled samples. MW: PAGERuler

Chromatographic purification steps were accomplished as discussed in Methods section, in the same way in case of unlabelled and isotope labelled proteins. The results were nearly the same, except for SEC – that was not performed on unlabelled samples, due to the lack of NMR. Here I discuss results on ¹⁵N-labelled ones. On the chromatogram of immobilized metal affinity chromatography (Figure 21A) the absorbance of nucleic acid at 255 nm is detectable with dark blue, while of 280 nm with dark green. First flow through (FT) comes down. The absorbance of this fraction is higher at 255 nm, than at 280 nm. Since the absorbance maximum of nucleic acids is at 260 nm, while that of proteins is at 280 nm, the high density at 255 nm of the FT fraction suggests, that it contains more nucleic acids than proteins. Then, by the course of washing (WASH) with 50 mM Imidazole, the nucleic acid absorbance is still higher, but, when eluting with 250 mM Imidazole, protein absorbances were about 3200 mAU at 280 nm in case of all constructs, that means significantly higher protein content. Elution fractions came down in 13 ml volume and contained ca. 26 mg protein per construct.

IMAC was followed by the Desalting step in order to reduce salt concentration that would prevent binding of the protein to the IEC resin. Only 10 ml sample could be loaded on the column from each. Elution fractions came down in 12 ml volume, that could be injected to IEC.

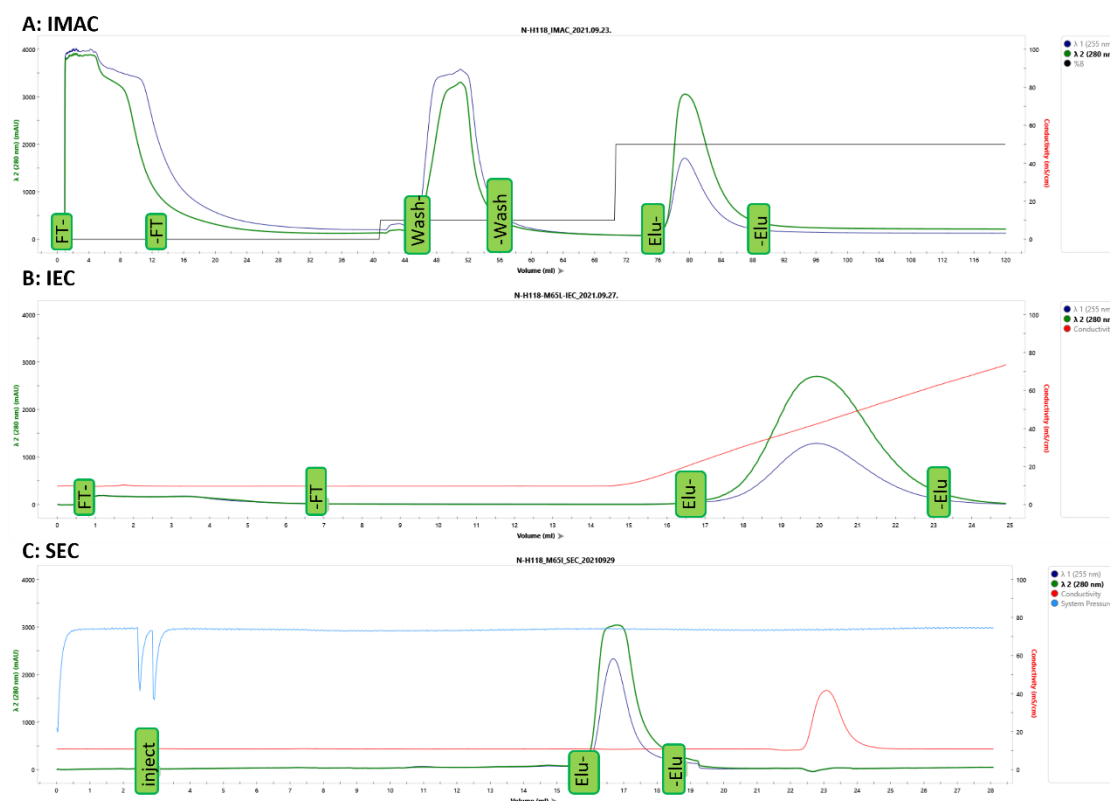


Figure 21. Chromatograms of M65I mutant's purification steps: A: IMAC, B: IEC: C: SEC

At the beginning of IEC, proteins were at low salt concentration buffers due to the Desalting step. Chromatogram of IEC is illustrated on Figure 21B, where absorbance curves of 255 nm and 280 nm are represented with dark blue and dark green respectively. Conductivity is labelled with red, that corresponds to salt content – conductivity is increasing with the increment of salt during the gradient elution. Elution fractions peaked at about 3000 mAU and came down in 6 ml at 300 mM NaCl.

By the course of SEC, samples were injected in 1ml per each. Elution fractions came down 13ml after injection in 3 ml, with 3000mAU at 280nm. They were concentrated for NMR measurement into 1ml. Detailed parameters of NMR samples are described in Table 3.

	concentration	molar concentration	molecular weight
N-H118 (WT)	4,5mg/ml	323,6 μ M	13906,1260Da
N-H118-M65I	2,89mg/ml	206,65 μ M	13888,0938Da
N-H118-S97L	3,18mg/ml	228,25 μ M	13932,2080Da

Table 3. Parameters of 15 N-labelled samples used for NMR measurements

4.4 Examination of known substitutions of the EVH1 domain

Mutant proteins were sent to NMR measurements, where HSQC and HNHA spectra were recorded along with T_1 and T_2 relaxation and heteronuclear NOE. First HSQC spectra were analysed, since that is the “fingerprint” of a protein. M65I showed higher differences from the wild-type, affecting almost the entire domain considerably, whereas amide shifts of S97L were largely more or less in accordance with those of the native molecule.

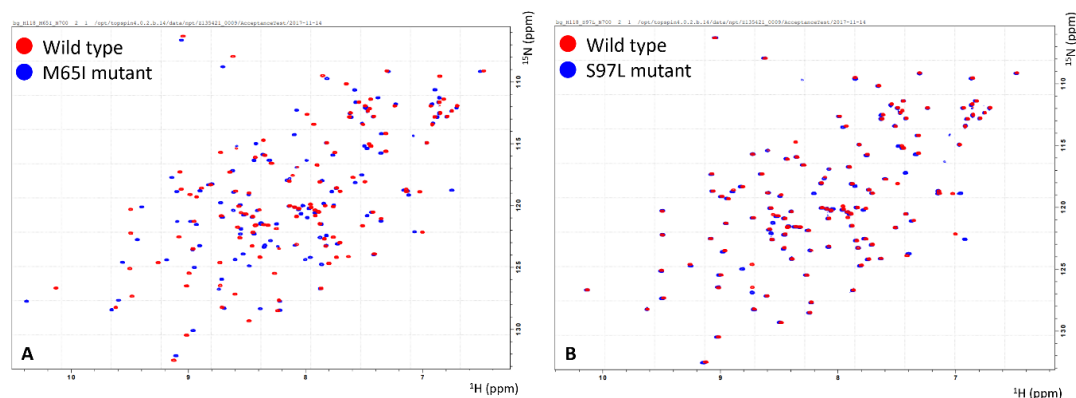


Figure 22. HSQC spectra of mutant and wild-type samples. A: M65I, B: S97L

A preliminary assignment was done for the mutant variants based on the known NH shifts of the wild-type. In case of M65I, this assignment has to be confirmed with NMR measurement and assignment of a double labelled sample, but it is certain that spin systems corresponding to the residues in the centre of the amino acid sequence (close to the mutation site) are displaced. By S97L, the main alterations were, that spin systems of residues L97, K98 and Q101 are missing from the HSQC. Since this spectrum has good quality signal dispersion and these residues are the involved and neighbouring ones to the mutation, it is probable, that the alterations are due to the mutation.

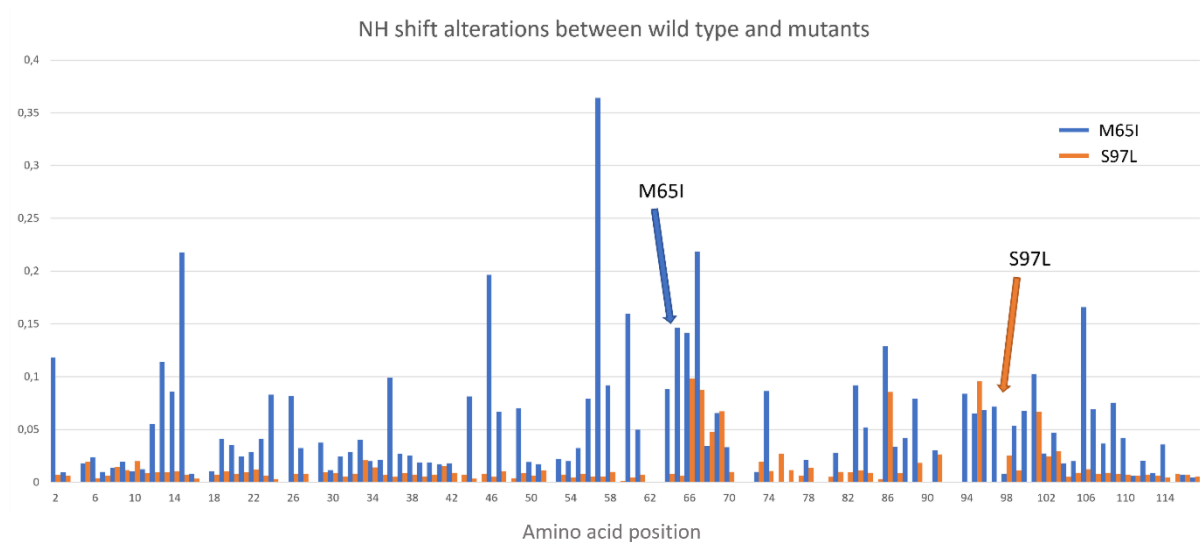


Figure 23. NH shift alterations from wild-type based on preliminary assignment

For further investigation of the effects of point mutations, circular dichroism (CD) spectroscopy and tryptophan fluorescence were measured with temperature dependence on native and mutant samples. Based on literature data the secondary structure of the wild-type EVH1 domain β -strands are

dominating, but on the carboxy-terminal an α -helix is present. The mutation site of M65I takes place between two β -strands, while of S97L is in the helix (Figure 24). Temperature dependence of secondary structural elements were also tested. As a result, we could observe, that CD curves of wild-type and S97L samples were similar at all temperatures, M65I showed similarity at lower temperatures, but the alterations increased with temperature. The presence of the α -helix can be clearly read out from the curves based on the long negative band in the region of 210-230nm and also the positive and negative peaks of β -strands are observable at 195 nm and 216 nm respectively, therefore we could conclude, that all

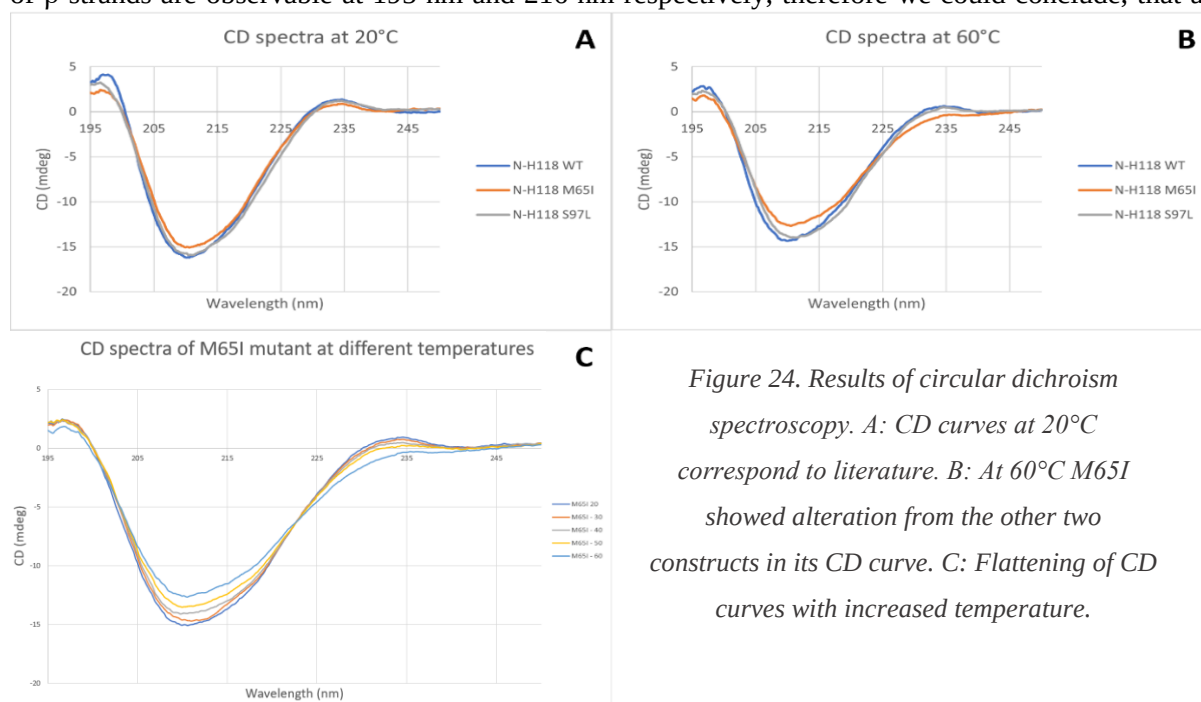


Figure 24. Results of circular dichroism spectroscopy. A: CD curves at 20°C correspond to literature. B: At 60°C M65I showed alteration from the other two constructs in its CD curve. C: Flattening of CD curves with increased temperature.

samples possess the cited secondary structure. At higher temperatures the curves got “flatter”, which means the increment of β -strand content, based on BESTSEL¹⁰. It is notable, that this deviation is the most significant in case of M65I, differences from the other two are clearly visible at 60°C.

Tryptophan fluorescence was measured at different temperatures and urea concentrations. The Homer EVH1 domain contains two tryptophane (W) residues in its sequence. As temperature increased, fluorescence spectra showed peak intensity reduction at all urea concentrations for all constructs. Absorbances usually had a maximum at 330 nm, but in case of M65I a significant alteration could be observed in this measurement as well, namely, that by 2M urea concentration peak absorbances shifted into the direction of longer wavelength elevating the temperature (Figure 26). This outcome somehow consistent with the temperature sensitivity observed by CD spectroscopy as well. By the second experiment, where

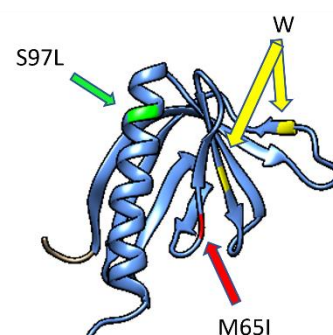


Figure 25. Tryptophans(W) and mutation sites on the EVH1

¹⁰ <https://bestsel.elte.hu/index.php> Date of download: 23. 10. 2021

samples were incubated at each temperature for a longer time, red-shift of the emission spectra of the wild-type sample was observed as well. For the quantification of this outcome, ratios of absorbances at 350 nm and 320 nm were calculated, where significant increase can be observed at 42°C for the M65I mutant and a smaller increment for the wild-type. But it has to be mentioned, that fluorescent intensities were significantly higher in the M65I emission spectra. After temperature elevation, reversibility of the unfolding process was tested by an additional measurement at 18°C. Based on the results, we could conclude that the observed changes are irreversible, that can also refer to an aggregation induced by the heat effect.

To sum up, it seems by CD spectroscopy that the secondary structure of the Homer EVH1 domain is not so sensitive to mutations and less dependent on temperature, while its tertiary structure shows sensitivity to mutations based on tryptophan fluorescence results.

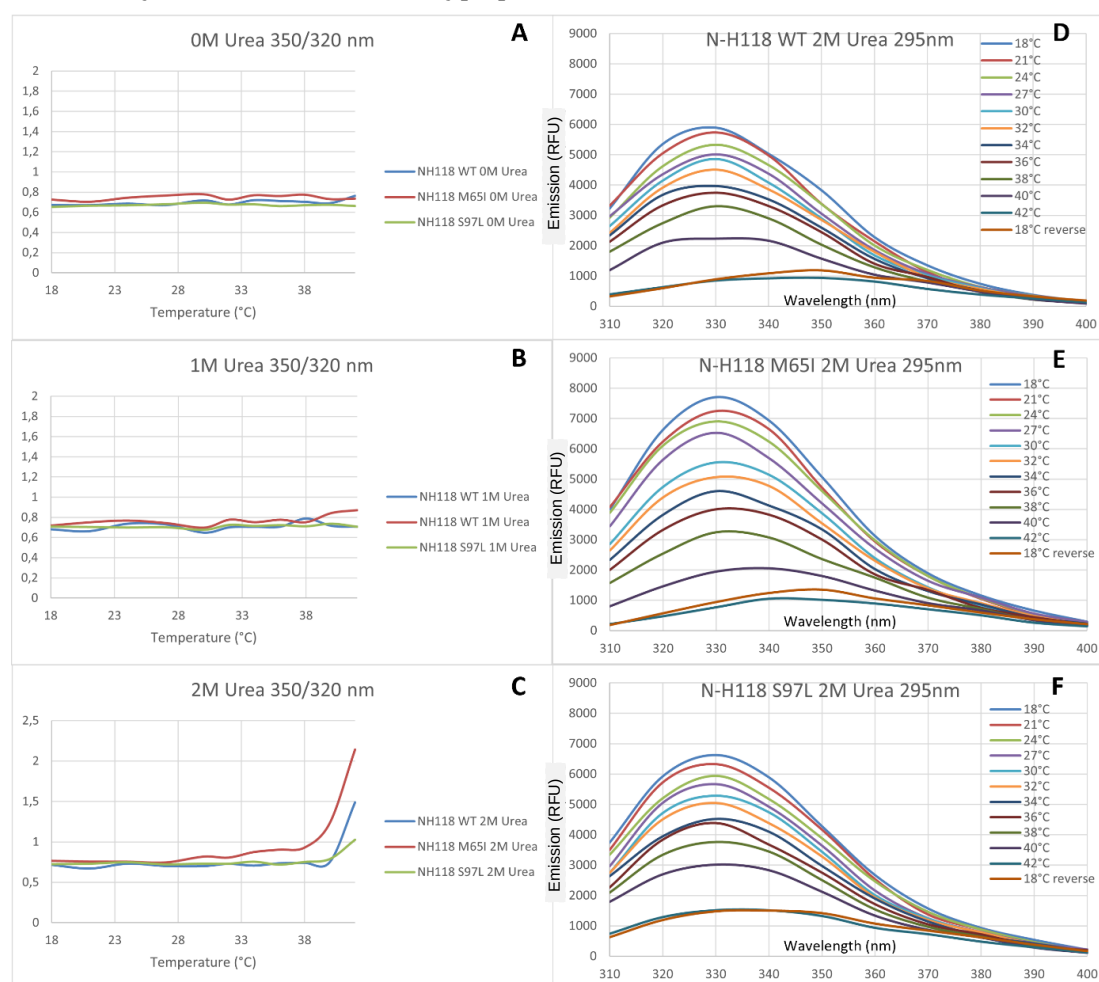


Figure 26. Results of tryptophan fluorescence spectroscopy: A-C: 350 nm/320 nm fluorescence ratios, significant increment by M65I at 2M urea concentration. D-F: temperature dependence of fluorescence emission at 2M urea concentration for each construct.

5. SUMMARY

5.1 Discussion

The protein assignment of the wild-type, TEV-cleaved Homer EVH1 protein has been successfully performed, that seems to be improved relative to the published one from 2000 (Tanaka et al., 2000). From the obtained NMR data, the internal dynamics was inferred, and a well-defined structural fold could be established. Although both calculations remain to be fine-tuned, based on the results we can safely conclude that Homer EVH1 is a stable, globular protein with a well-defined backbone.

The *in vitro mutagenesis* of the EVH1 domain was successful, hence ASD related mutant proteins could be expressed and isolated. Protein production of all three constructs (wild-type and mutants) was done parallel, therefore the effects of the substitutions could be examined well. NMR spectra of ¹⁵N-labelled M65I mutant showed significant alterations from the wild-type, while of S97L, only the spectral environment of the mutation site has been changed. Based on circular dichroism results, secondary structure of the domain is not seriously affected by the mutations. On the other hand, tertiary structure is influenced by mutations based on tryptophan fluorescence (especially that of the M65I). In case of the latter experiment the possible temperature range was limited by our instrument (18°C-42°C) and did not allow us to examine the structural stability of the proteins at higher temperatures, although this could provide us with additional important information.

5.2 Future plans

Our major aim for the future is to finalise the solution NMR-based protein structure of the EVH1 domain. The fine-tuning of protein dynamics is also needed, after which the internal dynamics of mutant proteins could also be calculated. Based on the results, we could observe significant alterations of M65I mutant from wild-type, but until now we can only guess about the underlying reasons. For detailed investigation additional measurements, and probably a double labelled sample will be needed. Additionally, interaction measurements are planned with the proline-rich region (PRR) of Shank and the Homer binding motif (HBM) of Debrin, for which the constructs are in progress.

5.3 Acknowledgements

First of all, I would like to thank my supervisors, Dr. Zoltán Gáspári and Dr. Bálint Ferenc Péterfia the possibility of working in this research group. In the last three years, I always found myself lucky to have a protein to work on like the Homer EVH1 domain and to be part of such an amazing group. I am extremely grateful to Dr. Zoltán Gáspári for all of his advises in protein assignment and for the explanation on protein dynamics; and extremely grateful to Dr. Bálint Ferenc Péterfia for all the help and guidance in the laboratory, especially for his help provided in *in vitro mutagenesis* during the pandemic crisis.

I would like to thank to Dr. András Czajlik his work in side chain assignment and structure calculation and all the instructions he gave me regarding these topics. I believe, his expertise is outstanding, and I am happy that I could learn from him.

I am also grateful to Professor Dr. Gyula Batta from the University of Debrecen for the NMR measurements.

I am thankful to Dr. Viktor Farkas from the Separation Science Research and Education Laboratory of Eötvös Lóránd University, for the guidance he gave in circular dichroism and for István Pap for mass spectrometry measurements. I would also like to thank Enora Mountin the Homer1 mouse cDNA.

Finally, I am extremely grateful for the PhD students of our laboratory. Special thanks to Fanni Farkas, for the help she gave me in the parallel isolation of the three Homer proteins, Zsófia Kálmán for the bioinformatics advises and Anna Sánta for the help in CCPNMR. I would like to thank to some old colleagues as well, to József Hegedüs for the introduction to CARA and to István Pap and Csilla Hommonay for the protein construct and protocol optimisations.

Last but not least, I would like to thank the financial support obtained for my research.

„Supported by the ÚNKP-21-2 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.”



REFERENCES

- Ball, L. J., Jarchau, T., Oschkinat, H., & Walter, U. (2002). EVH1 domains: structure, function and interactions. *FEBS letters*, 513(1), 45-52.
- Beneken, J., Tu, J. C., Xiao, B., Nuriya, M., Yuan, J. P., Worley, P. F., & Leahy, D. J. (2000): Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition. *Neuron*, 26(1), 143-154.
- Banerjee, A., Luong, J. A., Ho, A., Saib, A. O., & Ploski, J. E. (2016). Overexpression of Homer1a in the basal and lateral amygdala impairs fear conditioning and induces an autism-like social impairment. *Molecular autism*, 7(1), 1-15.
- Bax, A., & Clore, G. M. (2019): Protein NMR: Boundless opportunities. *Journal of Magnetic Resonance*, 306, 187-191.
- Blanpied, T. A., Kerr, J. M., & Ehlers, M. D. (2008): Structural plasticity with preserved topology in the postsynaptic protein network. *Proceedings of the National Academy of Sciences*, 105(34), 12587-12592.
- Bucher, M., Niebling, S., Han, Y., Molodenskiy, D., Nia, F. H., Kreienkamp, H. J., ... & Mikhaylova, M. (2021). Autism-associated SHANK3 missense point mutations impact conformational fluctuations and protein turnover at synapses. *Elife*, 10, e66165.
- Dawson, N. L., Orengo, C., & Gáspári, Z. (2020). Assessing Protein Function Through Structural Similarities with CATH. In *Structural Bioinformatics* (pp. 43-57). Humana, New York, NY.
- Denkena, J., Zaisser, A., Merz, B., Klinger, B., Kuhl, D., Blüthgen, N., & Hermey, G. (2020). Neuronal activity regulates alternative exon usage. *Molecular brain*, 13(1), 1-24.
- Dosemeci, A., Weinberg, R. J., Reese, T. S., & Tao-Cheng, J. H. (2016). The postsynaptic density: there is more than meets the eye. *Frontiers in synaptic neuroscience*, 8, 23.

- Drahos L., Schlosser G., Vékey K. (2018): Tömegspektrometria. In: Dr. Buday László, Dr. Nyitray László, Dr. Perczel András (Eds.) (2018): „Ezerarcú fehérjék”. Budapest, Semmelweis Kiadó és Multimédia Stúdió (2018) pp.: 335-356 (Buday et al., 2018, B)
- Farkas V., Jáklí I., Kardos J., Vass E. (2018): Elektron- és rezgési spektroszkópia a fehérjekutatásban. In: Dr. Buday László, Dr. Nyitray László, Dr. Perczel András (Eds.) (2018): „Ezerarcú fehérjék”. Budapest, Semmelweis Kiadó és Multimédia Stúdió (2018) pp.: 277-300 (Buday et al., 2018, A)
- Feng, W., & Zhang, M. (2009). Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nature Reviews Neuroscience*, 10(2), 87-99.
- Ferella L., Rosato A. and Turano P. (2012): Determination of Protein Structure and Dynamics In: Bertini, I., McGreevy, K. S., & Parigi, G. (Eds.). (2012). *NMR of biomolecules: towards mechanistic systems biology*. John Wiley & Sons. (2012) pp: 53-173 - B
- Gomari, M. M., Saraygord-Afshari, N., Farsimadan, M., Rostami, N., Aghamiri, S., & Farajollahi, M. M. (2020). Opportunities and challenges of the tag-assisted protein purification techniques: Applications in the pharmaceutical industry. *Biotechnology Advances*, 107653.
- Guang, S., Pang, N., Deng, X., Yang, L., He, F., Wu, L., ... & Peng, J. (2018). Synaptopathology involved in autism spectrum disorder. *Frontiers in cellular neuroscience*, 12, 470.
- Hayashi, M. K., Tang, C., Verpelli, C., Narayanan, R., Stearns, M. H., Xu, R. M., Li H., Sala C. & Hayashi, Y. (2009): The postsynaptic density proteins Homer and Shank form a polymeric network structure. *Cell*, 137(1), 159-171.
- Hayashi, Y. (2019). Drebrin-Homer Interaction at An Atomic Scale. *Structure*, 27(1), 3-5.
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1-8.
- Hermey, G., Bluethgen, N., & Kuhl, D. (2017): Neuronal activity-regulated alternative mRNA splicing. *The international journal of biochemistry & cell biology*, 91, 184-193.
- Kaizuka, T., & Takumi, T. (2018). Postsynaptic density proteins and their involvement in neurodevelopmental disorders. *The Journal of Biochemistry*, 163(6), 447-455.
- Kelleher III, R. J., Geigenmüller, U., Hovhannisyan, H., Trautman, E., Pinard, R., Rathmell, B., ... & Margulies, D. (2012). High-throughput sequencing of mGluR signaling pathway genes reveals enrichment of rare variants in autism. *PloS one*, 7(4), e35003.
- Keller, Rochus L. J. (2015): *The Computer Aided Resonance Assignment Tutorial*, Institute for Molecular Biology and Biophysics, The Swiss Federal Institute of Technology, Zürich
- Lee, W., Tonelli, M., & Markley, J. L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, 31(8), 1325-1327.
- Li, Z., Liu, H., Li, J., Yang, Q., Feng, Z., Li, Y., ... & Zhang, M. (2019). Homer tetramer promotes actin bundling activity of drebrin. *Structure*, 27(1), 27-38.
- Luo, P., Li, X., Fei, Z., & Poon, W. (2012): Scaffold protein Homer 1: implications for neurological diseases. *Neurochemistry international*, 61(5), 731-738.
- Madl T. & Sattle M. (2012): NMR Methodologies for the Analysis of Protein-Protein Interactions In: Bertini, I., McGreevy, K. S., & Parigi, G. (Eds.). (2012). *NMR of biomolecules: towards mechanistic systems biology*. John Wiley & Sons. pp: 173-194 - A
- O'Connor, E. C., Bariselli, S., & Bellone, C. (2014). Synaptic basis of social dysfunction: a focus on postsynaptic proteins linking group-I m G l u R s with AMPAR s and NMDAR s. *European Journal of Neuroscience*, 39(7), 1114-1129.
- Ortega, A., Amorós, D., & De La Torre, J. G. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic-and residue-level models. *Biophysical journal*, 101(4), 892-898.
- Shiraishi-Yamaguchi, Y., & Furuichi, T. (2007). The Homer family proteins. *Genome biology*, 8(2), 1-12.
- Szabad János, Lippai Mónika (2016): *Az idegsejtek és az érzékelés molekuláris biológiája*. Pázmány Péter Catholic University, Faculty of Informationtechnology and Bionics, Molecular Biology and Genetics of the Cell; course material from the University's website

- Tanaka, T., Sugai, M., Ito, Y., Inokuchi, K., & Kohno, T. (2000): Letter to the Editor: Assignments of the ^1H , ^{13}C , and ^{15}N resonances of the 21 kDa Vesl/Homer family protein, Vesl-1S. *Journal of biomolecular NMR*, 18(2), 181-182.
- Tu, J. C., Xiao, B., Naisbitt, S., Yuan, J. P., Petralia, R. S., Brakeman, P., Doan A., Aakalu V.K., Lanahan A.a., Sheng M. & Worley, P. F. (1999): Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins. *Neuron*, 23(3), 58
- Wang, F., Zhang, Y., Wang, L., Sun, P., Luo, X., Ishigaki, Y., Sugai T., Yamamoto R. & Kato, N. (2015): Improvement of spatial learning by facilitating large-conductance calciumactivated potassium channel with transcranial magnetic stimulation in Alzheimer's disease model mice. *Neuropharmacology*, 97, 210-219
- Wang, W., Weng, J., Zhang, X., Liu, M., & Zhang, M. (2009). Creating conformational entropy by increasing interdomain mobility in ligand binding regulation: a revisit to N-terminal tandem PDZ domains of PSD-95. *Journal of the American Chemical Society*, 131(2), 787-796.
- Zeng, M., Chen, X., Guan, D., Xu, J., Wu, H., Tong, P., & Zhang, M. (2018): Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell*, 174(5), 1172-1187.

References quoted as footnote

- Bio-Rad, 2021: <https://www.bio-rad.com/en-hu/applications-technologies/affinity-chromatography?ID=LUSMJIDN>, Date of download: 10. 10. 2021
- CCPNMR, 2021: <https://www.ccpn.ac.uk/>, Date of download: 11. 11. 2021
- ELTE BESTSEL, 2021: <https://bestsel.elte.hu/index.php> Date of download: 23. 10. 2021
- Higman, V. A., 2012: Protein NMR: <https://www.protein-nmr.org.uk/solution-nmr/spectrum-descriptions/> Date of download: 02. 11. 2021
- InterPro, 2021: <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011993/>, Date of download 16. 09. 2021, <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000697/>, Date of download 16. 09. 2021
- Modeller, 2021: <https://salilab.org/modeller/> Date of download: 25. 10. 2021
- NCBI GENE, 2021: <https://www.ncbi.nlm.nih.gov/gene/9456>, Date of download: 13. 09. 2021, <https://www.ncbi.nlm.nih.gov/gene/9455>, Date of download: 13. 09. 2021, <https://www.ncbi.nlm.nih.gov/gene/9454>, Date of download: 13. 09. 2021
- NMRFAM-Sparky, 2008: <https://www.cgl.ucsf.edu/home/sparky/manual/index.html>, Date of download: 11. 11. 2021
- Protein Calculator: <http://protecalc.sourceforge.net/>, Date of download: 01. 12. 2021