ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY

INSTITUTE OF SCIENCES

RATIONAL PEPTIDE DESIGN FOR TARGETING HOMODIMERIZATION OF BACH2 BTB DOMAIN

EFE ACAR

M.Sc. THESIS

DEPARTMENT OF MOLECULAR AND TRANSLATIONAL BIOMEDICINE

SUPERVISOR

Assoc. Prof. Dr. Ahmet Can Timuçin

ISTANBUL-2024

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF SCIENCES

RATIONAL PEPTIDE DESIGN FOR TARGETING HOMODIMERIZATION OF BACH2 BTB DOMAIN

EFE ACAR

M.Sc. THESIS

DEPARTMENT OF MOLECULAR AND TRANSLATIONAL BIOMEDICINE

SUPERVISOR

Assoc. Prof. Dr. Ahmet Can Timuçin

ISTANBUL-2024

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethical rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

29/11/2024

Efe Acar

PREFACE AND ACKNOWLEDGEMENT

To start, I am grateful for the excellent supervision I received throughout my project. My special thanks go to Assoc. Prof. Dr. Ahmet Can Timuçin for providing me the necessary education on advanced computational analysis, laboratory and required equipment. The discussions and the pieces of advice I took from him were especially inspiring and valuable. Furthermore, his guidance and help throughout my project on every single step eased my understanding of the concept and improved me more than I imagined.

I also would like to thank Prof. Dr. Emel Timuçin for the laboratory she provided and valuable advice throughout bacterial culture step of the project.

My sincere thanks go Acıbadem University for providing me with the basic equipment and laboratory, which eased the project's finance dramatically.

Finally, I would like to thank my family and close friends for supporting and encouraging me. Wherever I go, they created the best familial atmosphere as best as they could.

This work was supported by TÜBİTAK (The Scientific and Technological Research Council of Turkey) [Grant No: KBAG-221Z368]. The numerical calculations reported in this work were fully performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Rational Peptide Design for Targeting Homodimerization of BACH2 BTB Domain

BACH2 is a transcriptional repressor, an antagonist of the cellular antioxidant regulator NRF2 transcription factor, which dimerizes through its BTB domain. BACH2 has also been implicated in tumor immunosuppression. BTB domain homodimerization of BACH2 is known to be necessary for its functionality. Despite its roles in NRF2-mediated antioxidant response and immunosuppression, up until now, no studies have been implemented to design peptide-based BTB domain homodimerization inhibitors, which prompted this thesis. Here, a peptide library has been computationally designed to target the homodimerization region of the BACH2 BTB domain monomer. To achieve this goal, the currently available BACH2 BTB domain structure was retrieved and subjected to sequence-based optimization. Following sequence-based optimization, peptide structures were predicted, and these structures were globally docked to the BTB domain of the BACH2 monomer. These globally docked BACH2 BTB domain-peptides structures were subjected to local redocking and simulated 500 ns in triplicates. Following MD simulations, relative binding free energy analyses indicated that one of these peptides, peptide 13 had higher relative binding free energy levels compared with positive control. Moreover, MD analyses revealed that this relative binding free energy estimation is continuous through the last 250 ns of MD simulations through RMSF, SASA, distance, hydrogen bonding, salt bridge, hydrophobic interaction, and essential dynamics analyses. In addition to these analyses, a supernatant depletion assay has been implemented to retrieve dissociation constant values for high affinity and control peptides. Overall, it has been concluded that the BACH2 BTB domain homodimerization region could be targeted via a high-affinity peptide.

Keywords: BACH2, Homodimerization, NRF2, Peptide, Affinity

ÖZET

BACH2 Proteinin BTB Bölgesinden Dimerizasyonunu Hedefleyen Rasyonel Peptit Tasarımı

BACH2, transkripsiyonel bir baskılayıcı, hücre anti-oksidasyon regülatörü olan NRF2 transkripsiyion faktörünün rakibi ve BTB domain bölgesinden dimerize olan bir proteindir. Bu protein aynı zamanda tümör imunosupresyonunda da rol almaktadır. BACH2 BTB domain homodimerizasyonu aynı zamanda da fonksiyonelite açısından da gerekli olduğu bilinmektedir. Bu proteinin NRF2 tabanlı anti-oksidan tepki elemanı ve immunosupresyonundaki rollerine rağmen, literatürde herhangi bir peptit tabanlı BACH2 BTB domain homodimerizasyonu engelleme araştırılması gerçekleştirilmemiştir. Bu tezde de bu peptit tasarımı amaçlanmaktadır. Bu tezde BACH2 BTB domain monomerinin homodimerizasyonun bölgesini hedefleyen, bilgisayar hesap temelli bir peptit kütüphanesi oluşturulmuştur. Bu amaca ulaşmak için, mevcut BACH2 BTB domain yapısı literatürden alınmış ve sekans temelli bir optimizasyonla deneyde kullanılmıştır. Bu sekans temelli optimizasyona, peptit tasarımı ile devam edilmiştir ve bu tasarlanan peptitler global olarak BACH2'nin monomerinin BTB domain bölgesine kenetlendirilmiştir. Bu global kenetlenmiş BACH2 BTB-peptit yapıları sonrasında yerel kenetlenmeye tabi tutulmuş ve 500ns'lik sürelerle kenetlenme sürekliliği 3 defa simüle edilmiştir. Bu simulasyonlar, bağıl bağlanma serbest enerji analizi hesaplamaları ile takip edilmiş, peptit 13'ün pozitif kontrole göre daha yüksek bir bağıl bağlanma serbest enerji değerijne sahip olduğu görülmüştür. Bununla birlikte, bu enerji değerinin son 250 ns'de sabit olduğu RMSF, SASA, uzaklık, hidrojen bağı, tuz köprüsü, hidrofobik etkileşimler ve gerekli dinamik analizleri ile desteklenmiştir. Bu analizlere ek olarak, supernatant tüketme analizi yardımıyla peptit 13 ile kontrol peptiti için ayrışma sabiti değerleri elde edilmiştir. Sonuç olarak, BACH2 BTB domain homodimerizasyon bölgesinin yüksek afiniteli bir peptit tarafından hedeflenebileceği gösterilmiştir.

Anahtar Sözcükler: BACH2, Homodimerizasyon, NRF2, Peptit, Afinite

1 INTRODUCTION AND AIM

1.1 NRF2 Transcription Factor Mechanism, Related Chemical Procedures

Nuclear factor erythroid 2-related factor 2 (NRF2) is a cap-n-collar transcription factor, which is coded by NFE2L2 gene (1). It has a similar sequence homology with its other orthologs; therefore, it is stated that NRF2 has 7 different "Nrf2-ECH homology domains" with various functions (1, 2). The first domain, Neh1, is located on the C-terminus of the protein which has the bZip domain (4, 5). Neh2 domain promotes KEAP1 binding, which is the NRF2 repressor (5). Neh3 plays a role in protein stability next to bZip domain and it is a transactivation domain (6). Neh4 and Neh5 contribute to histone acetyltransferase activity by binding to cAMP Response Element Binding Protein (CREB) (5). Neh6 domain works for the degradation of NRF2 by involving a redox-insensitive process if the cell is under stress (7). In unstressed conditions, this degradation is suppressed. It contains a basic leucine zipper domain (bZip) on its C-terminus and this domain is known for creating a potential heterodimer with other bZIP domains of muscle aponeurosis fibromatosis (MAF) proteins such as MAFK, MAFG, and MAFF (1, 8, 9). These MAF proteins have a bZip domain for both DNA binding and dimerization, yet do not contain a transcription activation region unlike large MAFs (MAFA, MAFB, c-MAF) (10). In addition to this, these heterodimer complexes play huge roles in the regulation of 250 genes in humans, which have Antioxidant Response Elements (ARE - 5′-TGACNNNGC-3′). The NRF2 - MAF complexes control homeostasis functions such as inflammation, redox metabolism, and proteostasis (11, 12, 13). In light of these roles, NRF2 is aimed to be a potential target in vascular, neurodegenerative, metabolic, and cancer-related pathologies, that have oxidative stress and inflammation. Theoretically, supporting homeostasis through NRF2 activation is foreseen as a therapeutic effect in many other diseases.

1.2 NRF2 Targeted Genes

In the case of NRF2 being activated, there are several genes that it targets, which encode drug-metabolizing enzymes (9). The first gene is GST, which is responsible for encoding glutathione S-transferase (14). This enzyme is a mitochondrial enzyme and plays a role in conjugating glutathione with endogenous electrophiles. When GSTs catalyze the glutathione toxic compounds are removed from the cell. Therefore, NRF2 plays a huge role in detoxification. The second gene is the NQO1 gene. This gene encodes NAD(P)H quinone oxidoreductase 1 and it catalyzes the reduction of reactive quinones (15). These quinones are reactive oxygen species and contribute to oxidative stress. GCLC and GCLM genes are also NRF2's target genes (16). They encode the Glutamate cysteine ligase catalytic subunit (GCLC) and glutamate-cysteine ligase regulatory subunit (GCLM), which form a heterodimer with each other. It also plays a role in glutathione synthesis. In addition to those, NRF2 also targets UGT1A1 and UGT1A6 genes that encode Uridine 5'-diphosphate-glucuronosyltransferase. It plays a role in the conjugation of glucuronic acid and makes it water-soluble (17). Furthermore, the ABCC2 gene is targeted by NRF2, and it encodes membrane transporters like multidrug resistance-associated proteins (MRPs). These proteins alter the toxicity of compounds and pharmacokinetics (18, 19). The final gene that is targeted is KEAP1 and it encodes Kelch-like ECHassociated protein 1 (KEAP1) (20).

The stability of NRF2 is controlled by E3 ligase adaptor "Kelch-like ECH-associated protein 1 (KEAP1)" inside the cell (1). It has 2 protein binding regions, which are BTB (bric-a-brack, tramtrack, broad-complex) in the N terminus and Kelch domain (also called double glycine repeat) in the C terminus (21). Between these 2 regions, there is also another region called the intervening region (IVR) or linker region (LR) with a high level of cystine residues (22, 23). The Kelch domain of KEAP1 interacts with the Neh2 domain of NRF2, which promotes ubiquitylation of NRF2 by Cullin3 and RBX1 E3 ligase complexes (24). This causes NRF2 degradation. In humans, KEAP1 contains 27 cystine amino acids, which acts as a redox sensor for electrophilic reactions and oxidative signals (1, 25). In situations like disturbance of intercellular redox balance, glutathione and thioredoxin work as redox tampons and

reduce reactive oxygen species along with protein glutathionylation (1). KEAP1 cystine amino acids on the other hand are modified with increased reactive oxygen species, thiol oxidation, and macromolecule glutathionylation (1, 26). These modifications disturb the functionality of KEAP1 and eventually cause NRF2 stabilization, nuclear translocation, and targeted gene expression.

In the literature, NRF2 activators are also named KEAP1 inhibitors (1). These compounds are subcategorized as; "Electrophiles", "Protein-protein interaction inhibitors" and "Multiple targeted drugs" (1). At this point, several various chemical modulators, drug candidates, and molecules are discovered and undergo different analyses. Electrophilic compounds generally aim for KEAP1 to be a redox sensor by directing its cystine amino acids to electrophilic reactions. This ends up with NRF2 molecular stabilization and activation (1, 27, 28, 29). The second approach for NRF2 activation via KEAP1 inhibition is cutting the "Cys151" amino acid in KEAP1 if there is a KEAP1- Cullin3/RBX1 E3 ligase complex. This approach prevents NRF2 ubiquitylation but activates newly produced NRF2 molecules (30), (31, 32). The most renowned drug, that is approved by FDA for relapsing-remitting multiple sclerosis, is dimethyl fumarate and it targets Cys151 in KEAP1 (33, 34), (35, 36). Furthermore, this drug is also authorized for psoriasis disease. Although targeting this residue is a successful approach, electrophilic compounds do not have enough selectivity (37). Targeting the protein-protein interaction surface of the NRF2/KEAP1 complex resulted with second-class KEAP1 inhibitors. At this point, NRF2 amino acid residues on KEAP1 with the highest interaction role are targeted with NRF2 peptides (39, 40, 41). The peptide contains HIV Tat sequence and calpain cleavage site sequence for cellular penetration, which is used on mouse models with cerebral ischemia and assured neurological protection and cognitive function protection (42). In the literature, the most familiar BACH-based NRF2 is applied for BACH1. It has a very similar structure compared to BACH2 in terms of creating heterodimers by interacting with small MAF proteins. This complex binds NRF2-targeted genes through ARE sites and acts as NRF2 nuclear suppressor. This work also shows that a compound called "HPP-4382" binds to the bZip site of BACH1 and prevents its activation, along with resulting NRF2 to reach its target genes (43). All these works point out the importance

of a peptide design targeting the BACH2 BTB domain to prevent homodimerization of it.

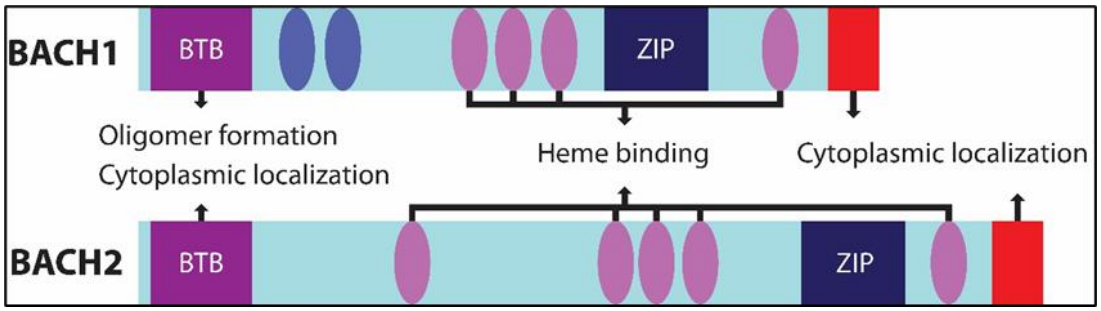

Figure 1: Identified regions of BACH1 (above) and BACH2(below). BTB domain BTB domain (Broad-Complex, Tramtrack and Bric a brac) and CNC (Cap`n Collar) homology 2), bZip domain (Basic leucine zipper)

1.3 Targeting BACH2 and Its Potential Role in NRF2 Regulation

BACH transcription factor family members are BACH1 and BACH2. They can interact with small MAF proteins and create heterodimers, which cause transcriptional repression. Both BACH1 and BACH2 interact with MAF proteins in the nucleus and suppress nucleic NRF2 (44, 45, 46, 47, 48, 49). BACH1 can be found in all human cells and is highly expressed especially in hematopoietic cells (50), yet BACH2 expression is only limited to B cells, T cells, alveolar macrophages and neural cells (51). This differentiates BACH2 as a more selective target than BACH1 and makes it more approachable as a cellular selective molecule. Besides, BACH1 and BACH2 consist of 3 main regions, which are the BTB domain, bZip domain, and intrinsically disordered region (52, 53, 54). Their interaction with MAF and heterodimer creation occurs through BTB domain. BACH1's BTB domain is known as essential for cytoplasmic–nuclear translocation (55). This result is obtained from a project, in which the BTB domain of BACH1 is truncated and an isoform is created. This truncated isoform interacts with the full-length BACH1 BTB domain and causes BACH1 cytoplasmic–nuclear translocation (55). It is also suggested a similar homodimerization mechanism for BACH2 cytoplasmic–nuclear translocation. In the same project, Leu24Pro mutation on the BTB domain of BACH2 is done, yet it is

unable to interact with wild-type BACH2 (55). This is discussed as the mutation might disturb the hydrophobic dimerization surface and cause false folding. Additionally, it is stated that there is an isoform of BACH2 in which the BTB domain is truncated (56). To sum up, if the BTB domain homodimerization occurs, homodimer BAH2 can translocate from cytoplasm to nucleus and cause BACH2-MAF interaction. This inactivates NRF2 targeted genes. A peptide-based BTB domain homodimerization targeting can provide the protein to stay monomeric and therefore it cannot translocate into the nucleus. At this point, it cannot compete with the NRF2 binding site and the NRF2-MAF complex can continue activating NRF2-targeted genes.

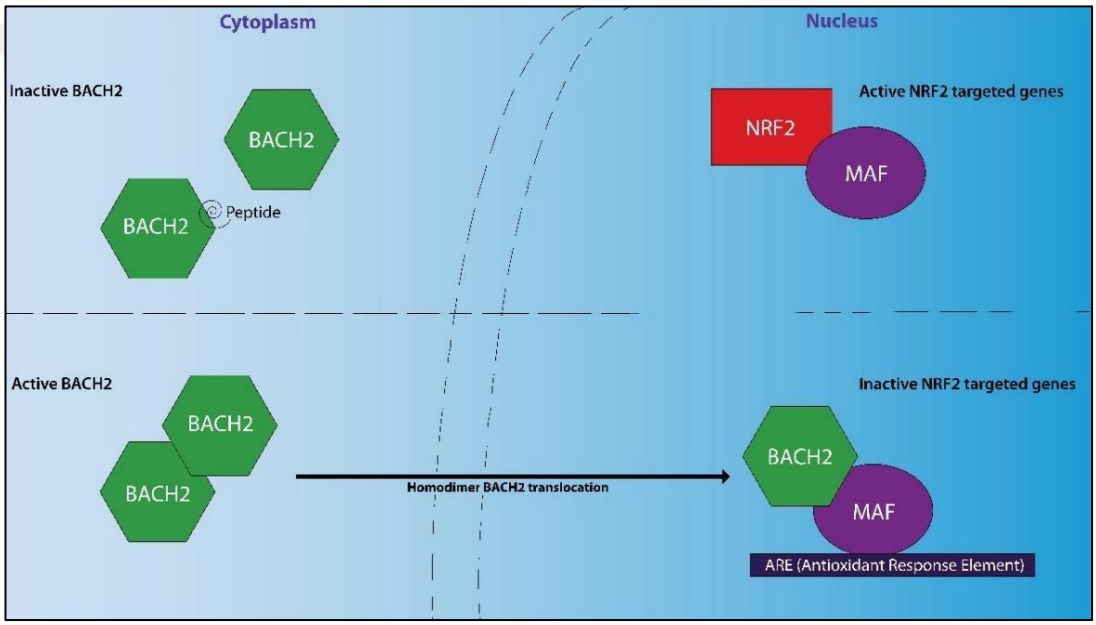

Figure 2: The predicted peptide mechanism on BACH2 BTB domain homodimerization. Activation of NRF2 targeted genes via BACH2 homodimer alongside nucleus shuttling prevention with peptide interaction. NRF2 (Nuclear factor erythroid 2-related factor 2), MAF (Transcription factor Maf), ARE (Antioxidant Response Element).

1.4 Targeting BACH2 BTB Domain and Role in Tumor Immunosuppression

BACH2 plays a huge role in immunologic tolerance, memory, and immunosuppression as a transcriptional repressor (54, 57). In light of these characteristics, several mutations in this gene are found to be related to autoimmune

diseases like asthma (58), multiple sclerosis (59), Crohn disease (60) and Hashimoto thyroiditis (61). These relations are documented with the help of BACH2-depleted mouse samples. These BACH2 knockout mice have high levels of eosinophil, macrophage, and effector CD4+ T cell populations compared to the negative control group (62, 63). Apart from that, BACH2's presence at the cellular level is essential in the differentiation of cytotoxic and peripheral T-cell populations (57, 64). Therefore, knocking out the BACH2 gene initiates a series of fatal inflammatory cellular actions. All this evidence shows that BACH2 plays an essential role in immune cell homeostasis. On the other hand, it is known that regulatory T cells limit the effector T cells CD4+ and CD8+ activities (64), which results in developing immunosuppression. In this context, deficiency of BACH2 results in regression of tumor growth related to CD4+ and CD8+ T cell populations (65). Parallel to this, BACH2's absence also results in antitumor activity by effector lymphocytes along with the decreased regulator T cell populations (65). This information in the literature shows that targeting BACH2 might be an approach for recognition by the immune system and result in a potential immunomodulator.

Up to this date, there is not any designed BACH2 activator or inhibitor (66). In this regard, there is only one entry for drug development efforts specifically to BACH2's role in immunosuppression (66). In that project, induced BACH2 expression by tetracycline is applied and the PMA/ionomycin-based interferon-gamma genes activation is prevented (66). This system results in HDAC-3 inhibitor RGFP966 as a BACH2 inhibitor. This inhibitor is used as the positive control because it is observed that BACH2-based BLIMP-1 (B lymphocyte-induced maturation protein-1) repression, is related to HDAC-3 in inhibitor B cell populations.

1.5 Aim of The Study

The aim of this project is to design, optimize, and test the affinity of peptides, which have the potential to prevent BACH2 BTB domain homodimerization through dimers interaction surface via using in silico and in vitro approaches.

2 BACKGROUND

2.1 Previous Studies Towards NRF2 Activation

There are 3 known NRF2 activators in the literature, also called KEAP inhibitors. These are electrophiles, protein-protein interaction inhibitors, and multi-target drugs (1).

2.1.1 Electrophiles

Electrophiles are positively loaded particles that bind to electrons of another molecule. In this specific circumstance, the majority of pharmacological NRF2 activators mostly modify cysteine residues of thiol-rich KEAP1 protein covalently. This modification is done through oxidation or alkylation. There are several cystine-rich residues on KEAP1 (28, 29). Cysteines Cys-151, Cys-273, and Cys-288 are the ones that are prone to electrophile reactions (1). Other residues are Cys-226, Cys-434, and Cys-613. As an example, electrophilic modification of Cys-151 by Bardoxolone-methyl (CDDO-Me) disrupts the KEAA1 and CUL3 interaction. In the end, KEAP1 is clogged in an NRF2 bound conformation, and newly formed NRF2 escapes ubiquitination (67, 68). Another electrophile is Dimethyl fumarate. It is also used with cystine-targeting electrophiles that were approved by the FDA in 2012 for relapsing-remitting multiple sclerosis (33, 36). It also targets cys-151 of KEAP1 in NRF2 activation. Other drugs for KEAP inhibition are omaveloxolone (RTA-408) (69), Oltipraz (70), Ursodiol (71, 72) etc. The side effect of electrophiles is that they are not selective enough. Therefore, targeting the protein-protein interaction surface is seen as essential (1).

2.1.2 Protein-protein interaction inhibitors

In the case of protein-protein interaction, inhibitors interact with NRF2 to the Kelch propeller of KEAP1, which makes them more selective than electrophiles. In this type, peptides that originated from NRF2 interact with NRF2 while it is dimerized

with KEAP1 (41). These peptides are combined with the HIV-Tat sequence and calpain cleavage sequence to form cell-penetrating characteristics (42). These peptides are found to be effective for cerebral ischemia mice by protecting their cognitive functions. Known peptides originated mainly from NRF2, while small molecules are also used with similar functions. These molecules are tetrahydroisoquinoline (73, 74), thiopyrimidine (75), naphthalene (76), carbazone (77) and urea derivatives.

2.1.3 Multi-target drugs

The NRF2 is phosphorylated by GSK-3, which eventually leads to ubiquitination by the E3 ligase β-TrCP. The abnormal activity of GSK-3 is also related to pathologies such as AD, cardiovascular diseases, or cancer (78, 79, 80, 81). These pathologies lead scientists to develop GSK-3 inhibitors for the treatment of these diseases. As an example, Tideglusib (82) can be given. This drug is studied for AD patients in phase 2 trials, and it is known for inhibiting GSK-3. This also prevents NRF2 phosphorylation and eventually ubiquitination. Furthermore, Enzastaurin (83) is used for the treatment of solid and hematological cancers, but it failed in phase 2 and 3 trials. Yet, the same drug is given for NRF2 activation.

2.2 Motivation for the Study

NRF2 activity can be inhibited by targeting its interaction with BACH1 in the nucleus. BACH1 competes with NRF2 and forms heterodimers with MAF proteins. As per the study, HPP-4382 is found to inhibit BACH1 (43). Yet this procedure cannot be applied for its sub-family protein BACH2. The BACH2 gene is essential for cytotoxic T-cell differentiation, therefore changing its dimeric activity without disrupting the gene expression is seen as required (62). Furthermore, it is a transcriptional repressor (84), but also plays a role in immunologic tolerance and immunosuppression (85). Also, BACH2-depleted mice show high levels of eosinophil, macrophage, and effector CD4+ T cell populations. This points out that a protein-protein interaction inhibitor is predicted to be applicable and selective. The lack of this

peptide in the literature targeting BACH2 leads us to research possible peptide candidates for this specific study.

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 HEK293T cell line

The presence of BACH1 and BACH2 proteins in different cells forces researchers to observe the concentration of these proteins in limited cell lines. One of these cell lines is Human Embryonic Kidney 293 (HEK293) which is used for transfection and protein isolation in this project. HEK293 cell line is an immortalized cell line obtained from a female fetus (86). Unlike the name "immortal" this cell line cannot proliferate indefinitely and rapidly, yet due to several mutations, it dodges the cellular senescence. This makes them usable while culturing and passaging in vitro. These cells are semi–adherent cells and can be used for transient and stable expression (86). Because of these characteristics, they are mainly used in cancer research, protein production, and protein–protein interactions (87), (88). A version of this cell is HEK293T. In this version, HEK293 cells are stably transfected with a plasmid containing SV40 large T antigen (pRSV1609). The SV40 large T antigen is expressed in cells and therefore the plasmid induces production of SV40 origin of replication in cells. This allows cells to increase their recombinant protein production (89).

The proliferation of HEK293 cells is an advantage, yet culturing and passaging them for long periods brings also a disadvantage with it, which is translation inefficiency. According to the literature, HEK293 cells' health is degraded and loses its growth rate and translation efficiency after 20 passages. It is also vulnerable to human–specific–viruses (90).

3.1.2 Peptide design

The therapeutic approach against cellular abnormalities varies due to differences in pathologies. In BACH2's case the problematic starts with its essential role in T – cell differentiation. This prevents scientists from knocking the gene off in tumor

presence, which makes it difficult to deal with. In addition to that, in the case of tumors, BACH2 gains a tendency to create dimers, which eventually shuttles the complex into the nucleus and causes its competition with NRF2.

3.1.3 Molecular Dynamics (MD) analysis

The computational technique known as Molecular Dynamics (MD) simulations uses Newton's laws to assess the motions of ions, water, small and macromolecules, or more complex systems to simulate and mimic the behavior of the natural biological environment. It includes lipid membranes and water molecules. Structural motions and movements, which are dependent on the temperature, pressure, solute, and solvent are very significant in determining protein–protein and protein–peptide complexes. For this reason, MD simulations are applied before in vitro approaches. Specific to MD, it mostly sheds light on drug design by helping to evaluate molecular affinity and stability between structures (91). In this sense, several tools and simulation platforms are used such as VMD (Visual Molecular Dynamics) and much more.

3.1.4 Visual Molecular Dynamics (VMD)

Visual Molecular Dynamics (VMD) is software, which is developed by developed by the Theoretical and Computational Biophysics Group for the 3-D visualization, animation, and analysis of large biomolecular systems, featuring built-in scripting functions. It can represent, render, and color many molecular assemblies simultaneously. It requires a PDB format of macromolecule for visualization and a PSF format for basic dissociation simulations (92).

3.1.5 Peptiderive

Rosetta Peptiderive is a protocol developed by Yuval Sedan at the Furman Lab, which is used for determining protein-protein interaction areas as peptide segments with desired length (93). The segment that contributes the most to interaction is given as a result. The algorithm requires the PDB format of the macromolecule and the sub-

chains to work. After the specification of the protein and its receptor chain, the system starts to predict a peptide from the opposing chain. The procedure starts with minimization of the structure with Rosetta, which removes local clashes. Then the amino acid, whose length is determined, slides on the opposing chain. At each window position, the peptide–receptor chain interaction is observed, and binding energy is calculated (94).

For the calculation, the following function is applied to the structure: ΔΔGAB = ΔGAB – ΔGA - ΔGB. This gives the interaction between chains A and B of the protein. Then it is followed by the formula: relative binding energy = ΔΔGPAB / ΔΔGAB. That gives the relative binding energy of the structure with a peptide that is predicted from chain A. As result, relative binding energy scores are given according to requests (93).

3.1.6 LIGPLOT+

LIGPLOT is a bioinformatics tool, that is used to analyze the molecular background of peptide/ligand–protein interaction and generates molecular structure images (95). As input, PDB files of the protein and peptide are required.

3.1.7 FoldX

FoldX is a macromolecule designing algorithm that uses an empirical force field. The algorithm is particularly useful in understanding the functional importance of specific mutations and their potential impact on protein function and overall cellular processes. Furthermore, it can also be used to design peptides with the same algorithm (96). FoldX calculates the free energy binding of peptide unfolding, which is important for protein stability. The structure is mutated one by one into every other amino acid and alternative conformations are checked (97). The FoldX algorithm for peptide design calculates the free energy binding by considering various energy contributions. These include the Van der Waals interactions (ΔGvdw) between atoms in the peptide and the solvent, and the difference in solvation energies for apolar and polar groups

(ΔGsolvH and ΔGsolvP), when transitioning from the unfolded to the folded state (97). The free energy difference between intra-molecular hydrogen-bond formation (ΔGhbond) and intermolecular hydrogen–bond formation with the solvent, the extra stabilizing free energy (ΔGwb) provided by water molecules forming multiple hydrogen bonds to the protein, the electrostatic contribution of charged groups (ΔGel), including the helix dipole, the entropy cost (ΔSmc) for fixing the backbone in the folded state, and the entropic cost (ΔSsc) of repairing a side chain in a particular conformation. In simpler terms, when using the FoldX algorithm for peptide design, the free energy binding is calculated by taking into account various energy factors such as interactions between atoms, solvation energies, formation of hydrogen bonds, stability provided by water molecules, electrostatic contributions, entropy costs (97). The formula for this free energy is $\Delta G = \Delta Gvdw + \Delta GsolvH + \Delta GsolvP + \Delta Ghbond + \Delta Gwb + \Delta Gel + \Delta Smc + \Delta Ssc$.

If the score is above 0,5 kcal/mol, it is an indicator for structural instability; if it is lower than -0,5 kcal/mol, it is an indicator for structural stability (96).

3.1.8 PEP-FOLD 3

A de novo method called PEP-FOLD is used to predict peptide structures from amino acid sequences. This approach combines the projected series of SA letters to a greedy algorithm and a coarse-grained force field, based on the structural alphabet SA letters to characterize the conformations of four successive residues (98). PEP-FOLD3 performs a sequence of 100 simulations, starting with a single amino acid sequence consisting of five to fifty standard amino acids. Each simulation takes a distinct conformational space sample. It yields an archive containing every model created, together with the cluster details and optimal conformation of the top five clusters. The system works by assuming pH as neutral (pH=7) (99). However, it has also some limitations. Amino acid sequence size is one of them. It is evaluated that the algorithm works for less than 5 and more than 50 amino acid peptides, yet the greater the size is, the more inaccurate the structure. Furthermore, PEPFOLD 3 merely takes the standard

20 amino acids. It is unable to process peptides containing D-amino-acids, or unusual L- amino acids. At the end 5 peptide predictions are given as results (100).

3.1.9 Cluspro

ClusPro is an algorithm for global protein-protein docking and is used for several different applications. It provides several advanced options to further customize the search, such as removing unstructured protein regions, implementing attractions or repulsion, taking pairwise distance constraints into account, creating homo-multimers, taking into account small angle X-ray scattering (SAXS) data, and identifying heparin binding sites. Specific to peptide docking, it retrieves data from PDB and executes a Fast Fourier Transform (FFT) based docking approach. This is a fast global rigid body docking of these fragments to the receptor (101). The method has two steps. The first is the rotation step. The peptide/ligand is rotated 70.000 times and translated in the x, y, and z axes relative to the receptor protein. 1000 of these translations with the highest and lowest scores of each rotation are chosen. The second step is clustering. These 1000 peptides/ligands are clustered with a 9 Å C-alpha RMSD (Root – mean – square – deviation) radius. In summary, the peptide or ligand position with the most neighbors in 9 Å is determined. This position is then centered as a cluster center and its neighbors are accepted as cluster members. This method is applied to another position once again and the procedure continues for further positions (101).

3.1.10 HADDOCK

HADDOCK (High Ambiguity Driven protein-protein Docking) is a versatile local docking method, which is information-driven and used to simulate biomolecular complexes. It differs from ab initio docking approaches by using ambiguous interaction restraints (AIRs) to drive the docking process by encoding information from known or expected protein interactions. Moreover, it supports a wide range of other experimental data, such as NMR (nuclear magnetic resonance) residual dipolar couplings, pseudo-contact shifts, and cryo-EM maps. It also enables the establishment of precise, unambiguous distance constraints. Proteins, small molecules, nucleic acids,

peptides, cyclic peptides, glycans, and glycosylated proteins can all be docked using the most recent version (v2.4) (102).

Stepwise, it is similar to ClusPro, yet the HADDOCK docking protocol automatically clusters output structures based on similarities and gives a score to each cluster that is created. At least four output structures with RMSD values within 2 Å apart comprise each cluster (102).

3.1.11 Anaconda

Anaconda is a distribution software of Python and R programming languages for scientific computer analysis.

3.1.12 Spyder

Python scientific programming can be done with the open-source, cross-platform integrated development environment called Spyder. It integrates with several well-known packages in the scientific Python stack, such as NumPy, SciPy, Matplotlib, pandas, IPython, SymPy, and Cython. It is also used for all-to-all RMSD correlation analysis.

3.2 Methods

3.2.1 Wild-type peptide prediction

Today, there are only 2 BACH2 BTB domain structures as dimers can be found in the Protein Data Bank (PDB). They are subdivided into 3OHU and 3OHV, which are differentiated according to their disulfide bond on the Cys20 residue. They both have 2 chains (Chain A and B), yet the one without a disulfide bond (PDB ID: 3OHU) is chosen, because it is more convenient for peptide design. Chosen BACH2 BTB domain structure is then observed under the online tool “Rosetta Peptiderive”, to predict a peptide with 15 amino acid length, that participates the most in homodimer

interaction. This tool uses a protocol, which calculates relative binding energy between two structures. After the given PDB file and desired amino acid length into the tool, the algorithm calculates the relative binding energy between chains. The function is applied to all 15 amino acid sequences, which contribute to the interaction. Finally, the one with the most is chosen as the targeted region and wild-type peptide. Apart from this, the predicted peptide does not change, if the receptor is selected as chain A or B, therefore receptor chain is selected as B randomly.

3.2.2 Interaction analysis

After the selection of wildtype peptide, dimeric interaction with the BACH2 BTB domain is analyzed with "Ligplot+" by determining hydrogen bonds and hydrophobic interactions. This shows any interaction between peptide and BTB domain. Apart from that, Ala97-Arg103 amino acids in chain A of the BACH2 BTB domain are isolated due to the nonspecific binding of hydrogen bonds.

3.2.3 Peptide design

The amino acids of the predicted wildtype peptide on the BACH2 BTB domain homodimer are converted into other amino acids and optimized with the "FoldX" algorithm. This method is approved as a half-empiric algorithm, that calculates the free energy of protein folding between wildtype (ΔGwt) and mutant (ΔGmt) peptide. The method processes all point mutations in the peptide in each residue and gives a stable result as kcal/mol. If this stability is above 0,5 kcal/mol, it is an indicator of structural instability; if it is lower than -0,5 kcal/mol, it is an indicator of structural stability. It uses the "pssm" command to predict interaction energy ΔΔG. Finally, 16 different peptides are designed.

3.2.4 Peptide structure prediction

Firstly, a cell-penetrating peptide is added into the structure, which is HIV Tat 47-57. After the "FoldX" gives results, these peptides' structures are predicted by an

algorithm called “PEP-FOLD3”. The algorithm runs 100 simulations, which sample different areas of the conformational space. In the end, it ends up with 5 structure predictions with the best conformations and clusters.

3.2.5 Global docking

The project is carried on with a global docking procedure with the online tool “ClusPro”, which tests if the predicted peptide structures target the BACH2 BTB domain dimerization region. That region includes Ala97, Lys98, Leu99, Leu100, Leu101, Ser102, Arg103 amino acids. All created peptides in the library are tested in terms of their targeted regions along with wild-type peptides. This method is also applied to the BACH1 BTB domain and BCL-6 BTB domain dimers.

3.2.6 Local docking

After the peptide structure and global docking profile are predicted, the local docking method is applied with the online tool “HADDOCK”. It uses a priori method and a combination of different docking algorithms and bioinformatic models. Therefore, it promotes conformational flexibility. The selected area of the peptide is between 12 and 26th residues and it targets the 97-103 residues of the BACH2 BTB domain because it yields more hydrogenic bonds. In the end, the protein–peptide complex with the highest HADDOCK score is used for molecular dynamics and free energy analysis in further steps.

3.2.7 Preliminary tests before molecular dynamics

At the end of docking procedures, the peptide library is reduced to 6 (peptide 1, peptide 4, peptide 12, peptide 13, peptide 14, peptide 15). Before the simulations, protonation states of amino acids in targeted proteins and peptides are tested with the “PROPKA” algorithm in the PDB2PQR program. All complexes are solved in water and neutralized in 150 mM NaCl.

3.2.8 Molecular dynamis analysis

These peptides undergo NAMD-based molecular dynamics (MD) analysis with CHARMM parameters. The system models water molecules in terms of the TIP3P model. In all simulations, constant temperature and pressure are applied. Long-range Coulomb interactions are calculated with the particle mesh Ewald algorithm. In order to determine unbound electrostatic and van der Waals interactions, a 17 Å range limit is set. The Langevin piston Nosé−Hoover Peptide method stabilized the pressure to 1 atm and temperature to 300 K. Hydrogen bond length is limited with the SHAKE algorithm and system coordinates are measured for every 2 picoseconds. The energy of all systems is minimized with the conjugated gradient method. This is followed by an increase in temperature from 10 K to 300 K on the 30th picosecond. This analysis is run 4 times with 500 nanoseconds for all peptides and negative controls. These simulations are run with TRUBA platform of TUBİTAK. In every MM-PBSA work, relative binding free energy (RBFE) levels of protein chains and protein–peptide interactions are calculated.

3.2.9 Data analysis of MD results

All protein–protein complex structures are analyzed with the program "Visual Molecular Dynamics (VMD)". Convergence of simulation to conformational stability is calculated with root-mean-square-displacement of core atoms (C, N, Cα, O).This method is run with the algorithm called "Spyder" in the program "Anaconda". RMSD is observed for 500 nanoseconds and repeated 4 times. In the case of an RMSD score lower than 3Å, the structure is considered conformationally stable and evaluated as eligible for deeper analysis. This procedure is followed by all-to-all RMSD and RMSF (Root-mean-square-fluctuation) analysis of carbon alpha flexibility. The analysis is finalized with solvent-accessible surface area (SASA) to verify the RMSD, RMSF, and range results. Finally, peptide 13 is chosen at the end with the highest affinity.

3.2.10 Principal component analysis

The collective dominant movement of atom groups during simulations is determined with principal component analysis (PCA). This method is also applied for positive control; BACH2 BTB domain monomer on chain A of homodimer and the same complex which is disturbed by peptide 13. Therefore, the collective dominant movement of atoms is checked for positive control complex, complex with low affinity (wildtype peptide (peptide 1) – protein) and with high affinity (mutant peptide (peptide 13) – protein). In the first step of PCA analysis, translational and rotational movements are eliminated with frame superimpose, and inner movements are isolated. The PCA analysis is later on diagnosed to determine the eigenvector and eigenvalue and undergo cartesian covariance matrix analysis. The eigenvalue is calculated as mean – square – fluctuation and the highest eigenvalue is considered as the most dominant collective movement. Results are shown in 2 ways, which are the movement interpolation on the structure and eigenvalue distribution according to the eigenvector index. Apart from that, every move is determined by cross-correlation analysis. The flexibility of BACH2 monomers is analyzed with an eigenvalue index. To achieve this, the last 250 nanoseconds of simulation are examined 4 times. All these PCA analyses are accomplished with Python-based software called ProDy.

3.2.11 Cross correlation

Finally, the movement of the N terminus region and amino acids near that region in the BACH2 BTB domain monomer is determined with amino acid alpha-carbon cross-correlation analysis. This analysis is run through the Python-based software ProDy.

3.2.12 Peptide synthesis

After the selected mutant and wildtype peptides are run through several simulations, they are purchased from BIOMATIK for in vitro approaches. These

peptides are asked to be designed as 100 milligrams with 95% purification. They are put under -20°C and sterile.

3.2.13 HEK293T cell line and its growth

In the supernatant depletion assay, the HEK293T cell line is used to overexpress myc – DDK tagged BACH2 and myc – DDK tagged BACH1 overexpression along with their protein lysate production. It is an embryonic epithelial cell line and contains SV-40 antigen. HEK293T cells are cultured in high glucose DMEM + %10 FBS medium with 37°C and 5% CO2. For cell growth, T-25 and T-75 flasks are used. Subculturing of this cell line is done with 1:5 dilution every 2 days. Further subculturing processes for long periods (3 or 5 days) are done with 1:20 dilution and 1:100 dilution. In this subculturing process, cells are detached from the flask with the help of 0.25% Trypsin-EDTA solution. The medium is replaced by a new medium every 3 days. Cell growth is controlled under the light microscope day by day, and the amount is counted with a hemocytometer when it is necessary. In T-25 flasks, it is seen that 5x106 cells are present in the case of 80% fullness.

3.2.14 Plasmid production and isolation

HEK293T cell growth is followed by plasmid production, which is purchased by ORIGENE. The first plasmid is a backbone plasmid, which can express myc - DDK tag (pCMV6-Entry, mammalian vector with C-terminal Myc - DDK Tag), the second one is BACH1 overexpression plasmid (BACH1 (NM_206866) Human Tagged ORF Clone- RC221628) and the third is BACH2 overexpression plasmid (BACH2 (NM_021813) Human Tagged ORF Clone – RC214061). They are purchased as ready for transfection with 10 micrograms. Its concentration is lowered to 0.1 µg/µ by solving with 100 microliters of distilled water and stored at -20°C. 1 picogram – 100 nanograms of plasmid DNA is used with chemically competent DH5α E. Coli cells for plasmid production. The transformation protocol starts with 30 minutes of incubation after the DNA and cell line are infused. It is followed by heat – shock for 30 seconds at 42°C and incubation in ice for 2 minutes. After that, cells are recovered with SOC

medium addition. The culture with SOC medium is centrifuged for 60 minutes with ~200 rpm at 37°C. Finally, 5 – 200µl of sample along with antibiotics (25ug/mL of kanamycin and ampicillin) are taken from SOC medium and spread onto a petri dish containing solid LB. These petri dishes are incubated at 37°C for 16-18 hours. This step is followed by taking a sample from produced cell colonies and growth with liquid LB containing antibiotics. The volume is selected according to the MiniPrep plasmid isolation protocol. Some of these cells are prepared for plasmid isolation, while the remaining are stored with glycerol stock at -80°C. Bacterial cell culture, which is taken from liquid LB, is subjected to a mini–prep isolation protocol for the preparation with HEK293T cell transfection. Plasmid isolation is applied with QIAprep® Spin Miniprep Kit. The culture from the day before is centrifuged for 3 minutes at higher than 8000 rpm. The cell pellet is taken and solved in microcentrifuge tubes with 250µl P1 solution. This is followed by a 250µl P2 tampon and inverted 4 to 6 times for cell lysis. After the 5-minute incubation time 350µl of N3 solution is added and centrifuged for 10 minutes with 13000 rpm. The supernatant is taken to QIAprep spin colon, and it is washed with 0.5 ml PB solution, and centrifuged for 60 seconds. Before elution, additional washing tampon is cleared with 1 more minute centrifugation. The sample inside the QIAprep spin colon is transferred into new microcentrifuge tubes with 1.5 ml volume, incubated with 50µl EB tampon, and centrifuged for 1 minute with maximum speed. This protocol is repeated one more time to acquire more plasmid DNA.

3.2.15 Overexpression of myc – DDK tagged BACH2 in HEK293T cell line

After the plasmid isolation, they are planned to be transfected to HEK293T cells. To achieve this, 60 mm cell culture plates are used. The cell is diluted with a 1:10 ratio from a fully confluent T25 flask, and then cultured in 60 mm plates. They are incubated overnight for cell adhesion to the plate surface. The day after, 50-70% confluency is reached. 3 µg of DNA is diluted in 750 µl medium without serum and added 9 µl of Turbofectin transfection agent. This is incubated for 15 minutes at room temperature. Later on, this sample is added onto 60 mm plates drop by drop and put in incubation for 48 hours at 37°C with %5 CO2. Two days later, the medium is taken away and the

cells are scraped with a scraper after washing with 2 ml PBS and transferred into 15 ml falcon tubes. These tubes are centrifuged for 5 minutes with 125 x g. For protein extraction, the M-PER® Mammalian Protein Extraction Reagent kit by Thermofisher is used. Before the application, complete protease and phosphatase inhibitor is given along with 1 mM PMSF. The medium is removed in the first step and washed with PBS. For each 60 mm plate, 250 µl extraction reagent is given and waited for 5 minutes. The lysate is transferred into microcentrifuge tubes and centrifuged for 15 minutes with 14000 x g. The supernatant is then taken and given into another tube for analysis. Four 60 mm plates are used in this process. The first plate contains untransfected HEK293T cells, the second contains cell with backbone plasmid, the third contains HEK293T cells transfected with full length wildtype myc – DDK tagged BACH1 plasmid, and finally the fourth contains same cell line transfected with full length wildtype myc – DDK tagged BACH2 plasmid.

The BACH2 and BAH1 protein overexpression is tested with western blot procedure. 500 mM of reducing agents (Dithiothreitol or %5 beta – marcaptoethanol) is added into gel solutions and mixed with 20 µg of protein. This protein is then dissociated with 10% SDS – polyacrylamide gel. It is later on transferred to PVDF membrane. This membrane is blocked with 5% non – fat milk, which is prepared with PBSTween-20. It is later on incubated with primary anti – myc anticore for 16 hours at 4°C. The next day, it is followed by washing with PBS-Tween20 and HRP-tagged secondary anticore incubation for 1 hour at room temperature. They are finally washed with PBSTween-20 once again. At the end, the bands are checked with the help of SuperSignal™ West Pico PLUS Chemiluminescent Substrate and their luminescence is analyzed. Primary anticore is anti – myc tagged (Cell signaling #2278) and secondary is HRP tagged anti – rabbit (Cell Signaling #7074), and they are purchased by the company called “Cell Signaling”. In immunoblotting, beta actin lecels are checked with anti-beta actin anticore as loading control (Cell signaling #4970).

3.2.16 Supernatant depletion assay

In this step, peptides with biotin on their C terminus interact with different concentrations of Dynabeads that contain streptavidin (Dynabeads M-280 Streptavidin). In the literature, there is no affinity data for peptide 13 or peptide1, therefore in vitro dissociation constant of BACH2 – heme interaction is selected as initiation point (BACH2 – heme dissociation constant Kd: 166.2 nM). For supernatant depletion assay, peptides with concentrations of 1200 nM (60 pmol), 600 nM (30 pmol), 300 nM (15 pmol), 150 nM (7.5 pmol), 75 nM (3.75 pmol), and 0 nM (0 pmol) in 50 µl are aimed to be attached with beads of streptavidin dynabeads. The graphics are sketched according to molecular weights and purity of peptides in detail. The protocol starts with 30 minutes of incubation at room temperature of peptide and dynabeads with PBS. Peptide-attached beads are separated with a magnet and washed 3 times with 0.1 %BSA (bovine serum albumin)-PBS. The receptor for supernatant depletion assay, protein lysate supernatant of HEK293T that overexpress myc – DDK tagged BACH2 is used as the receptor. In this context, the purification of total protein extract is not seen as essential. The concentration of Myc – DDK tagged BACH2 containing total protein extract is stabilized to 0.15 µg/ul for the whole process. It is determined, that dynabeads - streptavidin pre–clearing process is not required. Yet, depletion assay only with dynabeads – streptavidin is used for negative control. Therefore, it helps to predict the dissociation constant of the circumstance, which does not contain an unspecific binding level.

Approximately 7.5 µg of cell lysate is incubated with changing peptide concentrations. At this point, peptides with biotin on their C terminus along with streptavidin – dynabeads precipitated with a magnet after 30 minutes of incubation. Samples from the supernatant (2.27 µg) are analyzed within the scope of densitometric signal through western blotting. To determine the protein amount, an optimization experiment is performed. 0 nM of peptide containing myc signal is selected as 100%, which helps the analysis of increasing peptide volume effect through other myc – signaling. The graph that points out this is sketched with Graph-Pad Prism to predict

dissociation constant (GraphPad- Binding- Saturation binding to Total and Nonspecific). The analysis is repeated 3 times.

4 RESULTS

4.1 Wildtype Peptide Prediction

The project starts with choosing receptor and partner chains of the BACH2 BTB domain. The "Peptiderive" results show that the peptide is the same for both chains, therefore, chain A is selected as the receptor chain. The peptide is predicted from chain B. Its location is between Pro12 and Leu26. The interface score of the predicted wildtype peptide sequence from chain B is calculated as -49.807.

Table 1: The cell penetrating peptide (CPP) and predicted wildtype (control) peptide.

Cell Penetrating peptide	**Predicted Peptide**
YGRKKRRQRRR	PMYVYESTVHCTNIL

4.1.1 Interaction analysis of the wildtype peptide

The prediction of the peptide is followed by the interaction analysis via the Ligplot+ algorithm. The results show that the sequence Met11-Ser16 in the peptide interacts with Ala97-Arg103 through hydrogen bonds (Fig. 3). A few amino acids in the peptide sequence form a hydrophobic interaction with chain A, yet it is known that hydrogen bonds control the specific binding of structures. As shown in the same figure, Ile 23, Leu24, Cys20, and Ala57 on chain A form hydrophobic interactions with His19, Cys20, Ile23, and Leu24 of the peptide (Fig. 3). This predicted peptide is aimed at be used in vitro in further steps; therefore, it is combined with an already known cell-penetrating peptide (HIV Tat 47-57). The predicted peptide (PP) with cell-penetrating peptide is shown in Table 1 with their one-letter abbreviations.

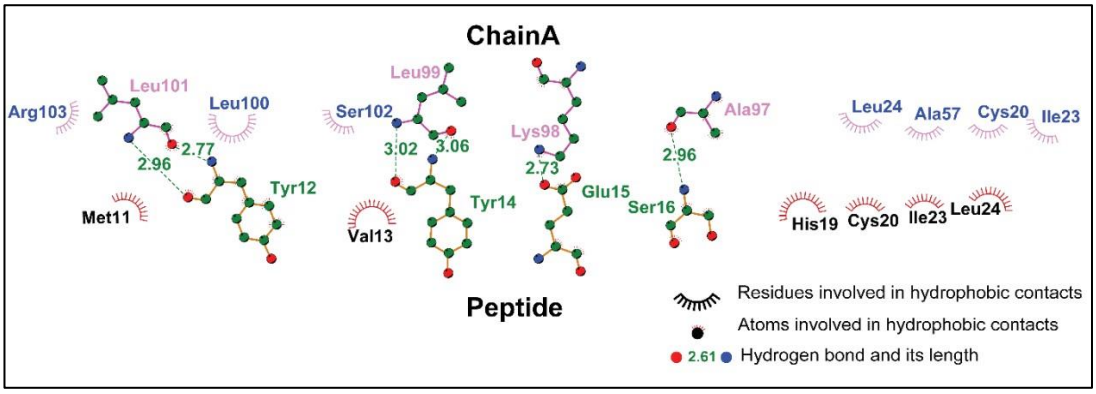

Figure 3: Ligplot+ analysis of hydrogen bonds and hydrophobic interactions between predicted peptide from chain B and protein (chain A). 97-103 sequences on chain A of the protein forms hydrogen bonds with Met11-Ser16 of the peptide.

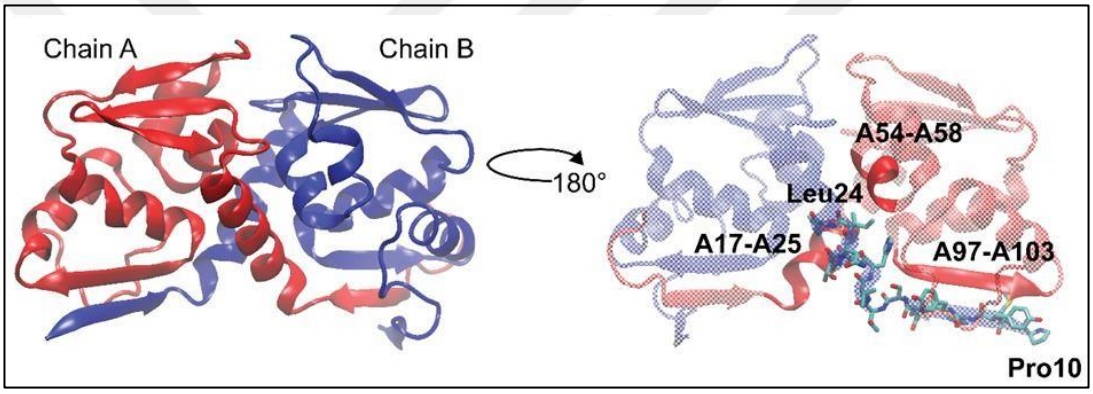

Figure 4: Crystal structure of BACH2 BTB domain (PDB ID: 3OHU) (left) and its 180°-degree rotated version to show Rosetta Peptiderive results (right). 15 amino acids that are predicted from chain B interact with A17- A25, A54-A58, A97-A103.

4.1.2 Wildtype peptide structure

In the next step, the wildtype peptide's structure is formed with the PEPFOLD 3.5 algorithm. It is observed that it does not create any secondary structure, yet the cell-penetrating peptide (CPP-HIV Tat47-57) on the N terminus gains an alpha-helical conformation as expected (Fig. 4).

4.1.3 Docking via HADDOCK

The predicted wildtype peptide is then locally docked onto the BACH2 BTB domain monomer with the HADDOCK algorithm. The active sites are chosen as amino acids 1226 of the peptide and 97 – 103 of the chain A. The structure with the highest HADDOCK score is selected for further MD analysis (Fig. 5A).

4.1.4 MD and trajectory analysis

The docked peptide and the monomer are analyzed with MD analysis for 100ns. This analysis predicts free binding energy. In the meantime, trajectory analysis is applied to determine the convergence of conformational stability by calculating the RMSD score along with all-to-all RMSD. Between 10 and 100ns BACH2 BTB domain monomer-peptide complex has similar conformations in all simulations (Fig. 5B, C, D), while this result is supported by an RMSD score which is below 3Å between 10 and 100ns (Fig. 5F). The complex keeps its conformational stability through the simulation time. For comparison, homodimer BACH2 BTB domain and its RMSD / all-to-all RMSD scores are given for positive control (Fig. 5E). This homodimer has also an RMSD score lower than 3Å in 10-100 ns and similar all-to-all RMSD pattern.

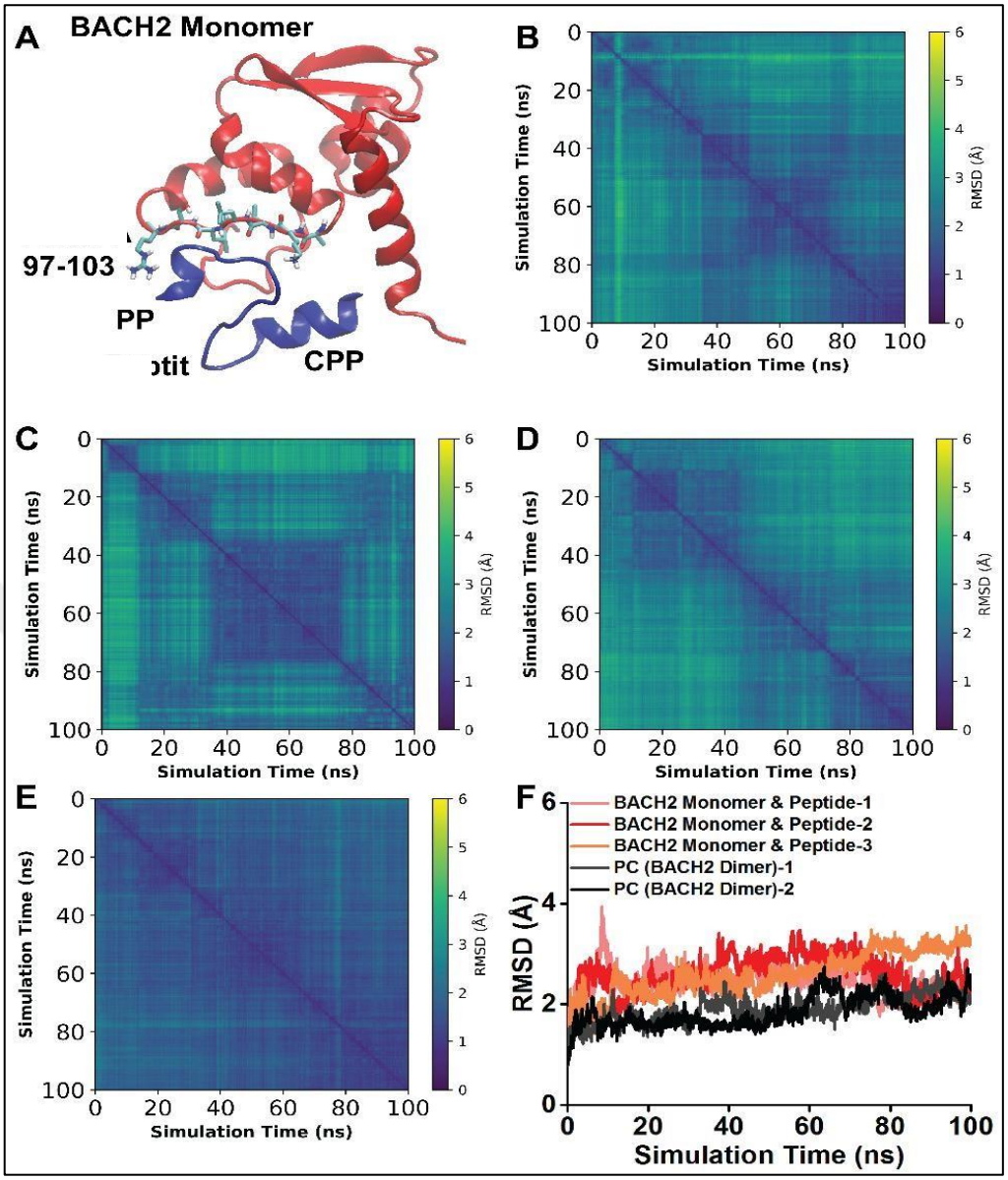

Figure 5: A) Structure prediction of wildtype peptide and CPP with PEPFOLD 3.5. Secondary conformation is not observed for the peptide, yet the CPP forms alpha helical conformation as expected. The image is shown after the local docking with HADDOCK on to residues 97.-103. These residues are chosen due to the high hydrogen bond amount compared to other residues. Blue: BACH2 monomer (chain A), Red: Peptide (CPP+PP). B, C, D) All-to-all RMSD (Root mean square displacement) graphics, that show MD simulations for 100ns of BACH2 BTB domain monomer-peptide and its similar conformations after 10^{th} ns (RMSD<~3 Å), (replicates 1, 2, 3 are shown in B, C and D respectively). E) All-to-all RMSD graphics, that show MD simulations for 100ns of homodimer BACH2 BTB domain (positive control), (RMSD<~3 Å). F) RMSD-time graphics of B, C, D and E results. RMSD and all-to all RMSD results confirm each other. Protein-peptide complex is replicated 3 times and positive control is replicated twice.

4.2 Relative Free Binding Energy Calculation with Wildtype Peptide

The last 10 ns of 100ns simulations (90-100ns) are selected for predicting relative binding free energy (ΔG (Interaction)) with the MM-PBSA method. It is observed that the relative binding free energy of peptide–monomer is 21 times greater than the homodimer BACH2 BTB domain complex ($p<0.05$) (Table 2). In the same table ΔE(VDWAALS), ΔE (EEL), ΔG (Gas), ΔE (EPB), ΔE (ENPOLAR), ΔE (EDISPER), and ΔG (Solvation) are calculated as well.

Table 2: Relative Free Binding Energy principles for BACH2 BTB domain monomer-peptide complex and positive control BACH2 BTB domain complex

	BACH2 BTB domain monomer–peptide complex		**Homodimer BACH2 BTB domain complex**	
	Energy (kcal/mol)	±SEM	Energy (kcal/mol)	±SEM
ΔE (VDWAALS)	-64.0	4.4	-227.4	2.2
ΔE (EEL)	-878.1	67.7	-327.6	2.1
ΔE (EPB)	+869.1	64.7	+421.3	4.6
ΔE (ENPOLAR)	-67.1	2.7	-175.0	1.0
ΔE (EDISPER)	+114.7	6.7	+307.5	1.4
ΔG (Gas)	-942.1	69.1	-555.0	0.1
ΔG (Solvation)	+916.7	65.7	+553.8	5.0
ΔG (Interaction)	-25.4	5.2	-1.2	4.9

4.3 Mutant Peptide Predictions and Library

At this point, several point mutations are tried for the wildtype peptide to estimate the contribution to the stability of the complex by different peptides. To do this FoldX algorithm is used by using pssm command. The amino acid mutations with the most electrostatic contribution are determined by the FoldX algorithm (Table 3).

Table 3: ΔΔG values of hydrophilic amino acid mutations, that are predicted to change BACH2 BTB domain homodimer interaction via FoldX algorithm

	Chain A and B mutations			
	Mutation	AVG1	SD2	SEM3
1	PA10R[4]	-0.61	0.19	0.09
2	SA16E	-0.59	-0.58	0.26
3	CA20H	-1.43	0.10	0.05
4	N22BR	-1.00	0.20	0.09

These point mutations are applied to wildtype peptide and a peptide library consisting of 15 peptides is created (Table 4).

Table 4: Predicted peptide sequences for in silico docking

Cell penetrating peptide	**Designed peptide**
YGRKKRRQRRRA	PMYVYESTVHCTNIL (WT)
YGRKKRRQRRRA	**R**MYVYESTVHCTNIL
YGRKKRRQRRRA	PMYVYE**E**TVHCTNIL
YGRKKRRQRRRA	PMYVYESTVH**H**TNIL
YGRKKRRQRRRA	PMYVYESTVHCT**R**IL
YGRKKRRQRRRA	**R**MYVYE**E**TVHCTNIL
YGRKKRRQRRRA	**R**MYVYESTVH**H**TNIL
YGRKKRRQRRRA	**R**MYVYESTVHCT**R**IL
YGRKKRRQRRRA	PMYVYE**E**TVH**H**TNIL
YGRKKRRQRRRA	PMYVYESTVH**H**T**R**IL
YGRKKRRQRRRA	PMYVYE**E**TVHCT**R**IL
YGRKKRRQRRRA	**R**MYVYE**E**TVH**H**TNIL

The prediction of these peptides is followed by the structure prediction via PEPFOLD 3.5 (Fig. 6). Cell-penetrating peptide in peptides 10, 12, 13, 14, 15, and 16 form helical structures. Furthermore, these peptides show a secondary structure conformation in their designed positions. All the peptides are predicted, and their

secondary structures are shown. 12. Alanine is differentiated from CPP; therefore, it is shown in every peptide. The designed area does not form any secondary structure in P1, P2, P3, P4, P5, P6, P9, and P1, but CPP is formed as a helical structure (Fig. 6).

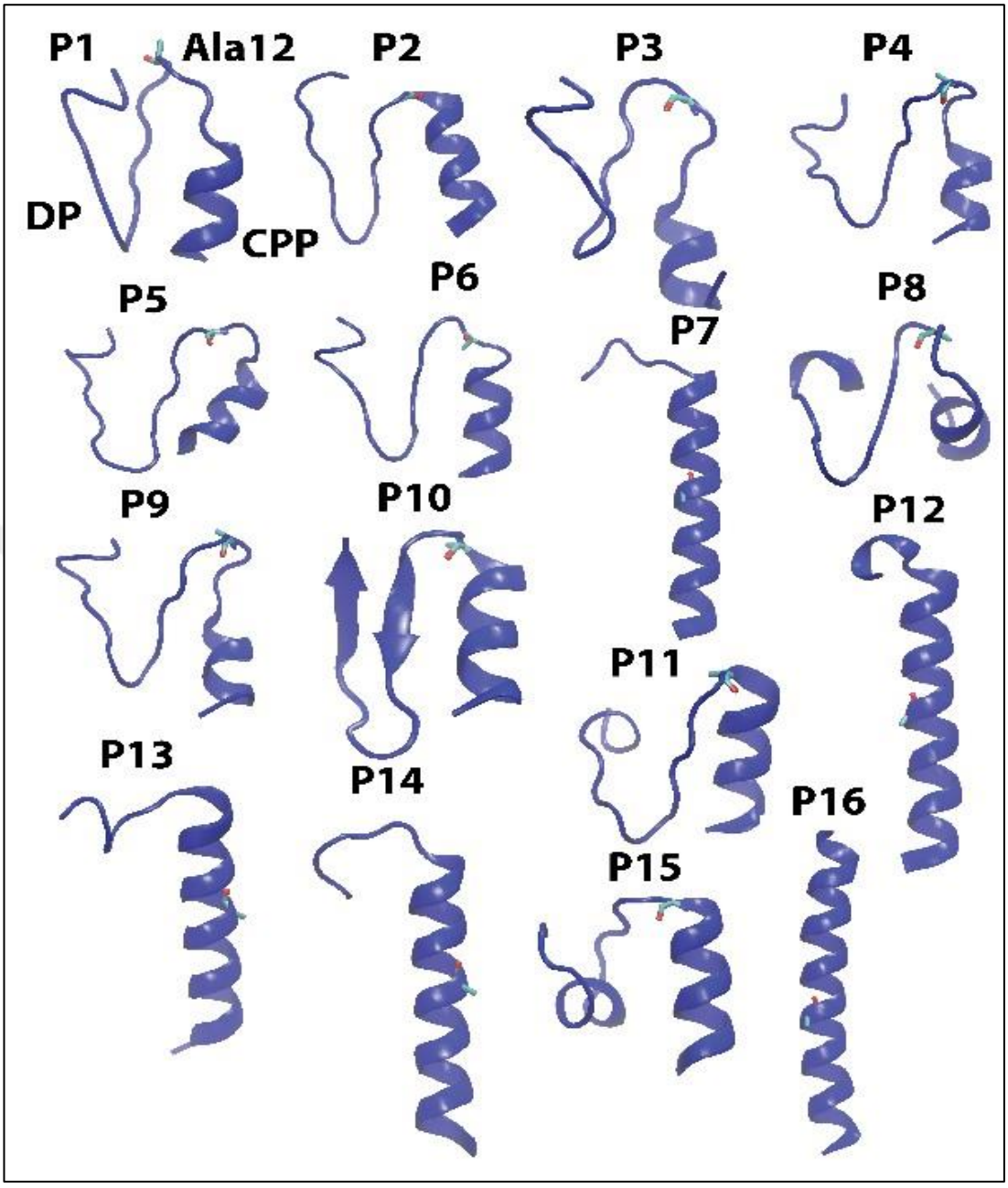

Figure 6: Structural predictions of peptides with PEPFOLD 3.5, whose sequences are shown in table 4. 12th alanine (Ala12) is the residue where designed peptides are differentiated from cell penetrating peptide. It is shown in all peptides. Secondary conformation is seen in peptides; P1, P2, P3, P4, P5, P6, P9 and P11. As expected CPP gets an alpha helical form as secondary conformation. (Peptide naming: P1 (Peptide 1), DP (Designed Peptide), CPP (Cell Penetrating Peptide)).

4.4 Global Docking of Peptides

After the structure prediction, these peptides are docked globally into the receptor chain via Cluspro. In addition to this, the BACH1 BTB domain and BCL- 6 BTB domain are used for negative control. The structure of these domains is shown (Fig. 7A, B, C).

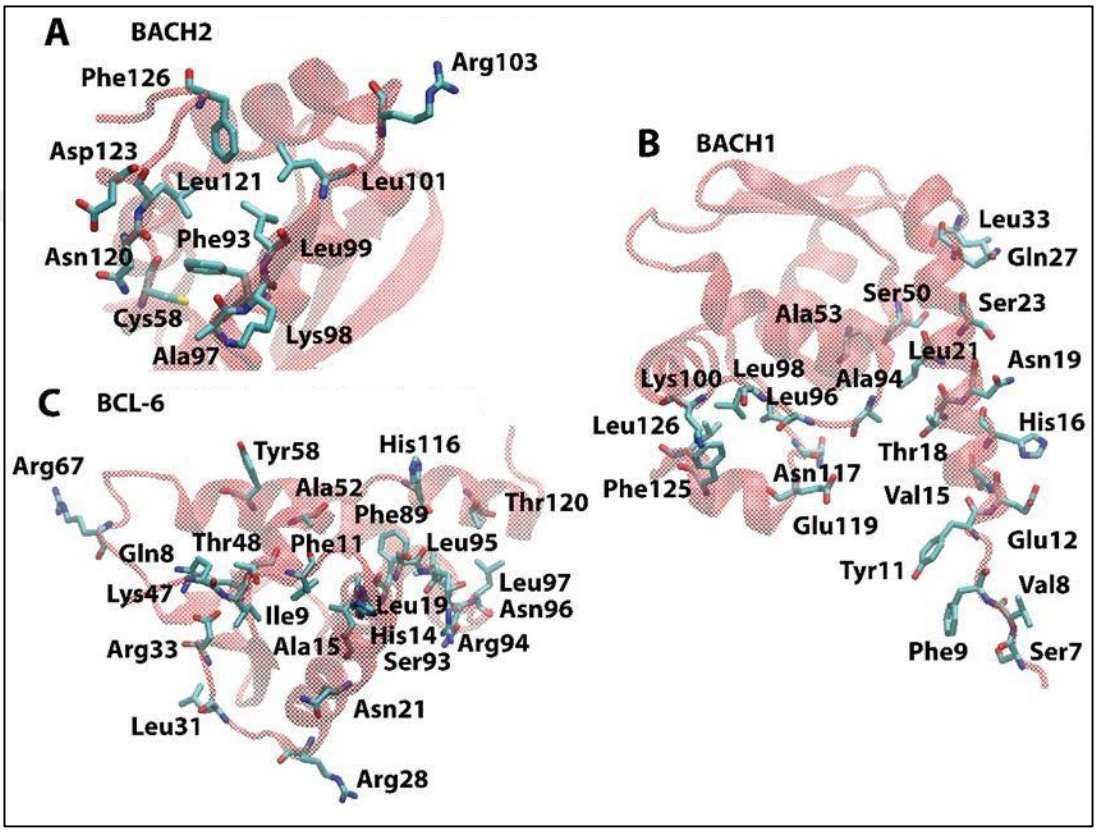

Figure 7: BACH2 BTB domain (A) and potential negative controls BACH1(B) and BCL 6 BTB(C) domains along with Ligplot+ image of amino acid residues that control dimerization. A) Ala97, Lys98, Leu99, Leu100, Leu101, Ser102, Arg103 amino acids are important for dimerization on BACH2 BTB domain. B, C) Amino acid residues for the dimerization of BTB domains on potential negative controls BACH1 and BCL 6 are shown.

4.5 PLIP Results

Peptides are docked onto BACH1 BTB domain, BCL- 6 BTB and BACH2 BTB domain with Cluspro. This global docking is followed by interaction analysis with PLIP (Protein – Ligand Interaction Profiler) (Table 5-23).

4.5.1 Control Peptide – BACH2 BTB domain (3HP (Hydrophobic interaction), 1HB (Hydrogen bond), 1SB (Salt bridge))

Table 5: Hydrophobic interactions and distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Control Peptide 1

Protein AA#	Peptide AA#	Distance (Å)
Ala57	*Tyr17*	*3.79*
Ala57	*Thr20*	*3.45*
Phe93	*Ile26*	*3.48*

Table 6: Hydrogen bonds, its distance and side chain presence between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Control Peptide 1

Protein AA#	Peptide AA#	Distance (Å)	Side Chain Interaction
Lys98	*Asn25*	*1.64*	*Yes*

Table 7: Salt bridge and its distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Control Peptide 1

Protein AA#	Peptide AA#	Distance (Å)
Glu60	*Arg19*	*3.47*

The control peptide forms 3 hydrophobic interactions, 1 hydrogen bond and 1 salt bridge with BACH2 BTB Domain. The residues are selected according to areas that contribute the most to interaction.

4.5.2 Peptide 4 – BACH2 BTB domain (5HB (Hydrogen bond), 1PS (Pi Stacking))

Table 8: Hydrogen bonds, its distance and side chain presence between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 4

Protein AA#	Peptide AA#	Distance (Å)	Side Chain Interaction
His119	*Tyr17*	*3.63*	*Yes*
Glu60	*Thr20*	*2.95*	*Yes*
Arg117	*Val21*	*2.93*	*No*
Arg117	*Val21*	*3.90*	*No*
Arg117	*Val21*	*2.67*	*No*

Table 9: Pi stacking and its distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 4

Protein AA#	Peptide AA#	Distance (Å)
His119	*Tyr17*	*5.26*

Peptide 4 forms 5 hydrogen bonds and 1 pi stacking with BACH2 BTB Domain. The residues are selected according to areas that contribute the most to interaction.

4.5.3 Peptide 12 – BACH2 BTB domain (1HP (Hydrophobic interaction), 2HB (Hydrogen bond), 1SB (Salt Bridge))

Table 10: Hydrophobic interactions and distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 12

Protein AA#	Peptide AA#	Distance (Å)
Ala57	*Tyr17*	*3.94*

Table 11: Hydrogen bonds, its distance and side chain presence between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 12

Protein AA#	Peptide AA#	Distance (Å)	Side Chain Interaction
Arg117	*Arg13*	*2.76*	*Yes*
Arg117	*Arg13*	*2.67*	*Yes*

Table 12: Salt bridge and its distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 12

Protein AA#	Peptide AA#	Distance (Å)
Glu60	*Arg13*	*4.39*

The peptide 12 forms 1 hydrophobic interaction, 2 hydrogen bonds and 1 salt bridge with BACH2 BTB Domain. The residues are selected according to areas that contribute the most to interaction.

4.5.4 Peptide 13 – BACH2 BTB domain (1HB (Hydrophobic interaction), 1SB (Salt Bridge))

Table 13: Hydrophobic interactions and distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 13

Protein AA#	Peptide AA#	Distance (Å)
Ala56	*Tyr17*	*3.90*

Table 14: Salt bridge and its distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 13

Protein AA#	Peptide AA#	Distance (Å)
Glu60	*Arg13*	*3.56*

The peptide 13 forms 1 hydrophobic interaction and 1 salt bridge with BACH2 BTB Domain. The residues are selected according to areas that contribute the most to interaction.

4.5.5 Peptide 14 – BACH2 BTB domain (1HB (Hydrogen bond), 1SB (Salt Bridge))

Table 15: Hydrogen bonds, its distance and side chain presence between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 14

Protein AA#	Peptide AA#	Distance (Å)	Side Chain Interaction
His119	*Arg13*	*2.94*	*Yes*

Table 16: Salt bridge and its distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 14

Protein AA#	Peptide AA#	Distance (Å)
Glu60	*Arg13*	*3.87*

The peptide 14 forms 1 hydrogen bond and 1 salt bridge with BACH2 BTB Domain. The residues are selected according to areas that contribute the most to interaction.

4.5.6 Peptide 15 – BACH2 BTB domain (5HP (Hydrophobic interaction), 4HB (Hydrogen bond))

Table 17: Hydrophobic interactions and distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 15

Protein AA#	Peptide AA#	Distance (Å)
Ala97	*Pro13*	*3.58*
His119	*Ty15*	*3.75*

Table 17: Hydrophobic interactions and distance between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 15 (Cont.)

Glu60	*Tyr17*	*3.43*
Ala57	*Arg25*	*3.75*

Table 18: Hydrogen bonds, its distance and side chain presence between BACH2 BTB Domain (PDB Code: 3OHU Chain A) – Peptide 15

Protein AA#	**Peptide AA#**	**Distance (Å)**	**Side Chain Interaction**
Asn120	*Pro13*	*2.87*	*No*
His119	*Tyr15*	*3.19*	*No*
His119	*Tyr15*	*2.84*	*No*
Glu60	*Tyr17*	*2.82*	*Yes*
Glu60	*Tyr17*	*2.82*	*Yes*

The peptide 15 forms 5 hydrophobic interaction and 5 hydrogen bonds with BACH2 BTB Domain. The residues are selected according to areas that contribute the most to interaction.

4.5.7 Peptide 13 – BACH1 BTB domain (1HP (Hydrophobic interaction), 2HB (Hydrogen bond))

Table 19: Hydrophobic interactions and distance between BACH1 BTB Domain (PDB Code: 2IHC Chain A) – Peptide 13

Protein AA#	**Peptide AA#**	**Distance (Å)**
Val8	*Ile26*	*3.87*

Table 20: Hydrogen bonds, its distance and side chain presence between BACH1 BTB Domain (PDB Code: 2IHC Chain A) – Peptide 13

Protein AA#	Peptide AA#	Distance (Å)	Side Chain Interaction
Arg49	*Thr20*	*2.85*	*No*
Ala10	*Leu27*	*3.73*	*No*

The peptide 13 forms 1 hydrophobic interaction and 2 hydrogen bonds with BACH1 BTB domain. The residues are selected according to areas that contribute the most to interaction.

4.5.8 Peptide 13 – BCL6 BTB domain (3HP (Hydrophobic interaction), 5HB (Hydrogen bond), 1SB (Salt bridge))

Table 21: Hydrophobic interactions and distance between BCL6 BTB domain (PDB Code: 6CQ1 Chain A) –Peptide 13

Protein AA#	Peptide AA#	Distance (Å)
Phe89	*Ile26*	*3.75*
Ile9	*Ile26*	*3.94*
Ile9	*Leu27*	*3.71*

Table 22: Hydrogen bonds, its distance and side chain presence between BCL6 BTB Domain (PDB Code: 2IHC Chain A) – Peptide 13

Protein AA#	Peptide AA#	Distance (Å)	Side Chain Interaction
Asp6	*Arg13*	*3.83*	*Yes*
His116	*Tyr15*	*2.78*	*Yes*
Arg94	*Glu19*	*2.76*	*No*
His116	*Thr24*	*3.87*	*Yes*
Thr120	*Arg25*	*2.82*	*Yes*

Table 23: Salt bridge and its distance between BCL6 BTB domain (PDB Code: 6CQ1 Chain A) – Peptide 13

Protein AA#	Peptide AA#	Distance (Å)
Asp6	*Arg13*	*3.18*

4.6 Local Docking with HADDOCK

The PLIP is followed by local docking by HADDOCK. The peptide 13 along with control peptide are docked onto BACH2 BTB domain (Fig. 8).

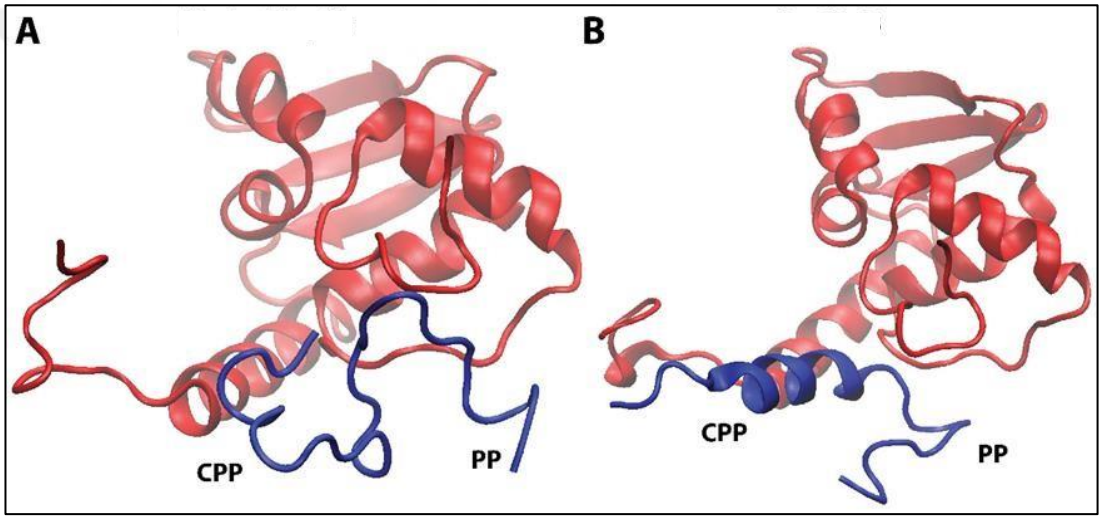

Figure 8: HADDOCK local docking results of BACH2 BTB domain with control peptide (A) and peptide 13 (B).

4.7 Molecular Dynamics Analysis and Results

Local Docking is continued with MD analysis for peptides 1, 4, 12, 13, 14 and 15 with BACH2, BACH1 and BCL6. These simulations are repeated 4 times for 500 ns (Fig. 9-17). The positive control can reach stability throughout all simulations (Fig. 9). Same simulations are also applied for control peptide (Fig. 10), peptide 4 (Fig. 11), peptide 12 (Fig. 12), peptide 13 (Fig. 13), peptide 14 (Fig. 16) and peptide 15 (Fig. 17). Control peptide reaches stability after 20th ns (<3 Å) in 2 simulations, while reaches the same stability after 250th ns in other two simulations (Fig. 10). The peptide

4 stabilizes the complex after 50th ns in all simulations (Fig. 11), while this stability cannot be seen in peptide 12 anyhow (Fig. 12). Peptides 13 and 15 show low RMSD results (< 3 Å) throughout simulations (Fig. 13 and Fig. 17). The same simulations are also used for peptide 13 – BACH 1 BTB domain and BCL 6 BTB domain (Fig. 14 and Fig. 15). They show low RMSD values, which are interpreted as the structure is stable.

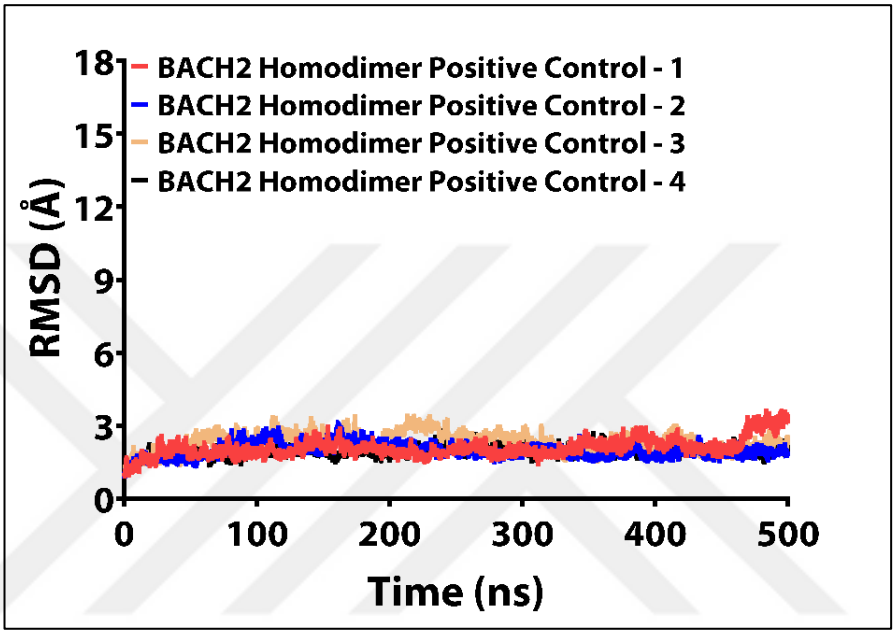

Figure 9: RMSD analysis of BACH2 BTB domain homodimer for 500 ns with molecular dynamics simulations (<3 Å and states as stable in all 4 simulations)

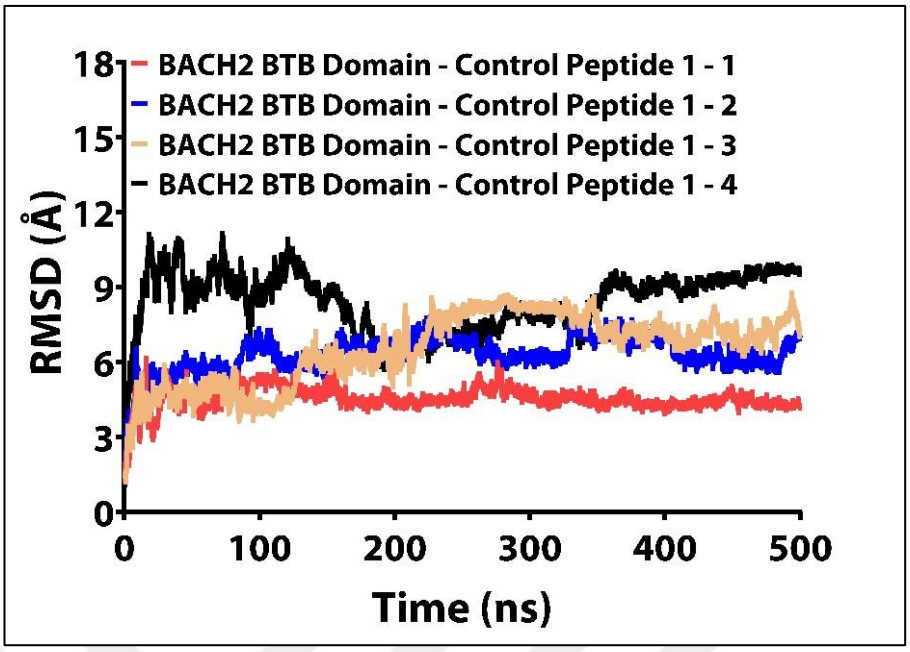

Figure 10: RMSD analysis of BACH2 BTB domain–control peptide (peptide 1) for 500 ns with molecular dynamics simulations (<3 Å in the first 20 ns of 2 simulations, reaches stability in other 2 simulations after 250 ns)

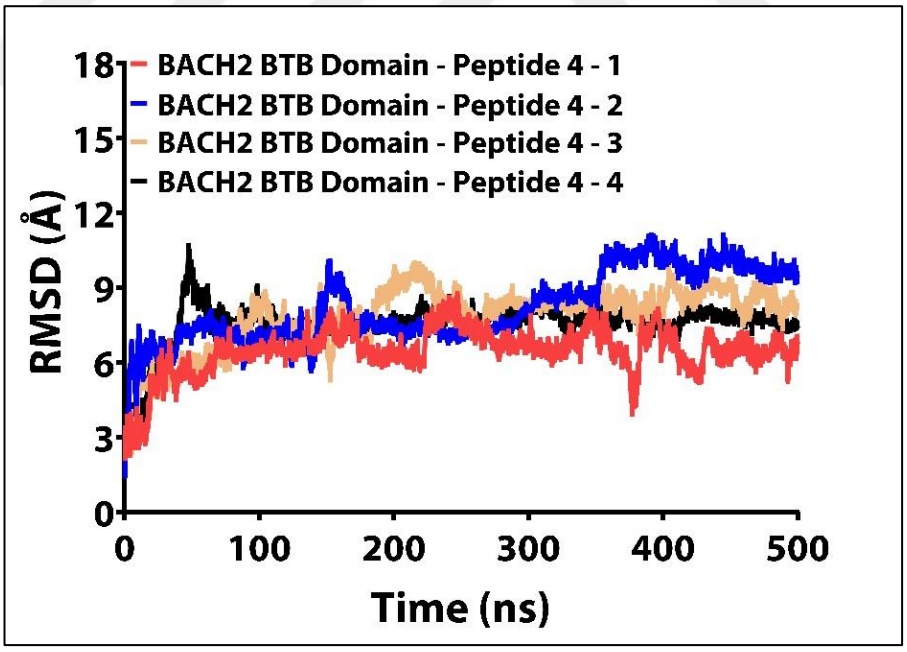

Figure 11: RMSD analysis of BACH2 BTB domain–peptide 4 for 500 ns with molecular dynamics simulations (<3 Å after 50 ns in all simulations)

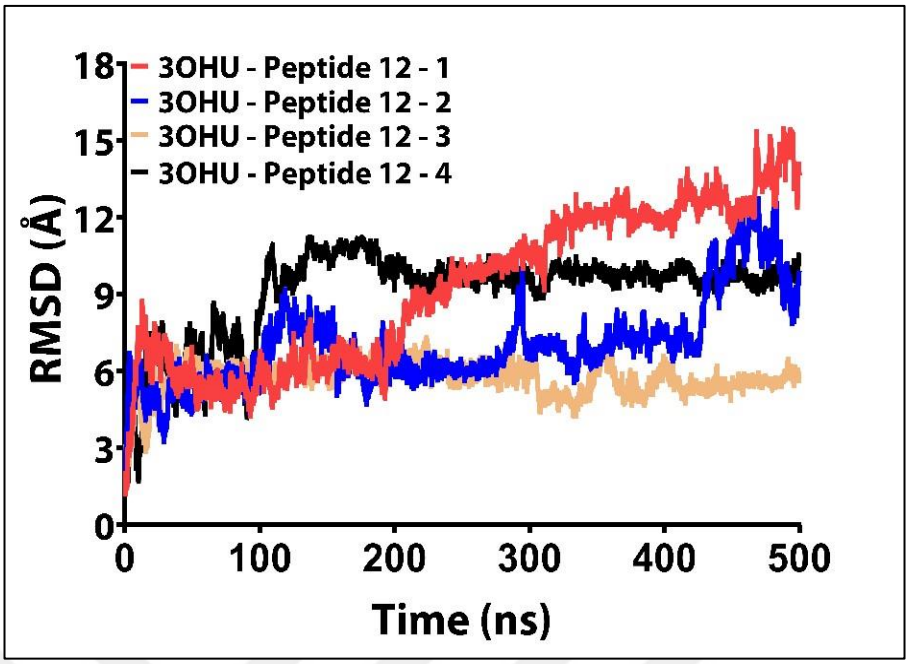

Figure 12: RMSD analysis of BACH2 BTB domain–peptide 12 for 500 ns with molecular dynamics simulations (<3 Å through first simulation, loses stability after 450th ns in second simulation and stable in other 2 simulations).

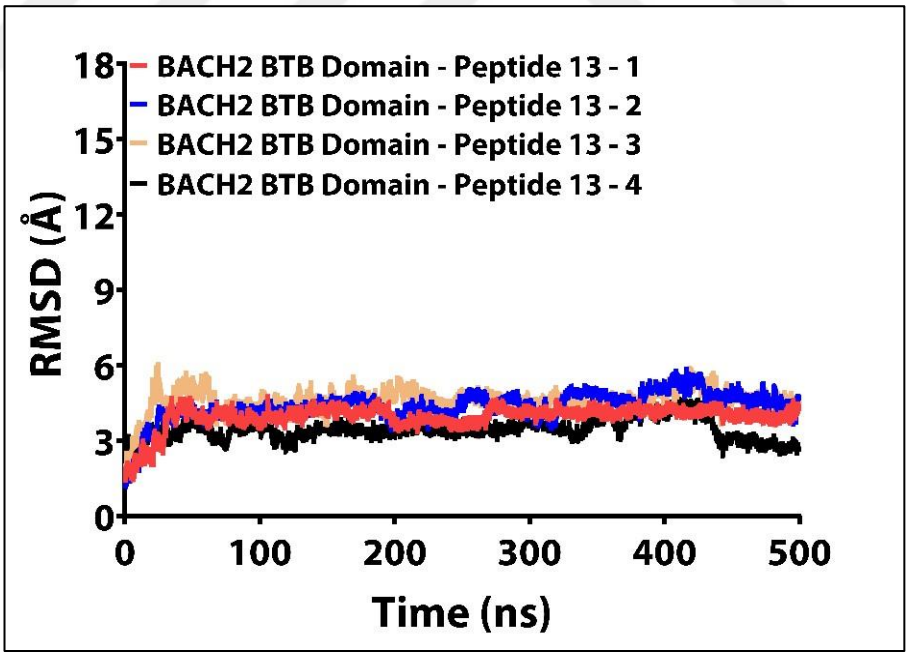

Figure 13: RMSD analysis of BACH2 BTB domain–peptide 13 for 500 ns with molecular dynamics simulations (<3 Å in all simulations).

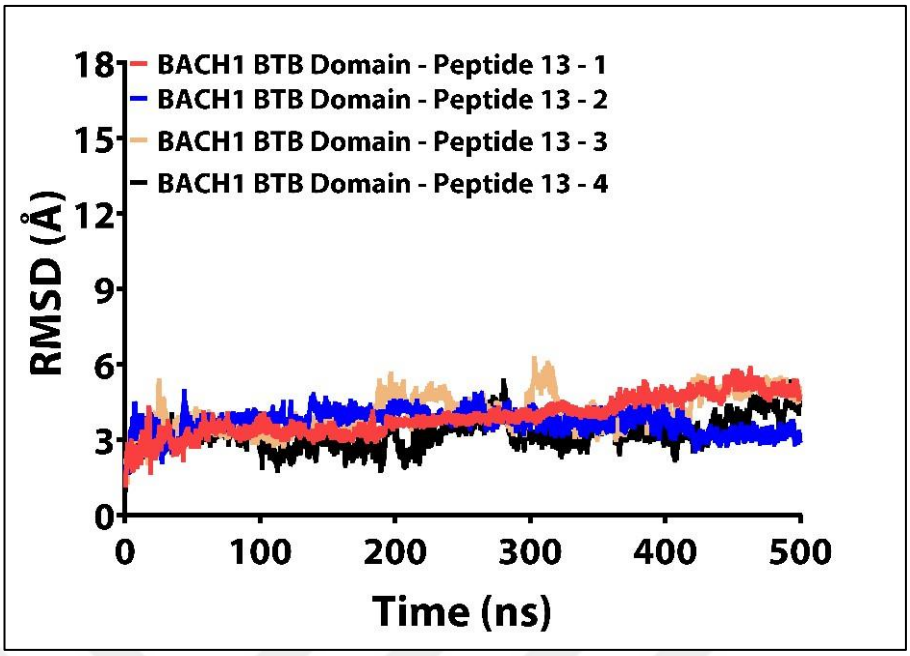

Figure 14: RMSD analysis of BACH1 BTB domain–peptide 13 for 500 ns with molecular dynamics simulations (<3 Å in all simulations).

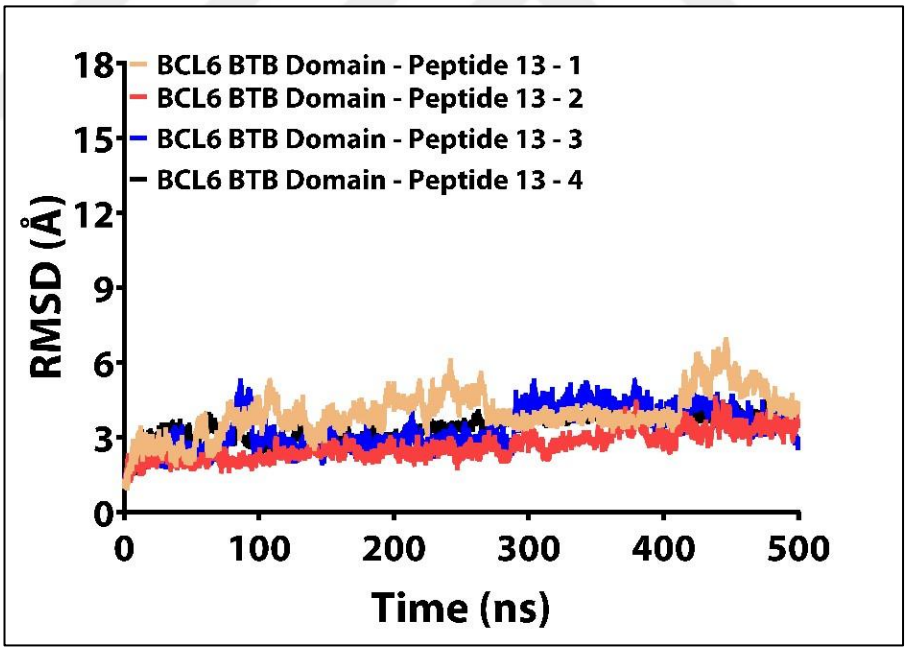

Figure 15: RMSD analysis of BCL6 BTB domain–peptide 13 for 500 ns with molecular dynamics simulations (<3 Å in all simulations)

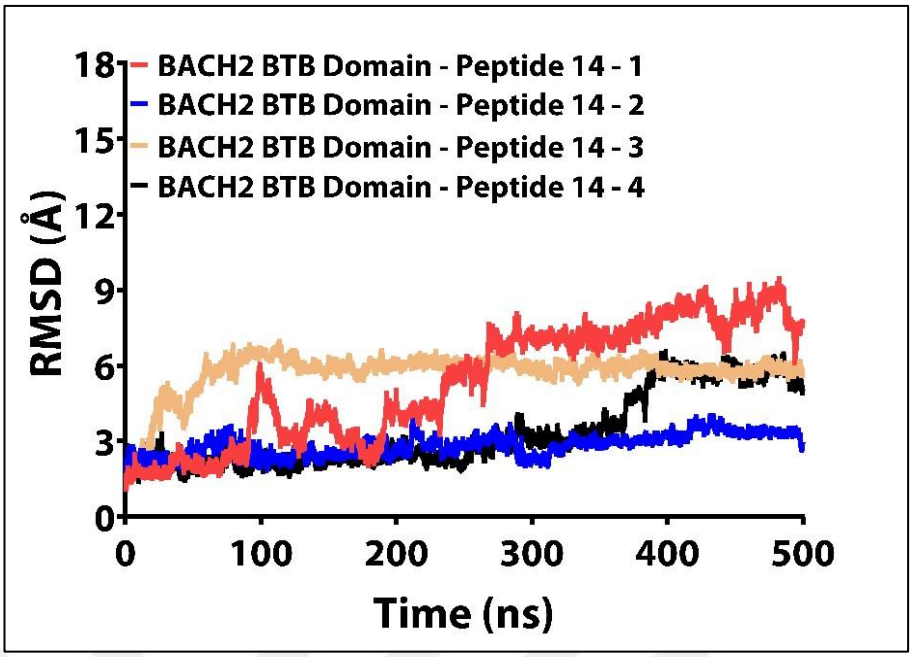

Figure 16: RMSD analysis of BACH2 BTB domain–peptide 14 for 500 ns with molecular dynamics simulations (<3 Å after 250ns in the first simulation, same for other simulations after 50th ns).

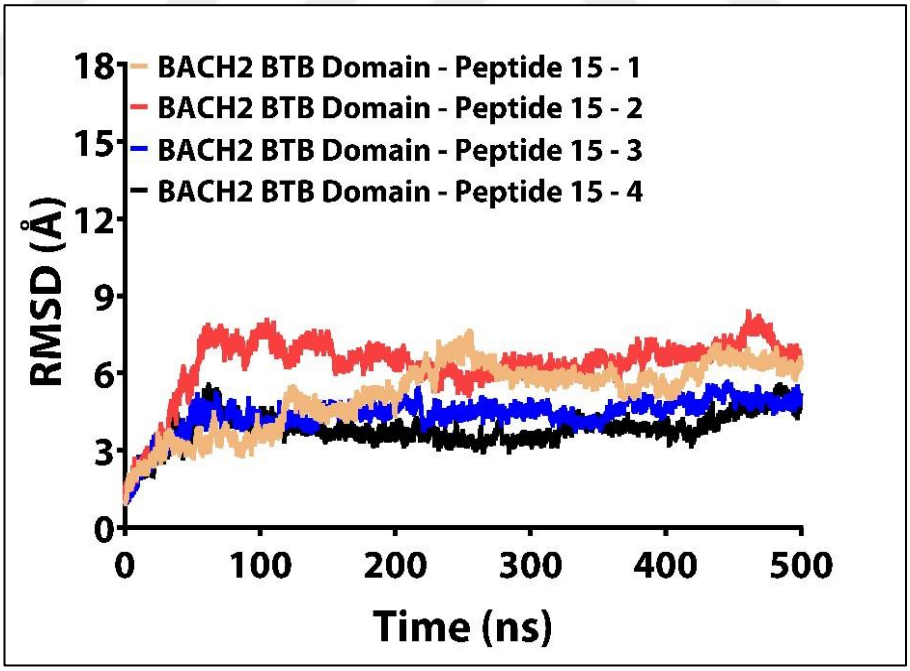

Figure 17: RMSD analysis of BACH2 BTB domain–peptide 15 for 500 ns with molecular dynamics simulations (<3 Å after 450 ns in all simulations).

4.8 MM-PBSA Results

Relative free binding energy (RBFE) is predicted after RMSD results (Fig. 18). The RBFE results show the difference between the binding affinities of the peptide and the complex. The RBFE score decreases the most with peptide 13 compared to other peptides and positive complex (Fig. 18).

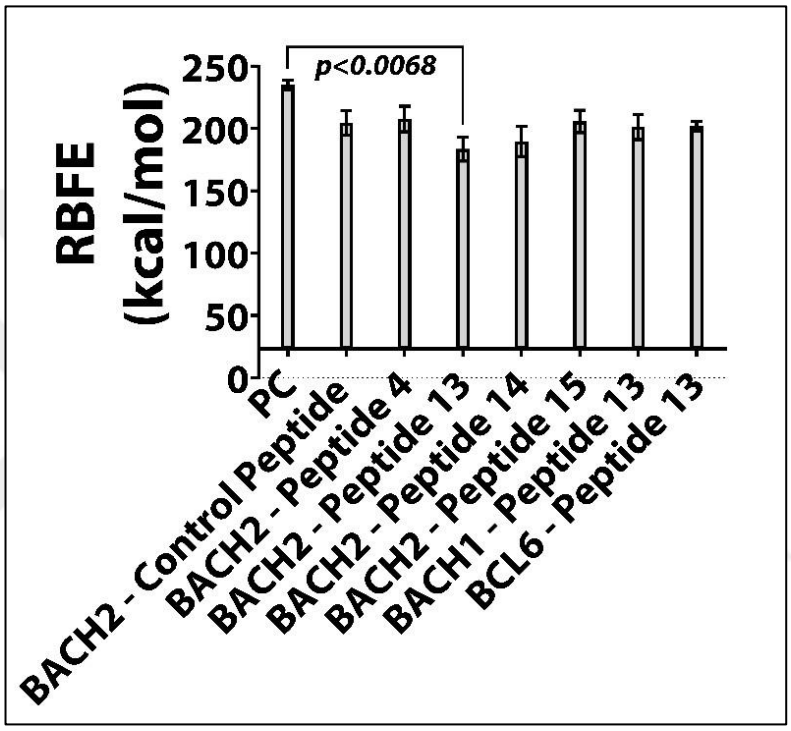

Figure 18: Relative free binding energy results of complexes obtained with MM-PBSA method. (Lowest with peptide 13)

Amino acid based RBFE contribution of positive control homodimer, BACH2 BTB domain monomer – peptide 1 and BACH2 BTB domain monomer – peptide 13 for chain A are shown (Fig. 19). Residues that contribute the most for positive control are MET11, TYR12, VAL13, TYR14, ILE23, LYS98 and ASN 120 on chain A.

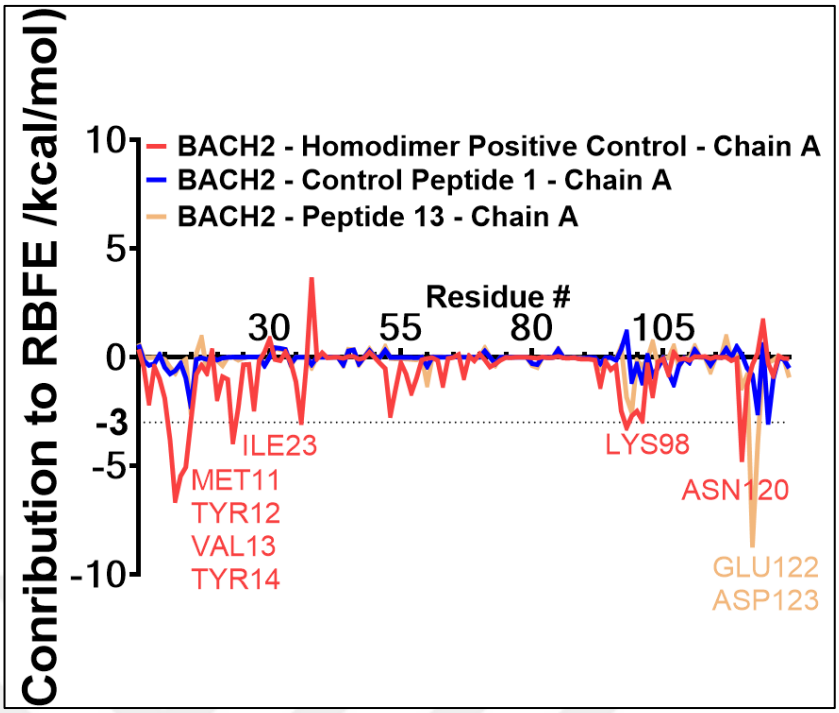

Figure 19: Amino acid-based contribution to relative free binding energy. (RBFE of BACH2 BTB domain homodimer positive control, complex with peptide 1 and complex with peptide 13)

Amino acid based RBFE contribution of positive control homodimer for chain B is shown (Fig. 20). Residues that contribute the most for positive control are TYR12, TYR14, LEU36, LYS98 and LEU 101 on chain B.

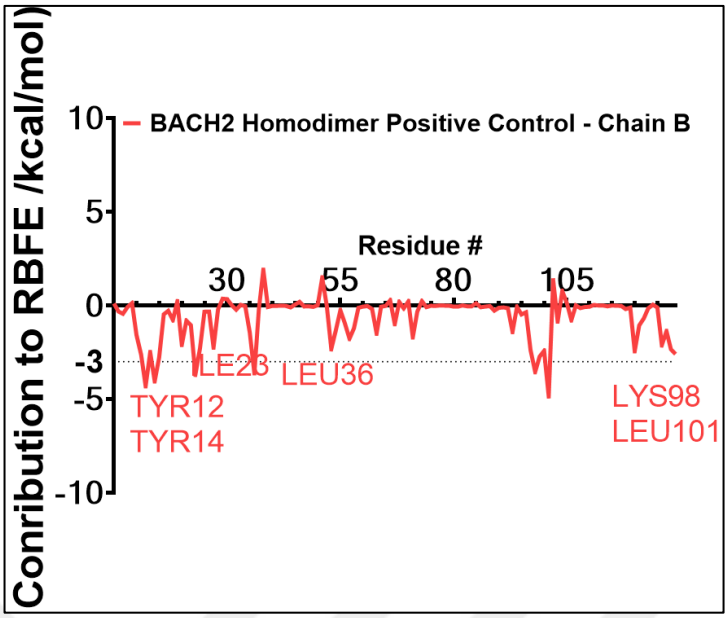

Figure 20: Amino acid-based contribution to relative free binding energy. RBFE of BACH2 BTB domain homodimer positive control.

Amino acid based RBFE contribution of peptide 1 and peptide 13 is shown (Fig. 21). These residues are ARG6, ARG9, ARG10 on CPP, while ARG13 and VAL16 on designed peptide.

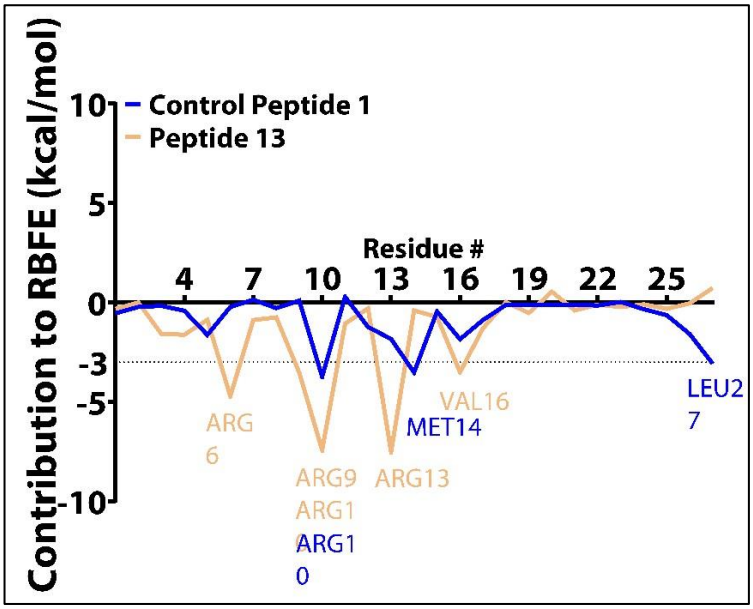

Figure 21: Amino acid-based contribution to relative free binding energy. (RBFE of BACH2 BTB domain monomer-peptide 1 / peptide 13)

The indicated area with a black line is the predicted wildtype peptide area and its target (Fig. 22). Apart from that, peptide 1 (control/wildtype peptide) interacts with the structure at most with residues Arg10, Met14, and Leu27. A similar pattern with higher affinity is seen with peptide 12 by using residues Arg6, Arg9, Arg10, Arg13, and Val16. In the control peptide, an amino acid with interaction scores lower than 3kcal/mol cannot be found, yet peptide 13 interacts with Glu122 and Asp123 in the BACH2 BTB domain lower than this score (Fig. 22).

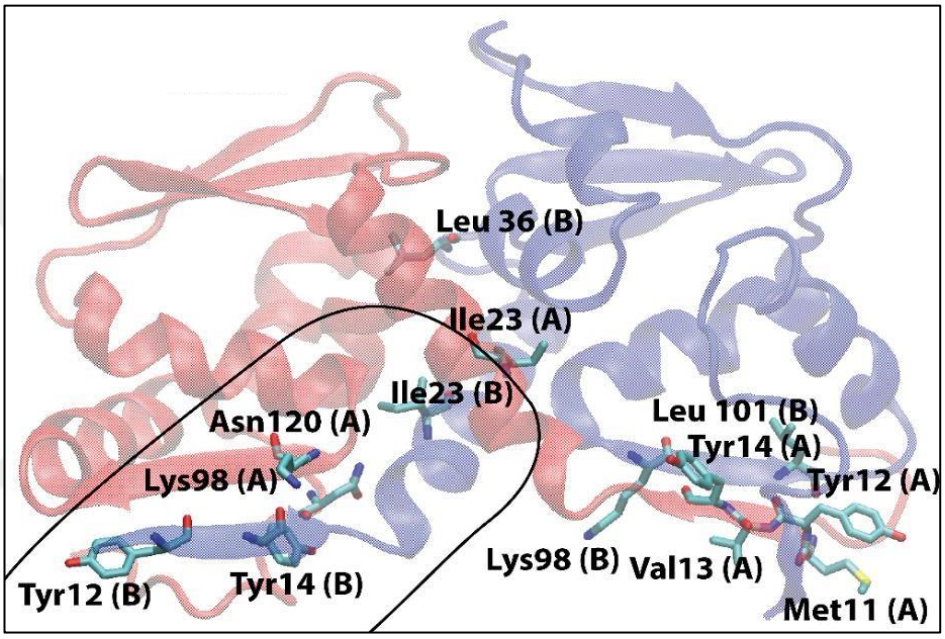

Figure 22: Amino acid residues represent the most to relative free binding energy in positive control complex simulations. The residues are obtained via MM-PBSA method. The area inside the black line is the position where the peptide is located.

The procedure is followed by hydrogen bond analysis. At this stage, uniquely formed hydrogen bonds of proteins or peptide side chains are sorted according to their occupancy ratio (Table 24). In positive control, hydrogen bonds between LYS98 - GLU15, ASN22 - ASN120 and TYR14 - GLU122 are parallel to results that are obtained from MM-PBSA based amino acid contribution (Fig. 20). In peptide 1 – BACH2 BTB-domain-complex, hydrogen bonds between ARG10 - GLU15 is observed, which is also parallel to MM-PBSA based amino acid contribution results (Fig. 21). Finally, in complex with peptide 13, it is observed that peptide 13 uses ARG6, ARG9, ARG10, ARG13 and VAL16 as amino acid contribution, while these

residues are GLU122 and ASP123 for the BACH2 BTB domain monomer (Fig. 21). When these residues are taken account of, peptide 13 targets BACH2 BTB domain dimerization with hydrogen bonds: ARG10 - ASP123, ARG13 - GLU122, ARG9 - GLU122, TYR17 - ASP123 and ARG13 - ASN120. In Table 24, the hydrogen bonds between residues are indicated in bold if they are also observed in MM-PBSA-based amino acid contribution.

Table 24: The representation of hydrogen bonds and the repletion, between BACH2 BTB domain homodimer (PC), BACH2 BTB domain monomer-control peptide (peptide 1), BACH2 BTB domain monomer-peptide 13 complexes and with their side chains.

Donor	Receptor	Repletion (%)	Donor	Receptor	Repletion (%)	Donor	Receptor	Repletion (%)
PC-1			Pep1-1			Pep13-1		
LYS98-Side-NZ	**GLU15-Side-OE2**	53.20%	**ARG10-SideNH2 (P)#**	**GLU15-Side-OE2**	52.80%	**ARG10-Side-NH1 (P)**	**ASP123-SideOD1**	39.28%
ASN22-Side-ND2	**ASN120-SideOD1**	52.80%	**ARG10-Side-NE (P)**	**GLU15-Side-OE1**	49.92%	**ARG10-Side-NH2 (P)**	**ASP123-SideOD2**	37.20%
LYS98-Side-NZ	**GLU15-Side-OE1**	48.32%	LYS5Side-NZ (P)	ASP123-SideOD2	28.88%	**ARG10-Side-NH1 (P)**	**ASP123-SideOD2**	32.00%
ASN120-SideND2	HSE19-SideND1	20.80%	LYS5Side-NZ (P)	ASP123-SideOD1	24.88%	**ARG10-Side-NH2 (P)**	**ASP123-SideOD1**	30.48%
PC-2			Pep1-2			Pep13-2		
ASN22-Side-ND2	**ASN120-SideOD1**	57.44%	SER124-Side-OG	THR24-SideOG1 (P)	33.44%	**ARG13-Side-NH2 (P)**	**GLU122-Side-OE2**	31.92%
LYS98-Side-NZ	**GLU15-Side-OE1**	49.68%	**ARG10-Side-NE (P)**	**GLU15-Side-OE2**	13.52%	**ARG13-SideNE (P)**	**GLU122-Side-OE1**	24.56%
LYS98-Side-NZ	**GLU15-Side-OE2**	47.20%	**ARG10-Side-NE (P)**	**GLU15-Side-OE1**	12.88%	LYS98-Side-NZ	GLU18Side-OE2 (P)	20.80%

Table 24: The representation of hydrogen bonds and the repletion, between BACH2 BTB domain homodimer (PC), BACH2 BTB domain monomer-control peptide (peptide 1), BACH2 BTB domain monomer-peptide 13 complexes and with their side chains. (Cont.)

TYR14-Side-OH	**GLU122-Side-OE1**	13.04%	TYR15Side-OH (P)	GLU15-Side-OE2	11.84%	LYS98-Side-NZ	GLU18Side-OE1 (P)	19.20%
PC-3			Pep1-3			Pep13-3		
LYS98-Side-NZ	**GLU15-Side-OE2**	48.08%	LYS5Side-NZ (P)	GLU15-Side-OE1	14.16%	**ARG13-SideNE (P)**	**ASP123-SideOD2**	37.04%
LYS98-Side-NZ	**GLU15-Side-OE1**	46.96%	**ARG7-Side-NE (P)**	**GLU15-Side-OE2**	13.44%	TYR15Side-OH (P)	THR21-SideOG1	37.28%
ASN22-Side-ND2	**ASN120-SideOD1**	46.96%	**ARG7-Side-NH2 (P)**	**GLU15-Side-OE1**	11.76%	**ARG9-Side-NH1 (P)**	**GLU122-Side-OE1**	36.40%
ASN120-SideND2	HSE19-SideND1	26.00%	LYS5Side-NZ (P)	GLU15-Side-OE2	9.12%	THR17-SideOG1	THR24-SideOG1 (P)	36.00%
PC-4			Pep1-4			Pep13-4		
ASN22-Side-ND2	**ASN120-SideOD1**	58.40%	LYS5Side-NZ (P)	ASP123-SideOD2	14.08%	**ARG10-Side-NH1 (P)**	**ASP123-SideOD2**	44.00%
LYS98-Side-NZ	**GLU15-Side-OE2**	56.56%	LYS5Side-NZ (P)	ASP123-SideOD1	13.04%	**TYR17-SideOH (P)**	**ASP123-SideOD2**	38.80%
LYS98-Side-NZ	**GLU15-Side-OE1**	52.32%	THR20-SideOG1 (P)	GLU15-Side-OE1	6.24%	**ARG10-Side-NH1 (P)**	**ASP123-SideOD1**	30.08%
ASN120-SideND2	HSE19-SideND1	33.92%	THR20-SideOG1 (P)	GLU15-Side-OE2	5.92%	**ARG13-Side-NH1 (P)**	**ASN120-SideOD1**	28.96%

The average number of hydrogen bonds is calculated between peptide 13 and the BACH2 BTB domain through simulation courses (Fig. 23). Hydrogen bonds of amino acid side chains in peptide 13 that keep the structure intact are pointed out. It is observed that peptide 13 forms 5 hydrogen bonds, while peptide 1 forms 1 hydrogen bond and positive control forms 3 hydrogen bonds on average.

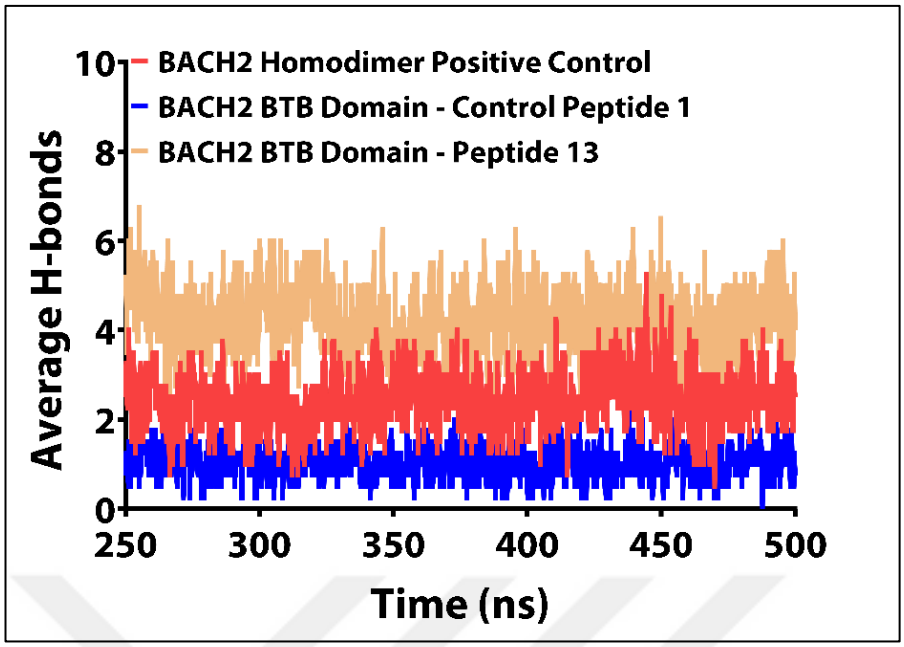

Figure 23: The average hydrogen bond amount between complexes and their side chains for the last 250 ns. All the hydrogen bonds are shown per frame. The average is taken after 4 replicates. (The average hydrogen bond amounts: Peptide ~ 5, Positive Control ~ 3, Control Peptide ~ 1).

In addition to hydrogen bond analysis, salt bridge distance is also analyzed (Fig. 24). Salt bridges between structures are observed compared to amino acid residues that contribute the most to MM-PBSA (Fig. 21). In this regard, the salt bridge between LYS98 (BACH2 BTB Domain Chain A) and GLU 15 (BACH2 BTB Domain chain B) is observed in every analysis for positive control. These residues are also essential due to results from hydrogen bond analysis, which makes the LYS98 - GLU15 bond an important element for affinity. TYR14 on chain B is one of the amino acids that contribute the most to RBFE and GLU15 in its $+1^{st}$ position is added into salt bridge analysis. In the complex with peptide 1, a salt bridge between LYS98 - GLU18 is detected, which is accepted as equivalent to LYS98 - GLU15. However, in one of the replicants, this salt bridge in LYS98 - GLU18 shows a long distance, and it is determined as a low contribution to affinity. For peptide 13, LYS98 - GLU15 is observed in 2 replicants, and the distance is measured approximately the same with the positive control complex. Furthermore, GLU19 – LYS98 forms another salt bridge in every replicant, yet its distance is longer than the one in the positive control.

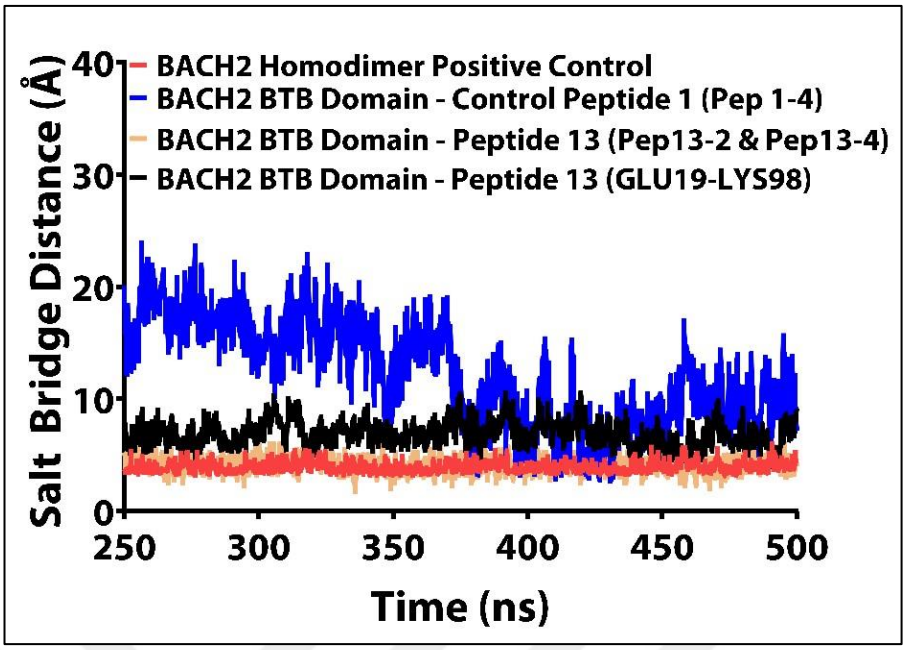

Figure 24: The distance of salt bridge between LYS98-GLU15 on positive control and similar salt bridges between similar residues for other complexes. The salt bridge analysis is applied with VMD. The distance is taken between the center of masses.

MM-PBSA analysis is continued with carbon–carbon contact analysis. The potential hydrophobic interaction is shown by considering the last 250 ns of the simulation (Table 25). In the analysis, a 50% repletion ratio is stated as the minimum limit. Firstly, in the positive control complex, MET11, TYR12, VAL 13, TYR14, LYS98, and ASN 120 on chain A form a carbon–carbon interaction with TYR12, TYR14, LEU36, LYS9,8 and LEU101 on chain B. In 90% of the last 250 ns simulations, it is observed that TYR12 - LEU101, TYR14 - LYS98, LYS98 - TYR14, and ASN120 - TYR14 interactions maintain the hydrophobic interactions. In the case of carbon – carbon interaction analysis for peptide 1, ARG10 – MET11 shows carbon – carbon interaction more than 50% occupancy twice, and LYS98 - MET14 shows %99.4 occupancy once. These residues are also taken from MM-PBSA-based amino acid contribution results. In the complex containing peptide 13, carbon-carbon contacts are observed for ASN120 - ARG13 (3 times), TYR12 - ARG9 (once), ASN120 - VAL16 (3 times), and LYS98 - VAL16 (4 times) with more than 80% occupancy.

Table 25: Carbon-carbon contacts inside the monomer-peptide complexes with lower than 5 Å distance for the last 250ns. ND (Non-Data): Data cannot be taken. 50% repletion ratio is the criteria for carbon – carbon contact. These contacts are shown in bold.

PC-1		**Repletion (%)**	**Pep1-1**		**Repletion (%)**	**Pep13-1**		**Repletion (%)**
TYR12	**LEU101**	100	**MET11**	**ARG10**	**54.2**	**ASN120**	**ARG13**	**100**
TYR14	**LYS98**	95.3	TYR12	ARG10	7.8	**TYR12**	**ARG9**	**100**
LYS98	**TYR14**	93.4	TYR14	ARG10	7.8	**ASN120**	**VAL16**	**99.4**
ASN120	**TYR14**	85.7	VAL13	ARG10	0.2	**LYS98**	**VAL16**	**90.8**
PC-2			**Pep1-2**			**Pep13-2**		
TYR12	**LEU101**	99.9	**LYS98**	**MET14**	**90.6**	**LYS98**	**VAL16**	**99.4**
ASN120	**TYR14**	98.8	MET11	MET14	4.9	**ASN120**	**ARG13**	**83.1**
TYR14	**LYS98**	95.8	ND	ND	ND	ASN120	VAL16	21.8
LYS98	**TYR14**	93.9	ND	ND	ND	TYR12	VAL16	2.9
PC-3			**Pep1-3**			**Pep13-3**		
TYR12	**LEU101**	99.8	**VAL13**	**ARG10**	**99.4**	**ASN120**	**VAL16**	**93.1**
TYR14	**LYS98**	97.1	ASN120	MET14	39.8	**LYS98**	**VAL16**	**79.3**
ASN120	**TYR14**	97.1	LYS98	MET14	5.2	TYR12	ARG9	25
LYS98	**TYR14**	89.4	TYR12	ARG10	2.2	ASN120	ARG13	7.1
PC-4			**Pep1-4**			**Pep13-4**		
LYS98	**TYR14**	93	**MET11**	**MET14**	**56**	**ASN120**	**ARG13**	**99.7**
TYR12	**LEU101**	99.7	TYR14	MET14	43.5	**ASN120**	**VAL16**	**99**
ASN120	**TYR14**	95.5	LYS98	LEU27	30.3	**LYS98**	**VAL16**	**96.5**
TYR14	**LYS98**	94.9	ASN120	MET14	8.7	TYR14	ARG9	0.5

The procedure is followed by the backbone alpha carbon flexibility analysis (Fig. 25). Firstly, alpha carbon flexibility on chain A of targeted BACH2 BTB domain monomer is analyzed with RMSF values for positive complex, complex with peptide 1 and 13. RMSF value is used for measuring atom flexibility, therefore low values are considered as positive results. In positive control and the complex with peptide 13, carbon alpha flexibilities on the N and C terminal are measured as equal, yet in peptide 1 these values are higher (Fig. 25).

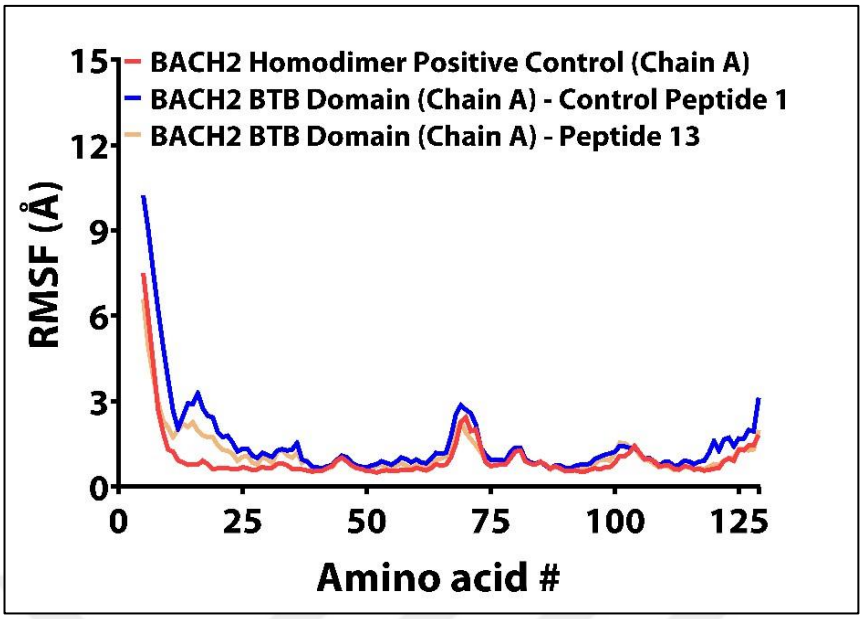

Figure 25: Alpha carbon flexibility of complexes in terms of RMSF Higher flexibility is observed in N and C terminus (lower flexibility with peptide 13).

RMSF results of carbon alphas in peptides can be seen. The flexibility of these carbon alphas is checked (Fig. 26). In the complex with peptide 13, rigidity is observed same for the peptide as it is observed in domain chain. Especially between 6 and 19th residues of peptide 13 show low RMSF values and low flexibility.

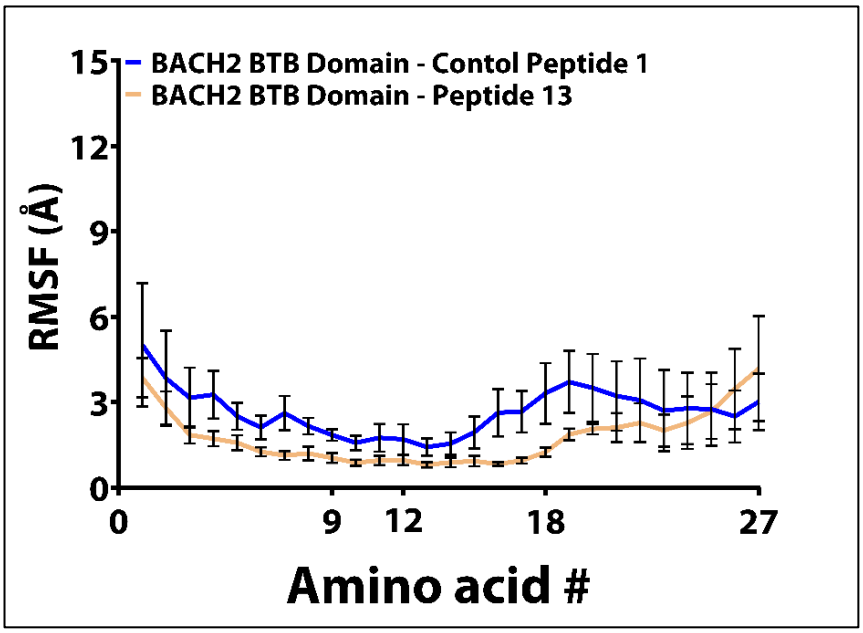

Figure 26: Alpha carbon flexibility of cell penetrating peptide in complexes in terms of RMSF. Residues between 6 and 19 show higher flexibility with peptide 1.

To understand the relation between backbone carbon alpha rigidity and interaction energy, distance analysis is applied (Fig. 27). Decrease in distance between two molecules is accepted as binding. At this point in positive control, the distance is low compared to the complex with peptide 1. In addition to that, residues that contribute the most to the affinity between peptide 13 and BACH2 BTB domain monomer have the lowest distance, which is interpreted as high affinity.

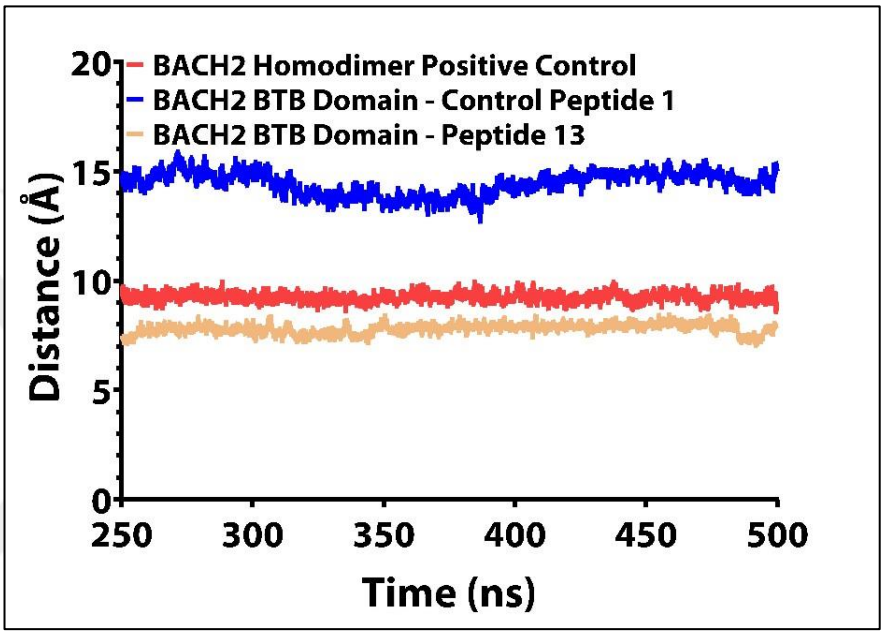

Figure 27: Distance of amino acids that contribute to the interaction of the complex. (Lowest in complex with peptide 13).

The last distance analysis is verified by using SASA (solvent accessible surface area) analysis (Fig. 28). The decrease in SASA value is interpreted as an indication of binding. This analysis shows that the complex with peptide 13 mimics the positive control complex in terms of SASA-based compactness for the last 250 ns. On the other hand, the SASA value remains the same for the complex with peptide 1 and it is observed with a higher SASA value compared to the other two structures.

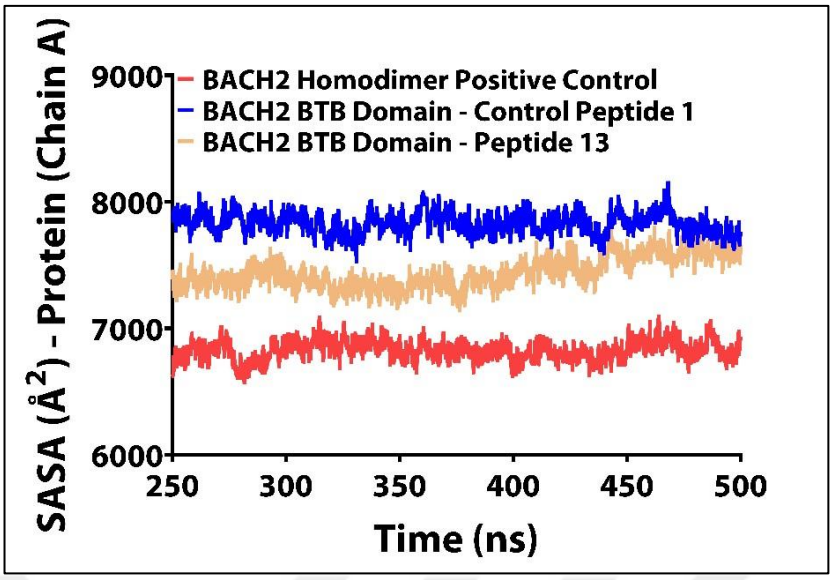

Figure 28: Solvent accessible surface area (SASA) analysis of the complexes. (Higher the SASA value, lower the interaction area).

The flexibility of all atoms of ARG9 in the cell-penetrating peptide and VAL16 in the designed peptide is calculated with MM-PBSA (Fig. 29, 30). These residues are chosen due to their contribution to RBFE. The results show that ARG9 significantly stays rigid in complex with peptide 13, while VAL16 stays rigid almost significantly.

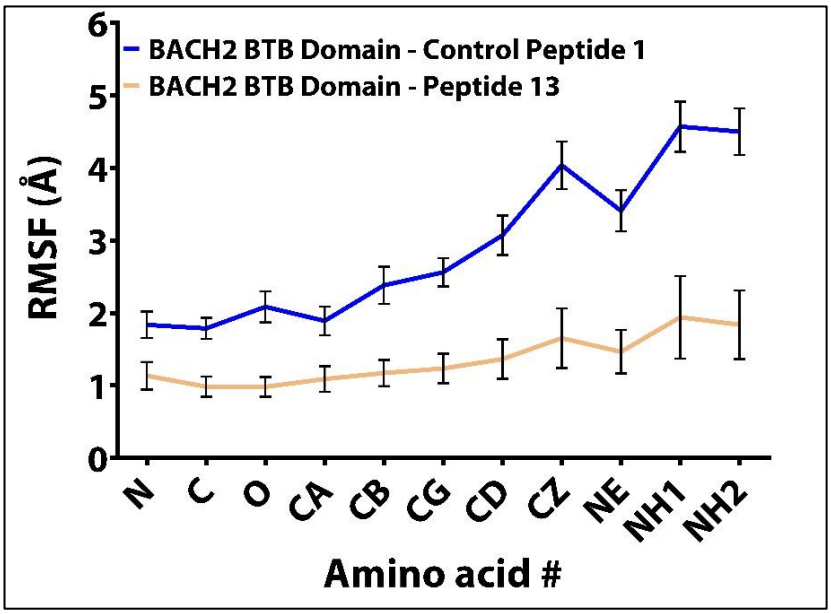

Figure 29: RMSF based flexibility analysis of Arg9 for all its atoms.

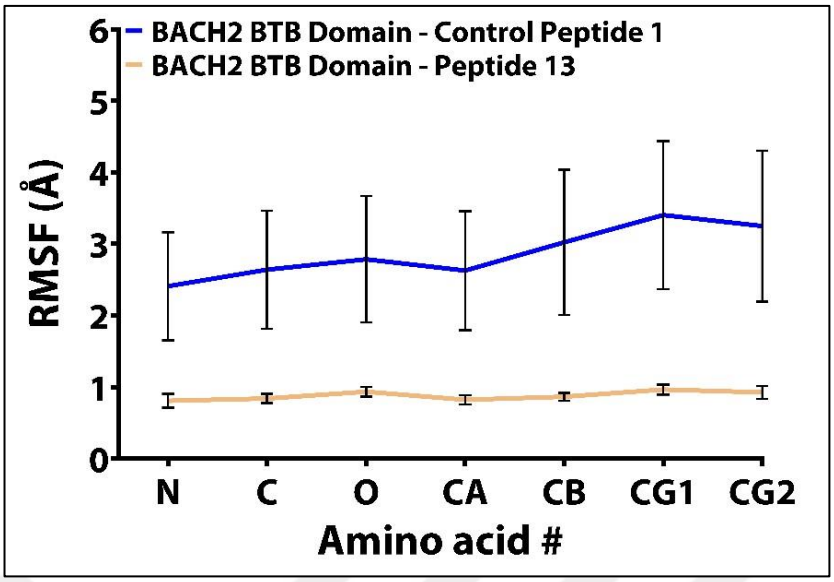

Figure 30: RMSF based flexibility analysis of Val16 for all of its atoms.

4.9 Eigenvector Index Results

The principal component analysis (PCA) is applied to determine major movements/shifts of atom groups. The movements of positive control BACH2 homodimer are shown with eigenvectors and their eigenvalue (Fig. 31, 32). The flexibility of peptide monomer complexes along with positive control is checked with another analysis which is the eigenvector index. Eigenvectors show the major movements of structure (> 2 Angstroms) through simulations. As it is calculated, the complex with peptide 13 shows a lower variance ratio (Fig. 31) compared to the complex with peptide 1 (Fig. 33), which is interpreted that peptide 13 provides more rigidity than peptide 1. Furthermore, PCA analysis is done by projecting these major movements (scatter plots) (Fig. 32, 34, 36). The peptide 13 - BACH2 BTB complex shows a similar pattern with positive control and scatters a lower level of conformational space compared to the complex with peptide 1.

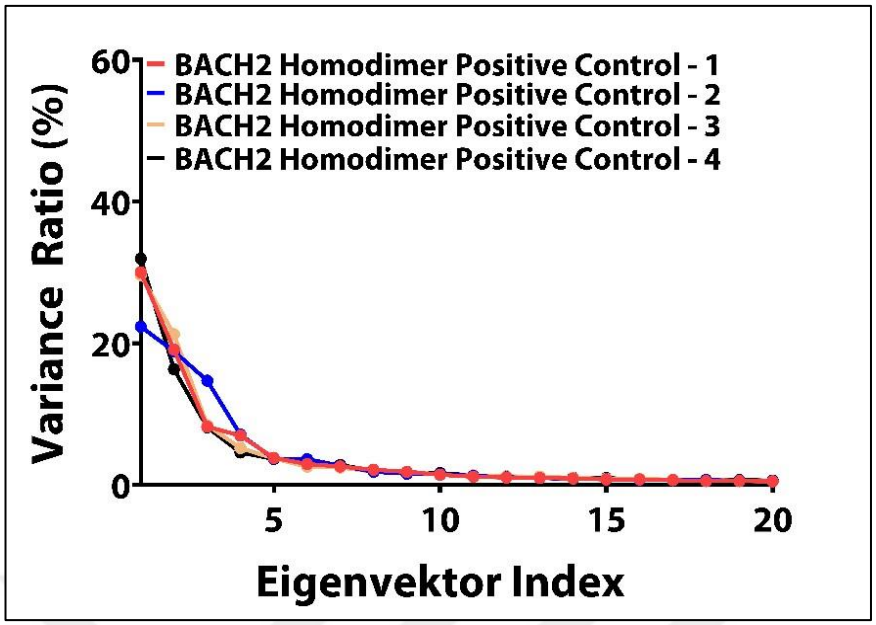

Figure 31: Scree plot of variance ratio to eigenvector index of positive control.

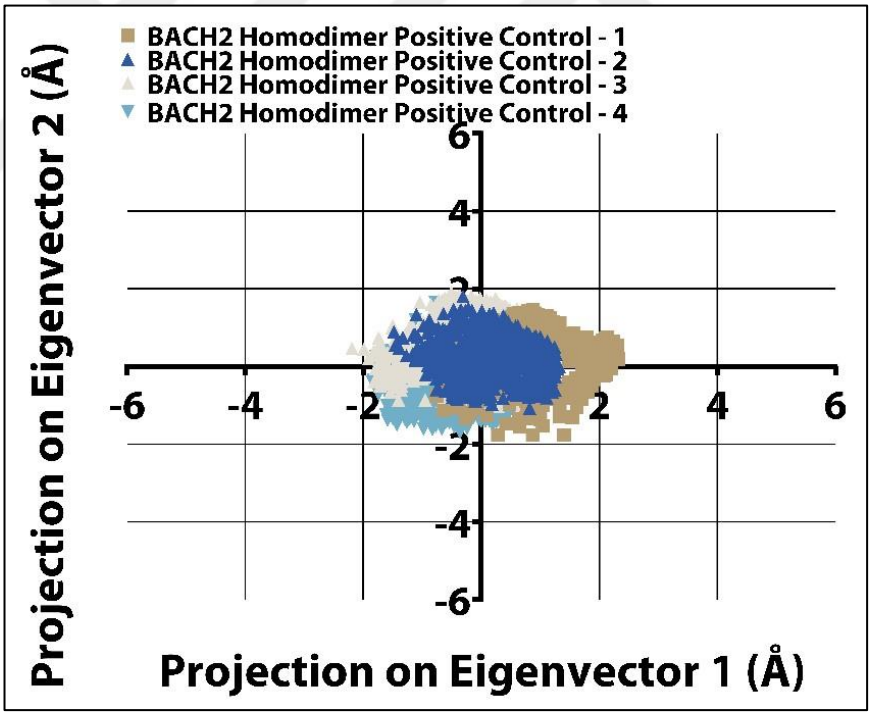

Figure 32: Scatter plot projection of trajectory on principal component analysis for positive control.

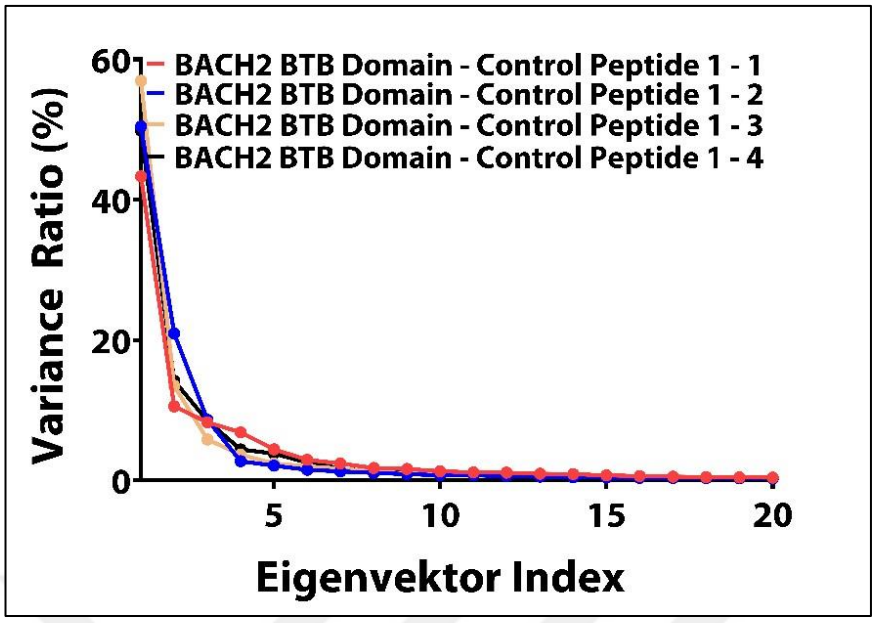

Figure 33: Scree plot of variance ratio to eigenvector index of complex with peptide 1.

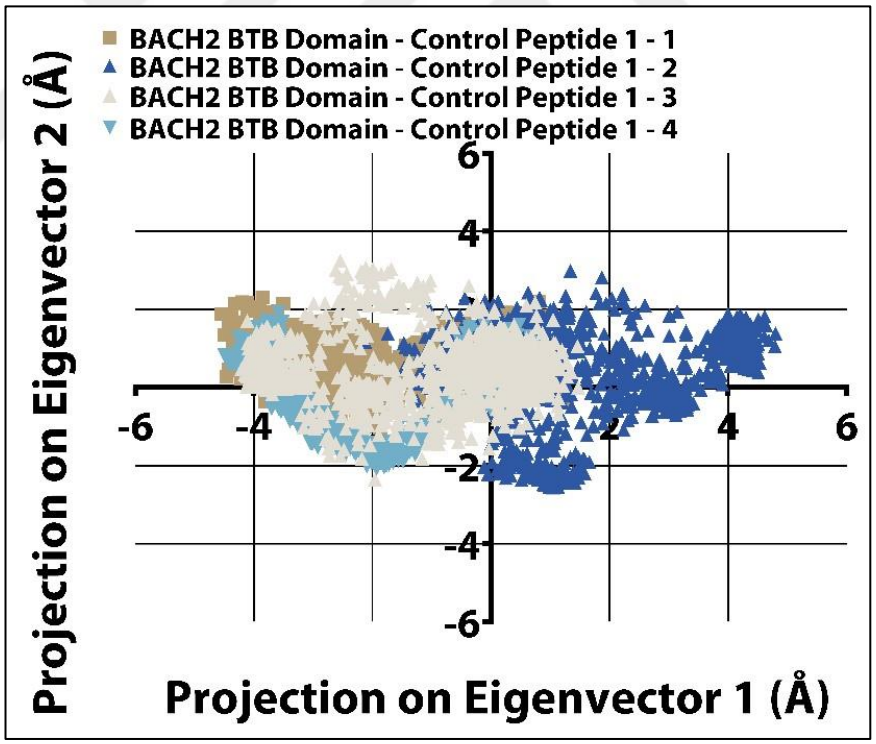

Figure 34: Scatter plot projection of trajectory on principal component analysis for peptide 1.

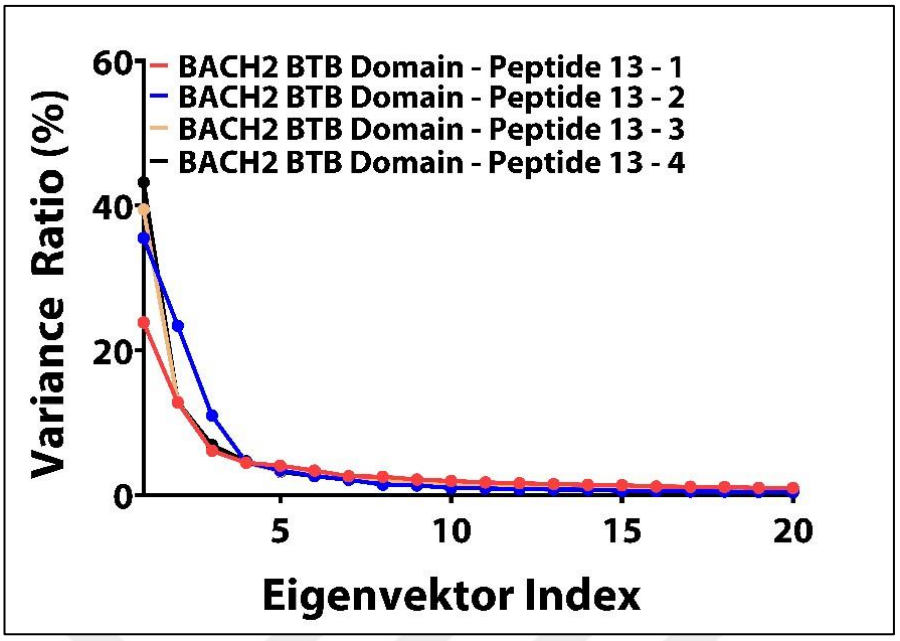

Figure 35: Scree plot of variance ratio to eigenvector index of complex with peptide 13.

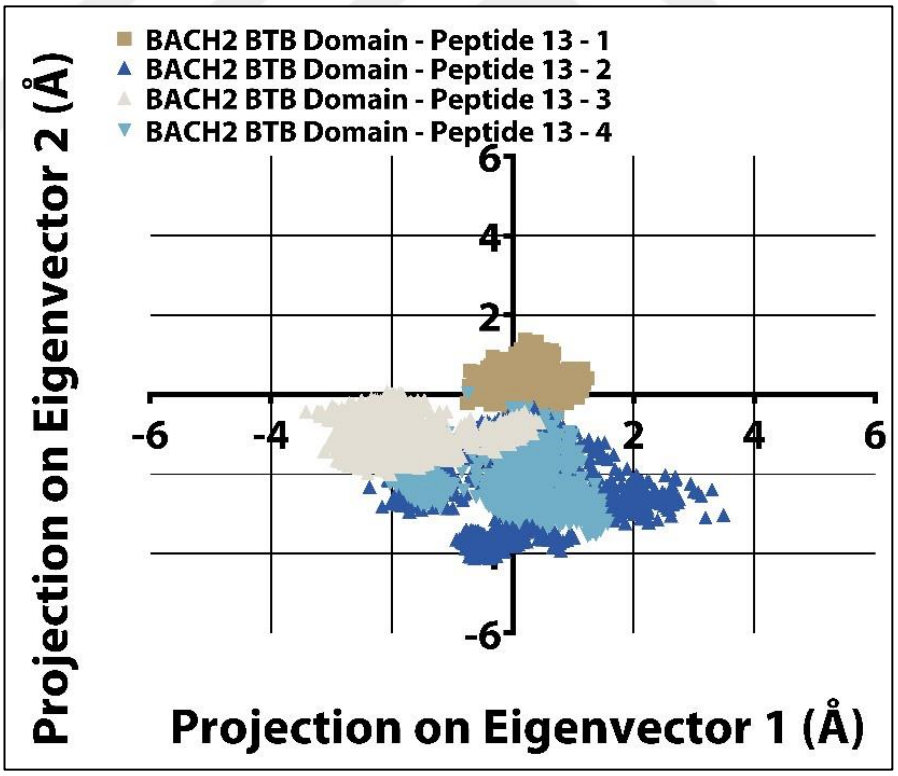

Figure 36: Scatter plot projection of trajectory on principal component analysis for peptide 13.

The alpha carbons of amino acids that contribute the most for eigenvectors 1 and 2 are also analyzed. To confirm the pattern that is seen for confirmational space this mobility assay is applied. N terminus of the BACH2 BTB domain is mostly mobile when it contains peptide 1 (Fig. 39, 40) compared to the other two complexes (Fig. 37, 38, 41, and 42). This shows that the complex with peptide 13 shows a similar pattern to the positive complex in terms of mobility analysis.

The mobility of these alpha carbons in positive control is pointed out (Fig. 37, 38). It is mostly rigid in all simulations. In terminus N it showed a small mobility.

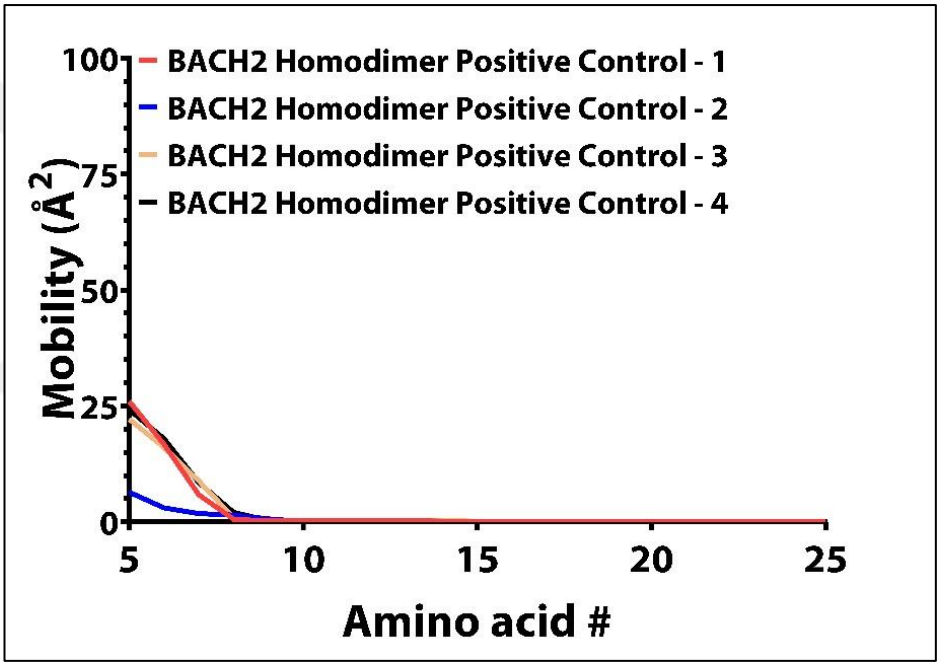

Figure 37: Mobility analysis of amino acids on positive control that contribute the most to the eigenvector 1.

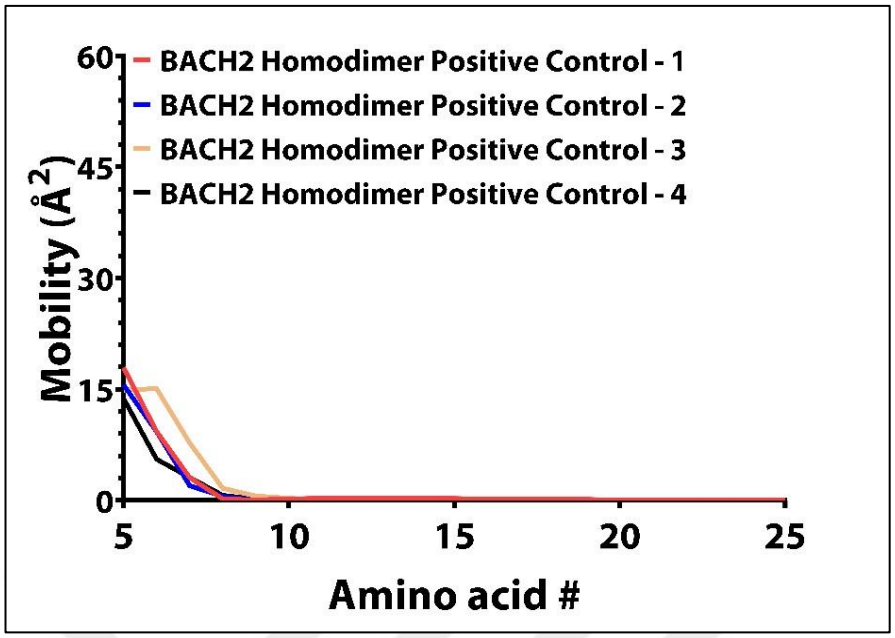

Figure 38: Mobility analysis of amino acids on positive control that contributes the most to the eigenvector 2.

The mobility of alpha carbons in BACH2 BTB domain – peptide 1 is pointed out (Fig. 39, 40). The structure is mobile in all simulations and shows low rigidity on nearly all residues.

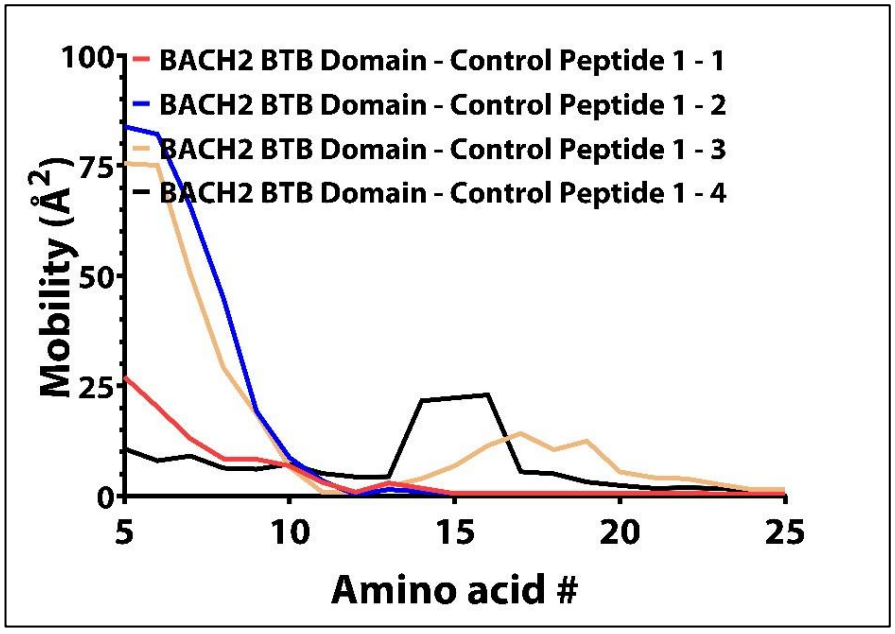

Figure 39: Mobility analysis of amino acids on complex with peptide 1 that contribute the most to the eigenvector 1.

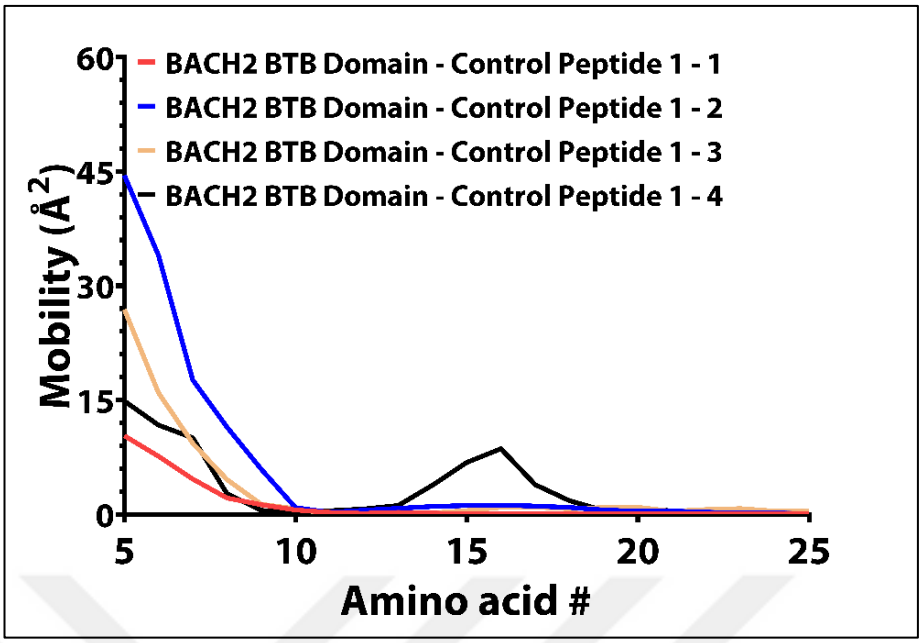

Figure 40: Mobility analysis of amino acids on complex with peptide 1 that contribute the most to the eigenvector 2.

The mobility of alpha carbons in BACH2 BTB domain – peptide 13 is shown (Fig. 41, 42). The structure is rigid and has similar pattern with positive control complex.

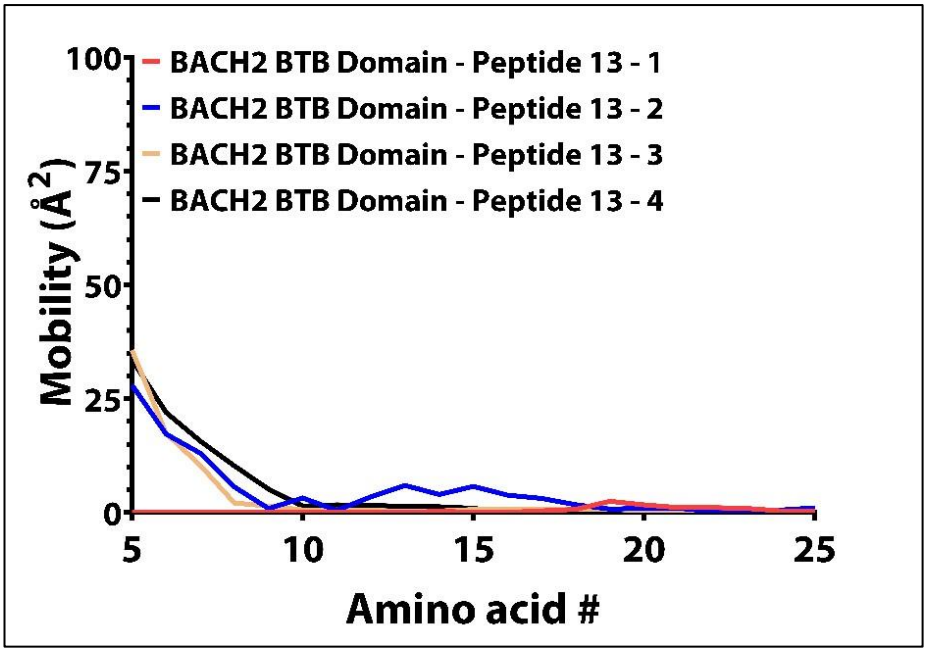

Figure 41: Mobility analysis of amino acids on complex with peptide 13 that contribute the most to the eigenvector 1.

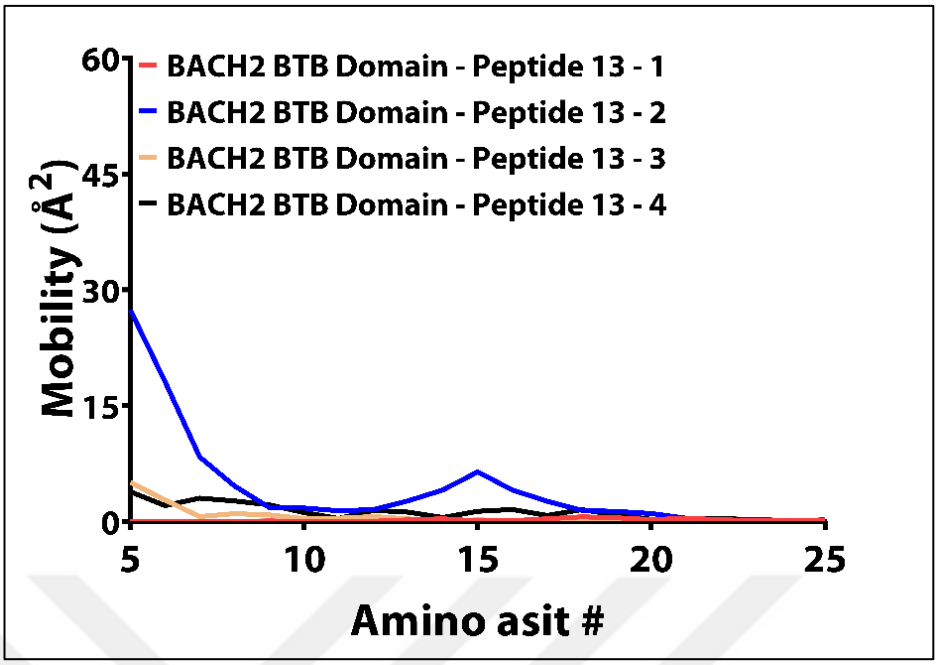

Figure 42: Mobility analysis of amino acids on complex with peptide 13 that contribute the most to the eigenvector 2.

The alpha carbon movements in the positive control, BACH2 BTB domain – peptide 1 and BACH2 BTB domain – peptide 13 along with structure demonstrated to understand the movements in the N terminus (Fig. 43). At this point eigenvectors have similar motions. It is observed that movement is limited with only the amino acids on the edge of the N terminus for positive control. The complex with peptide 13 on the other hand moves with similar patterns to positive control, but because it is a small molecule compared to the domain monomer, it can interact with the backbone of the monomer and stay stable. This shows the stabilization of peptide 13 via different interaction webs. Furthermore, residues between predicted region from positive control (97 – 103) and residues that contribute the most to affinity for peptide 13 (120 – 129) structure stay rigid. Although, this situation is different for peptide 1. More mobile amino acids on the N terminus are observed, it also approaches to backbone but cannot dampen its motion as well as peptide 13. In addition to that, it undergoes conformational change between 97 – 103rd and 120 – 129th positions.

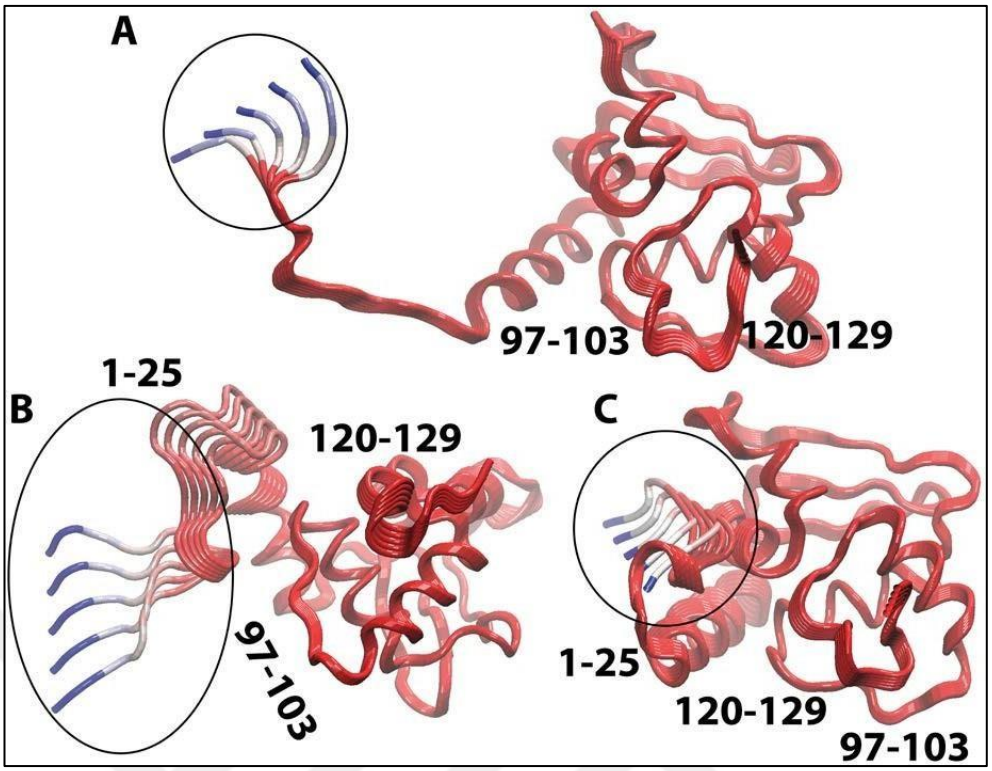

Figure 43: First eigenvector motions in complexes (A: positive control, B: BACH2 BTB domain monomer-peptide 1, C: BACH2 BTB domain monomer-peptide 13)

Amino acid movements in the BACH2 BTB domain are determined by cross-correlation of amino acid alpha carbons. The cross-correlation analysis of positive control complex, complex with peptides 1 and 13 are shown (Fig. 44, 45, 46). As a result, indicated areas with numbers 1, 2, and 3 are residues after a few amino acids on terminus N and they are positively correlated for positive control. This situation can also be observed for the complex with peptide 13 but not with peptide 1. This shows that movement on the N terminus for peptide 1 cannot be synchronized with the protein backbone, but the complex with peptide 13 has a similar correlation and pattern with positive control. This means that it can provide N terminus–backbone synchronization.

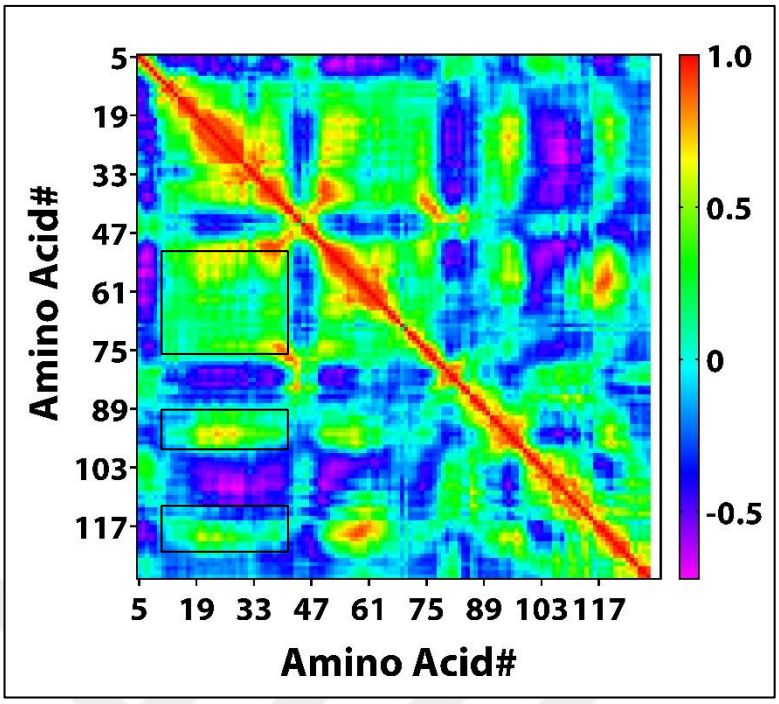

Figure 44: Cross correlation analysis of the movements on residues of positive complex.

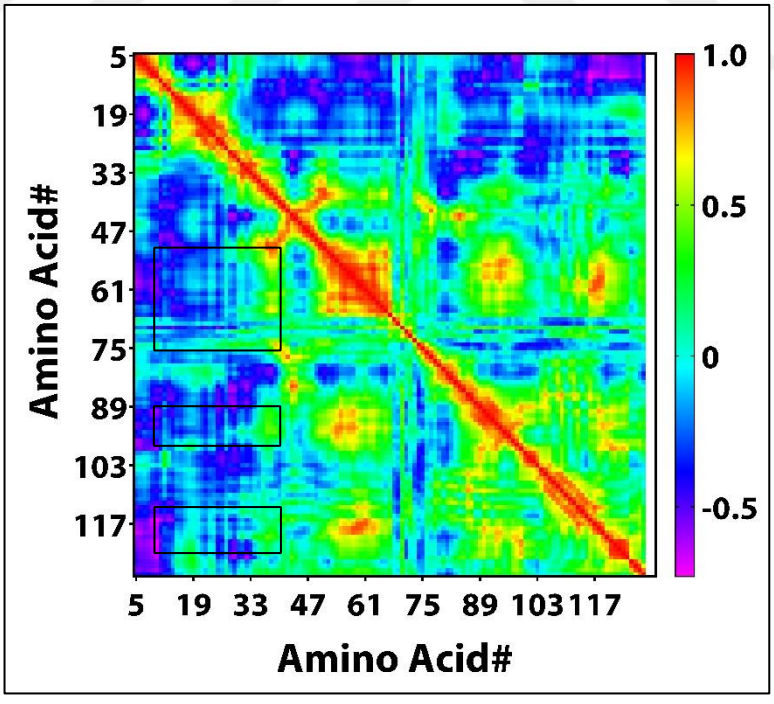

Figure 45: Cross correlation analysis of the movements on residues of complex with peptide 1.

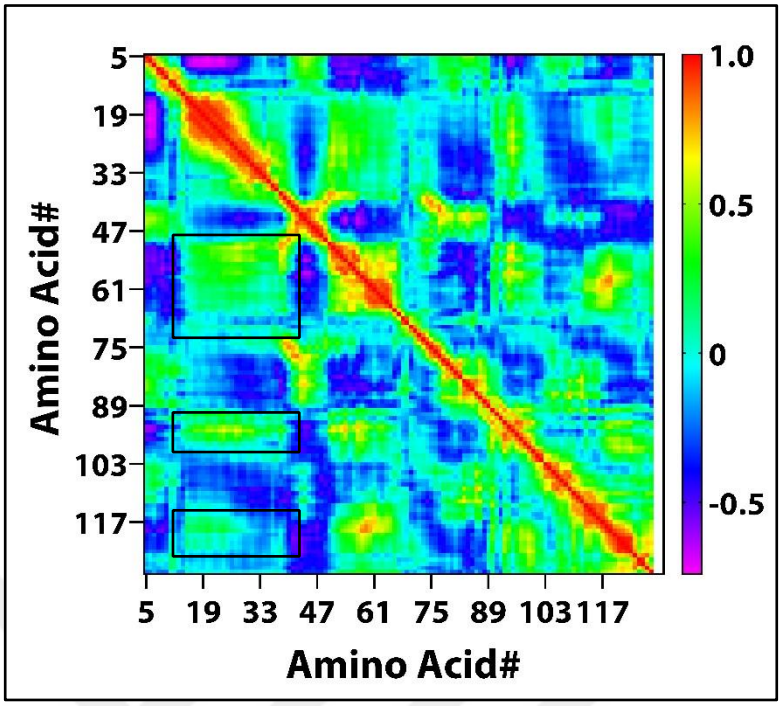

Figure 46: Cross correlation analysis of the movements on residues of complex with peptide 13.

In silico analysis and simulations are followed by the in vitro methods. The Hek293T cells are cultured, and their 80% confluency can be observed (Fig. 47). The overexpression of myc-tagged BACH1 and BACH2 in HEK293T cells is controlled with immunoblotting procedure (Fig. 48).

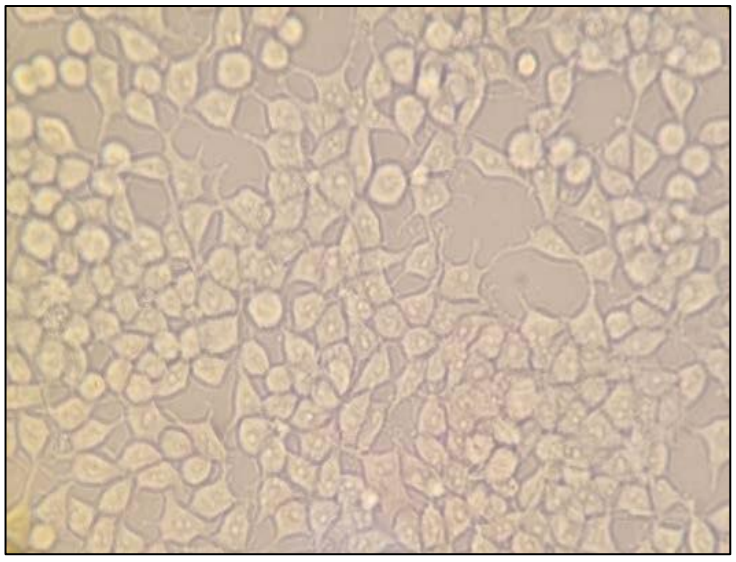

Figure 47: Image of HEK293T cells under 40x microscope with 80% confluency.

4.10 Overexpression of BACH2

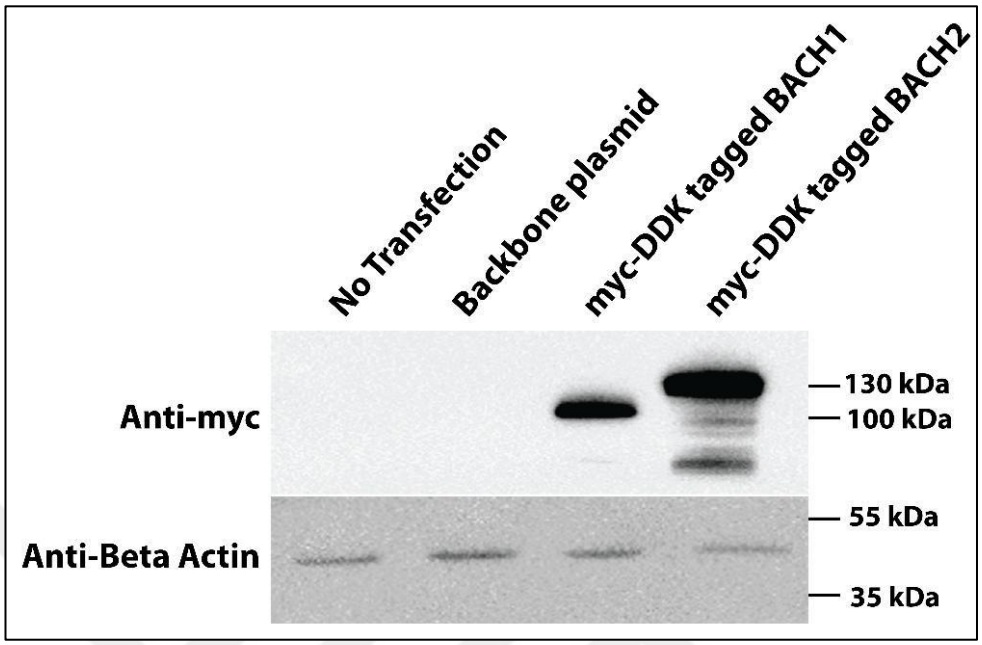

Figure 48: Immunoblotting of Myc-DDK tagged BACH1 and BACH2 overexpression in HEK293T cells.

The overexpression is continued with an optimization experiment with western blot. (Fig. 49). The concentrations are measured to be 5.7 mg/ml, 7.4 mg/ml and 8.3 mg/ml for BACH1, backbone, and BACH2 respectively. This is later on followed by an optimization experiment for the protein amount to be given (Fig. 49). As it is seen, the protein can be seen with 2.5 μg yet it fades with 1 μg. According to the optimization assay, 2.3 μg is seen as optimal for this experiment.

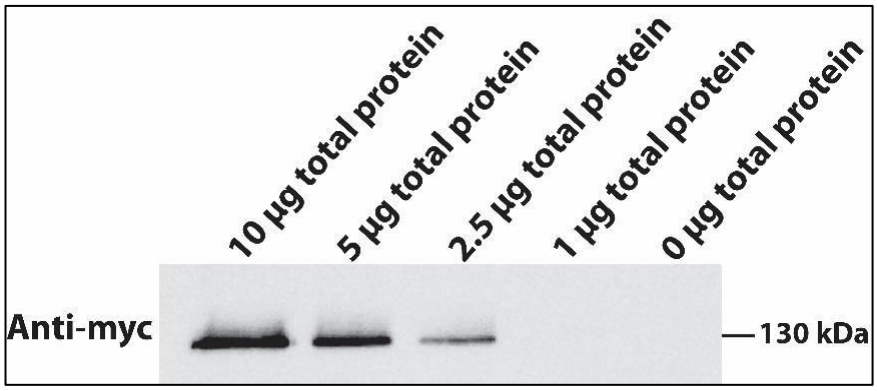

Figure 49: Immunoblotting of myc-DDK tagged ABCH2 overexpression with different amounts of protein for optimization.

The dissociation constants of peptide 1 (Fig. 50) and peptide 13 (Fig. 51) are shown. The peptide 1 and 13 effects on the concentration of myc-DDK tagged BACH2 are measured (Fig. 50, 51). According to the results, the dissociation constant of peptide 13 on myc-DDK tagged BACH2 BTB is 148.4 nM (Fig. 50). This constant is 493.3 nM for peptide 1 (Fig. 51). This means that there is 3.3 times higher affinity in the favor of peptide 13 in terms of dissociation constant. Dynabeads are used only for negative control.

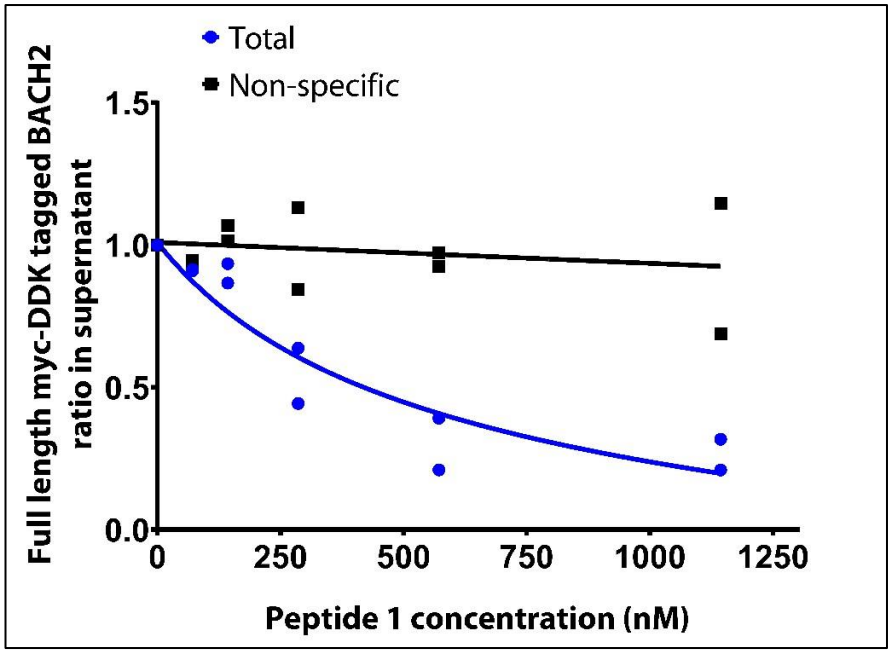

Figure 50: Graphic of supernatant depletion assay results for full length myc-DDK tagged BACH2 ratio in supernatant with increased peptide 1 volume.

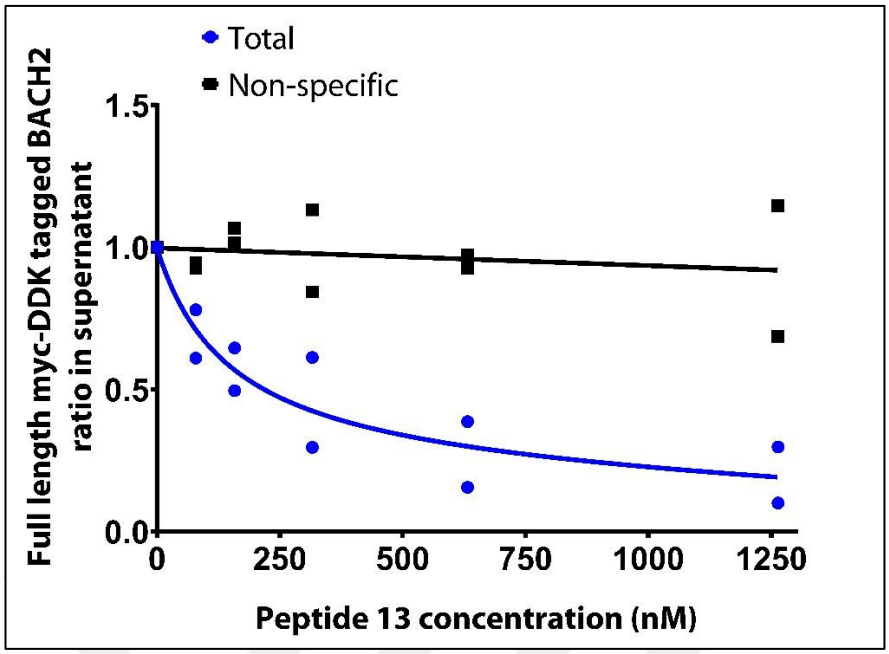

Figure 51: Graphic of supernatant depletion assay results for full length myc-DDK tagged BACH2 ratio in supernatant with increased peptide 13 volume.

The western blot results for myc-DDK tagged BACH2-peptide 13, myc-DDK tagged BACH2-peptide 1, and Dynabeads only are shown (Fig. 52). As can be seen, the concentration of the overexpressed BACH2 can be depleted with the lower amount of peptide 13 compared to peptide 1.

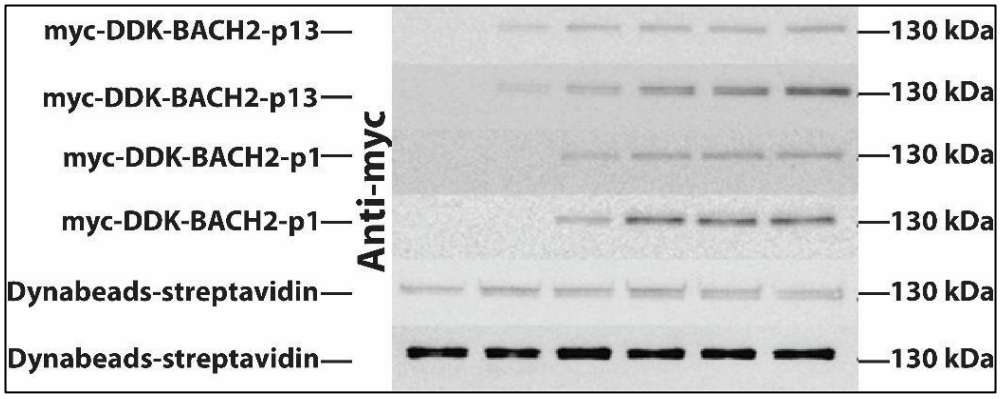

Figure 52: Supernatant depletion assay for myc-DDK tagged BACH2 with peptide 1, peptide 13 and dynabeads only.

5 DISCUSSION

Along these analyses, peptide 13 is considered a high-affinity binder to the BACH2 BTB domain with *in silico* and in vitro supernatant depletion assay. Furthermore, its affinity is higher than control peptide 1 and it is revealed that peptide 13 is more efficient for possible elimination of BACH2 homodimerization. As a result, a candidate peptide molecule is discovered for the first time, which might be a BACH2 homodimer inhibitor.

Inhibition of transcription factors by peptides is a common approach and concept which can be found in literature. Previously, it has been shown that peptide-based c-Jun antagonists bind to c-Jun and diminish its DNA binding property (Brennan et al.,2022). Previously again, it is reported that a peptide library is created to restrain BZLF1 transcription factor dimerization (Madden, Brennan, and Mason, 2024).

Another example is the report of blocking transcription factor E47 with a β-sheet peptide inhibitor; 2H10 (Ghosh, Issac, and Chmielewski, 1999). Furthermore, it is stated that there is a cyclic peptide discovery against C-terminal binding protein (CtBP) transcription factor dimerization (Birts et al., 2013). These peptide design examples can be given for other transcription factors such as STAT3 (Turkson et al., 2004) and FOXP3 (Lozano et al., 2021). Apart from that, our group predicted another peptide against NFAT5 DNA binding region dimerization with computational analyses, and the binding free energy was calculated (Timucin, 2021). All these background literatures prove the point that a transcription factor can be targeted with peptides and with our work an original peptide blocking BACH2 BTB domain dimerization is identified in this regard for the first time. After the literature research, any experimental absolute binding free energy calculation for BACH2 BTB domain homodimer was not found along with an in vitro dissociation constant for the same complex. Thereby, computationally predicted relative binding free energy of BACH2 BTB domain homodimer and BACH2 BTB domain monomer–peptide structures are used as positive controls relatively. The results are interpreted according to this relatively gained meaning. Within our work, the MM-PBSA method is used to predict

relative binding free energy, and the resulting entropic term is found as a positive value. At this point, false prediction of entropic terms via quasi-harmonic approximation is reported (Miller et al., 2012). This circumstance comprises an open field for further improvement, and advanced MD technologies such as free energy perturbation or steered MD simulations are suggested in future works. Besides, to subject, the predicted in vitro dissociation value for the conformation requires surface plasmon resonance and an isothermal titration calorimeter.

Within the MMPBSA work, during the prediction of polar solvation energy, the internal dielectric constant is valued as 1. As it is reported previously in different works, different internal dielectric constant values are used and the increased internal electric constant is coherent with computational work (Genheden and Ryde, 2015; Wang et al., 2017; Hu and Contini, 2019). Therefore, during future MMPBSA calculations, different internal dielectric constant values should be used, so that the relative free binding energy comparisons can be independent from the dielectric constant.

Another circumstance that needs to be drawn attention to is the characteristics of the cell-penetrating peptide. The peptide 13 has high affinity and functionality according to in vitro and in silico results, yet its cell-penetrating peptide amino acids contribute to binding as shown in Figure 19D. Despite the affinity contribution of the designed region is seen, HIV TAT 47 – 57 cell-penetrating peptide makes a huge contribution. This open research field should be tested with other cell-penetrating peptides so that the affinity effects of HIV TAT 47 – 57 can be tested. Cell-penetrating peptides can have off-target effects (Krautwald et al., 2016), therefore peptide design strategies must be evaluated in this regard.

Finally, within our research, all in vitro experiments are completed with protein extract. Despite the cell-penetrating peptide addition already completed, this experiment should be tested in live cells such as healthy and cancer cells. In these cells, BACH2 homo and heterodimerization should be checked with the peptide 13 addition. BACH2 is known as NRF2 transcription factor repressor, which leads to the

requirement of NRF2 targeted gene expression checks on mRNA and protein basis. Also, BACH2 plays a role in immunosuppression, and therefore analysis of peptide 13 in in vitro and in vivo immunosuppression models is suggested.

From the genetic perspective, some BACH2 mutations are known to have a relation of causality with diseases. Yet, the effect of some of these mutants on conformational stability changes in the protein structure is unknown. In the future, peptide 13's behavior will be understood better if it is checked in the case of these mutations.

6 CONCLUSION

As stated in the introduction, NRF2 activators are limited into 3 subcategories, which are electrophiles, protein-protein interaction inhibitors, and multi-target drugs. All these compounds are based on affecting NRF2 transcription via interacting with its dimer partner KEAP1, yet any therapeutic approach has not been applied to its competitor, BACH2. In the thesis, the hypothesis suggests that the BACH2 BTB domain can be targeted from its homodimerization region with a peptide, which is selective to this region. As a result, this hypothesis is approved, and it is seen that peptide 13 (RMYVYEETVHCTRIL) is efficient in this regard.

The BACH2 BTB domain is targeted by peptide 13 and molecular dynamics simulations for this domain–peptide complex are applied computationally by using several tools. At the end of these simulations, peptide 13 is opted due to its high affinity and better interaction profile compared to other peptides and positive control (homodimer complex). All these interaction analyses are later proved by determining the movements of the structures and it is carried on with in vitro experiments. The in-silico experiment results are stated as promising and followed by the in vitro experiments. For this, various plasmids containing BACH1 and BACH2 are used to transfect into HEK293T cells, which are overexpressed. The procedure is followed by the western blot analysis, which also showed promising results in terms of dissociation constant values.

To sum up, peptide 13 is a promising and efficient compound, which targets the homodimerization region of the BACH2 BTB domain and can be used as a potential drug if the necessary in vitro and in vivo experiments are applied in the future.

7 REFERENCES

1. Robledinos-Antón, N., Fernández-Ginés, R., Manda, G., & Cuadrado, A. (2019). Activators and Inhibitors of NRF2: A Review of Their Potential for Clinical Development. *Oxidative medicine and cellular longevity*, *2019*, 9372182. https://doi.org/10.1155/2019/9372182
2. Tonelli, C., Chio, I. I. C., & Tuveson, D. A. (2018). Transcriptional Regulation by Nrf2. *Antioxidants & redox signaling*, *29*(17), 1727–1745. https://doi.org/10.1089/ars.2017.7342
3. Li, W., & Kong, A. N. (2009). Molecular mechanisms of Nrf2-mediated antioxidant response. *Molecular carcinogenesis*, *48*(2), 91–104. https://doi.org/10.1002/mc.20465
4. Motohashi, H., Katsuoka, F., Engel, J. D., & Yamamoto, M. (2004). Small Maf proteins serve as transcriptional cofactors for keratinocyte differentiation in the Keap1-Nrf2 regulatory pathway. Proceedings of the National Academy of Sciences of the United States of America, 101(17), 6379–6384. https://doi.org/10.1073/pnas.0305902101
5. Motohashi, H., & Yamamoto, M. (2004). Nrf2-Keap1 defines a physiologically important stress response mechanism. *Trends in molecular medicine*, *10*(11), 549–557. https://doi.org/10.1016/j.molmed.2004.09.003
6. Nioi, P., Nguyen, T., Sherratt, P. J., & Pickett, C. B. (2005). The carboxyterminal Neh3 domain of Nrf2 is required for transcriptional activation. *Molecular and cellular biology*, *25*(24), 10895–10906. https://doi.org/10.1128/MCB.25.24.10895-10906.2005
7. McMahon, M., Thomas, N., Itoh, K., Yamamoto, M., & Hayes, J. D. (2004). Redox-regulated turnover of Nrf2 is determined by at least two separate protein domains, the redox-sensitive Neh2 degron and the redox-insensitive Neh6 degron. *The Journal of biological chemistry*, *279*(30), 31556–31567. https://doi.org/10.1074/jbc.M403061200
8. Hayes, J. D., & Dinkova-Kostova, A. T. (2014). The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends in biochemical sciences*, *39*(4), 199–218. https://doi.org/10.1016/j.tibs.2014.02.002
9. Ma Q. (2013). Role of nrf2 in oxidative stress and toxicity. *Annual review of pharmacology and toxicology*, *53*, 401–426. https://doi.org/10.1146/annurevpharmtox-011112-140320
10. Katsuoka, F., & Yamamoto, M. (2016). Small Maf proteins (MafF, MafG, MafK): History, structure and function. *Gene*, *586*(2), 197–205. https://doi.org/10.1016/j.gene.2016.03.058
11. Pajares, M., Jiménez-Moreno, N., Dias, I. H. K., Debelec, B., Vucetic, M., Fladmark, K. E., Basaga, H., Ribaric, S., Milisav, I., & Cuadrado, A. (2015). Redox control of protein degradation. *Redox biology*, *6*, 409–420. https://doi.org/10.1016/j.redox.2015.07.003
12. Pajares, M., Jiménez-Moreno, N., García-Yagüe, Á. J., Escoll, M., de Ceballos, M. L., Van Leuven, F., Rábano, A., Yamamoto, M., Rojo, A. I., & Cuadrado, A. (2016). Transcription factor NFE2L2/NRF2 is a regulator of macroautophagy genes. *Autophagy*, *12*(10), 1902–1916. https://doi.org/10.1080/15548627.2016.1208889

13. Pajares, M., Cuadrado, A., & Rojo, A. I. (2017). Modulation of proteostasis by transcription factor NRF2 and impact in neurodegenerative diseases. *Redox biology*, *11*, 543–553. https://doi.org/10.1016/j.redox.2017.01.006
14. Hayes, J. D., Chanas, S. A., Henderson, C. J., McMahon, M., Sun, C., Moffat, G. J., Wolf, C. R., & Yamamoto, M. (2000). The Nrf2 transcription factor contributes both to the basal expression of glutathione S-transferases in mouse liver and to their induction by the chemopreventive synthetic antioxidants, butylated hydroxyanisole and ethoxyquin. *Biochemical Society transactions*, *28*(2), 33–41. https://doi.org/10.1042/bst0280033
15. Venugopal, R., & Jaiswal, A. K. (1996). Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response elementmediated expression of NAD(P)H:quinone oxidoreductase1 gene. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(25), 14960–14965. https://doi.org/10.1073/pnas.93.25.14960
16. Solis, W. A., Dalton, T. P., Dieter, M. Z., Freshwater, S., Harrer, J. M., He, L., Shertzer, H. G., & Nebert, D. W. (2002). Glutamate-cysteine ligase modifier subunit: mouse Gclm gene structure and regulation by agents that cause oxidative stress. Biochemical pharmacology, 63(9), 1739–1754. https://doi.org/10.1016/s0006-2952(02)00897-3
17. Yueh, M. F., & Tukey, R. H. (2007). Nrf2-Keap1 signaling pathway regulates human UGT1A1 expression in vitro and in transgenic UGT1 mice. *The Journal of biological chemistry*, *282*(12), 8749–8758. https://doi.org/10.1074/jbc.M610790200
18. Maher, J. M., Dieter, M. Z., Aleksunes, L. M., Slitt, A. L., Guo, G., Tanaka, Y., Scheffer, G. L., Chan, J. Y., Manautou, J. E., Chen, Y., Dalton, T. P., Yamamoto, M., & Klaassen, C. D. (2007). Oxidative and electrophilic stress induces multidrug resistance-associated protein transporters via the nuclear factor-E2-related factor-2 transcriptional pathway. *Hepatology (Baltimore, Md.)*, *46*(5), 1597–1610. https://doi.org/10.1002/hep.21831
19. Reisman, S. A., Csanaky, I. L., Aleksunes, L. M., & Klaassen, C. D. (2009). Altered disposition of acetaminophen in Nrf2-null and Keap1-knockdown mice. *Toxicological sciences : an official journal of the Society of Toxicology*, *109*(1), 31–40. https://doi.org/10.1093/toxsci/kfp047
20. Lee, O. H., Jain, A. K., Papusha, V., & Jaiswal, A. K. (2007). An autoregulatory loop between stress sensors INrf2 and Nrf2 controls their cellular abundance. *The Journal of biological chemistry*, *282*(50), 36412–36420. https://doi.org/10.1074/jbc.M706517200
21. Canning, P., Sorrell, F. J., & Bullock, A. N. (2015). Structural basis of Keap1 interactions with Nrf2. *Free radical biology & medicine*, *88*(Pt B), 101–107. https://doi.org/10.1016/j.freeradbiomed.2015.05.034
22. Abed, D. A., Goldstein, M., Albanyan, H., Jin, H., & Hu, L. (2015). Discovery of direct inhibitors of Keap1-Nrf2 protein-protein interaction as potential therapeutic and preventive agents. *Acta pharmaceutica Sinica. B*, *5*(4), 285–299. https://doi.org/10.1016/j.apsb.2015.05.008
23. Lo, S. C., Li, X., Henzl, M. T., Beamer, L. J., & Hannink, M. (2006). Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *The EMBO journal*, *25*(15), 3605–3617.https://doi.org/10.1038/sj.emboj.7601243

24. Suzuki, T., Takahashi, J., & Yamamoto, M. (2023). Molecular Basis of the KEAP1-NRF2 Signaling Pathway. *Molecules and cells*, *46*(3), 133–141. https://doi.org/10.14348/molcells.2023.0028
25. Dayalan Naidu, S., & Dinkova-Kostova, A. T. (2020). KEAP1, a cysteinebased sensor and a drug target for the prevention and treatment of chronic disease. *Open biology*, *10*(6), 200105. https://doi.org/10.1098/rsob.200105
26. Holland, R., Hawkins, A. E., Eggler, A. L., Mesecar, A. D., Fabris, D., & Fishbein, J. C. (2008). Prospective type 1 and type 2 disulfides of Keap1 protein. *Chemical research in toxicology*, *21*(10), 2051–2060. https://doi.org/10.1021/tx800226m
27. Hur, W., & Gray, N. S. (2011). Small molecule modulators of antioxidant response pathway. *Current opinion in chemical biology*, *15*(1), 162–173. https://doi.org/10.1016/j.cbpa.2010.12.009
28. Satoh, T., McKercher, S. R., & Lipton, S. A. (2013). Nrf2/ARE-mediated antioxidant actions of pro-electrophilic drugs. *Free radical biology & medicine*, *65*, 645–657. https://doi.org/10.1016/j.freeradbiomed.2013.07.022
29. Wilson, A. J., Kerns, J. K., Callahan, J. F., & Moody, C. J. (2013). Keap calm and carry on covalently. *Journal of medicinal chemistry*, *56*(19), 7463–7476. https://doi.org/10.1021/jm400224q
30. Cleasby, A., Yon, J., Day, P. J., Richardson, C., Tickle, I. J., Williams, P. A., Callahan, J. F., Carr, R., Concha, N., Kerns, J. K., Qi, H., Sweitzer, T., Ward, P., & Davies, T. G. (2014). Structure of the BTB domain of Keap1 and its interaction with the triterpenoid antagonist CDDO. *PloS one*, *9*(6), e98896. https://doi.org/10.1371/journal.pone.0098896
31. Iso, T., Suzuki, T., Baird, L., & Yamamoto, M. (2016). Absolute Amounts and Status of the Nrf2-Keap1-Cul3 Complex within Cells. *Molecular and cellular biology*, *36*(24), 3100–3112. https://doi.org/10.1128/MCB.00389-16
32. Dayalan Naidu, S., Suzuki, T., Yamamoto, M., Fahey, J. W., & DinkovaKostova, A. T. (2018). Phenethyl Isothiocyanate, a Dual Activator of Transcription Factors NRF2 and HSF1. *Molecular nutrition & food research*, *62*(18), e1700908. https://doi.org/10.1002/mnfr.201700908
33. Schimrigk, S., Brune, N., Hellwig, K., Lukas, C., Bellenberg, B., Rieks, M., Hoffmann, V., Pöhlau, D., & Przuntek, H. (2006). Oral fumaric acid esters for the treatment of active multiple sclerosis: an open-label, baseline-controlled pilot study. *European journal of neurology*, *13*(6), 604–610. https://doi.org/10.1111/j.1468-1331.2006.01292.x
34. Fox, R. J., Miller, D. H., Phillips, J. T., Hutchinson, M., Havrdova, E., Kita, M., Yang, M., Raghupathi, K., Novas, M., Sweetser, M. T., Viglietta, V., Dawson, K. T., & CONFIRM Study Investigators (2012). Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. *The New England journal of medicine*, *367*(12), 1087–1097. https://doi.org/10.1056/NEJMoa1206328
35. Gold, R., Kappos, L., Arnold, D. L., Bar-Or, A., Giovannoni, G., Selmaj, K., Tornatore, C., Sweetser, M. T., Yang, M., Sheikh, S. I., Dawson, K. T., & DEFINE Study Investigators (2012).

Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *The New England journal of medicine*, *367*(12), 1098–1107. https://doi.org/10.1056/NEJMoa1114287

36. Xu, Z., Zhang, F., Sun, F., Gu, K., Dong, S., & He, D. (2015). Dimethyl fumarate for multiple sclerosis. *The Cochrane database of systematic reviews*, *2015*(4), CD011076. https://doi.org/10.1002/14651858.CD011076.pub2
37. Richardson, B. G., Jain, A. D., Speltz, T. E., & Moore, T. W. (2015). Nonelectrophilic modulators of the canonical Keap1/Nrf2 pathway. *Bioorganic & medicinal chemistry letters*, *25*(11), 2261–2268. https://doi.org/10.1016/j.bmcl.2015.04.019
38. Dubey, D., Kieseier, B. C., Hartung, H. P., Hemmer, B., Warnke, C., Menge, T., Miller-Little, W. A., & Stuve, O. (2015). Dimethyl fumarate in relapsingremitting multiple sclerosis: rationale, mechanisms of action, pharmacokinetics, efficacy and safety. *Expert review of neurotherapeutics*, *15*(4), 339–346. https://doi.org/10.1586/14737175.2015.1025755
39. Chen, Y., Inoyama, D., Kong, A. N., Beamer, L. J., & Hu, L. (2011). Kinetic analyses of Keap1-Nrf2 interaction and determination of the minimal Nrf2 peptide sequence required for Keap1 binding using surface plasmon resonance. *Chemical biology & drug design*, *78*(6), 1014–1021. https://doi.org/10.1111/j.1747-0285.2011.01240.x
40. Hancock, R., Bertrand, H. C., Tsujita, T., Naz, S., El-Bakry, A., Laoruchupong, J., Hayes, J. D., & Wells, G. (2012). Peptide inhibitors of the Keap1-Nrf2 protein-protein interaction. *Free radical biology & medicine*, *52*(2), 444–451. https://doi.org/10.1016/j.freeradbiomed.2011.10.486
41. Inoyama, D., Chen, Y., Huang, X., Beamer, L. J., Kong, A. N., & Hu, L. (2012). Optimization of fluorescently labeled Nrf2 peptide probes and the development of a fluorescence polarization assay for the discovery of inhibitors of Keap1-Nrf2 interaction. *Journal of biomolecular screening*, *17*(4), 435–447. https://doi.org/10.1177/1087057111430124
42. Tu, J., Zhang, X., Zhu, Y., Dai, Y., Li, N., Yang, F., Zhang, Q., Brann, D. W., & Wang, R. (2015). Cell-Permeable Peptide Targeting the Nrf2-Keap1 Interaction: A Potential Novel Therapy for Global Cerebral Ischemia. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *35*(44), 14727–14739. https://doi.org/10.1523/JNEUROSCI.1304-15.2015
43. Attucks, O. C., Jasmer, K. J., Hannink, M., Kassis, J., Zhong, Z., Gupta, S., Victory, S. F., Guzel, M., Polisetti, D. R., Andrews, R., Mjalli, A. M., & Kostura, M. J. (2014). Induction of heme oxygenase I (HMOX1) by HPP-4382: a novel modulator of Bach1 activity. *PloS one*, *9*(7), e101044. https://doi.org/10.1371/journal.pone.0101044
44. Dhakshinamoorthy, S., Jain, A. K., Bloom, D. A., & Jaiswal, A. K. (2005). Bach1 competes with Nrf2 leading to negative regulation of the antioxidant response element (ARE)-mediated NAD(P)H:quinone oxidoreductase 1 gene expression and induction in response to antioxidants. *The Journal of biological chemistry*, *280*(17), 16891–16900. https://doi.org/10.1074/jbc.M500166200
45. Chen, Z., Pittman, E. F., Romaguera, J., Fayad, L., Wang, M., Neelapu, S. S., McLaughlin, P., Kwak, L., & McCarty, N. (2013). Nuclear translocation of Bcell-specific transcription factor,

BACH2, modulates ROS mediated cytotoxic responses in mantle cell lymphoma. *PloS one*, *8*(8), e69126. https://doi.org/10.1371/journal.pone.0069126

46. Tsukumo, S., Unno, M., Muto, A., Takeuchi, A., Kometani, K., Kurosaki, T., Igarashi, K., & Saito, T. (2013). Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(26), 10735–10740. https://doi.org/10.1073/pnas.1306691110
47. Uittenboogaard, L. M., Payan-Gomez, C., Pothof, J., van Ijcken, W., Mastroberardino, P. G., van der Pluijm, I., Hoeijmakers, J. H., & Tresini, M. (2013). BACH2: a marker of DNA damage and ageing. *DNA repair*, *12*(11), 982–992. https://doi.org/10.1016/j.dnarep.2013.08.016
48. Tonelli, C., Chio, I. I. C., & Tuveson, D. A. (2018). Transcriptional Regulation by Nrf2. *Antioxidants & redox signaling*, *29*(17), 1727–1745. https://doi.org/10.1089/ars.2017.7342
49. Waxman E.A. Bach2 is a potent repressor of Nrf2-mediated antioxidant enzyme expression in dopaminergic neurons. *BioRxiv*. 2019:687590. https://doi.org/10.1101/687590
50. Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M., & Igarashi, K. (1996). Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Molecular and cellular biology*, *16*(11), 6083–6095. https://doi.org/10.1128/MCB.16.11.6083
51. Sun, J., Hoshino, H., Takaku, K., Nakajima, O., Muto, A., Suzuki, H., Tashiro, S., Takahashi, S., Shibahara, S., Alam, J., Taketo, M. M., Yamamoto, M., & Igarashi, K. (2002). Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. *The EMBO journal*, *21*(19), 5216–5224. https://doi.org/10.1093/emboj/cdf516
52. Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G., & Leprince, D. (1995). The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research*, *6*(9), 1193– 1198.
53. Kanezaki, R., Toki, T., Yokoyama, M., Yomogida, K., Sugiyama, K., Yamamoto, M., Igarashi, K., & Ito, E. (2001). Transcription factor BACH1 is recruited to the nucleus by its novel alternative spliced isoform. *The Journal of biological chemistry*, *276*(10), 7278–7284. https://doi.org/10.1074/jbc.M004227200
54. Afzali, B., Grönholm, J., Vandrovcova, J., O'Brien, C., Sun, H. W., Vanderleyden, I., Davis, F. P., Khoder, A., Zhang, Y., Hegazy, A. N., Villarino, A. V., Palmer, I. W., Kaufman, J., Watts, N. R., Kazemian, M., Kamenyeva, O., Keith, J., Sayed, A., Kasperaviciute, D., Mueller, M., … Laurence, A. D. J. (2017). BACH2 immunodeficiency illustrates an association between superenhancers and haploinsufficiency. *Nature immunology*, *18*(7), 813–823. https://doi.org/10.1038/ni.3753
55. Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M., & Igarashi, K. (1996). Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Molecular and cellular biology*, *16*(11), 6083–6095. https://doi.org/10.1128/MCB.16.11.6083

56. Liu, J., Sørensen, A. B., Wang, B., Wabl, M., Nielsen, A. L., & Pedersen, F. S. (2009). Identification of novel Bach2 transcripts and protein isoforms through tagging analysis of retroviral integrations in B-cell lymphomas. *BMC molecular biology*, *10*, 2. https://doi.org/10.1186/1471-2199-10-2

57. Igarashi, K., Kurosaki, T., & Roychoudhuri, R. (2017). BACH transcription factors in innate and adaptive immunity. *Nature reviews. Immunology*, *17*(7), 437–450. https://doi.org/10.1038/nri.2017.26

58. Ferreira, M. A., Matheson, M. C., Duffy, D. L., Marks, G. B., Hui, J., Le Souëf,

P., Danoy, P., Baltic, S., Nyholt, D. R., Jenkins, M., Hayden, C., Willemsen, G., Ang, W., Kuokkanen, M., Beilby, J., Cheah, F., de Geus, E. J., Ramasamy, A., Vedantam, S., Salomaa, V., … Australian Asthma Genetics Consortium (2011). Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet (London, England)*, *378*(9795), 1006–1014. https://doi.org/10.1016/S0140-6736(11)60874-X

59. International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C. C., Patsopoulos, N. A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S. E., Edkins, S., Gray, E., Booth, D. R., Potter, S. C., Goris, A., Band, G., Oturai, A. B., Strange, A., … Compston, A. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, *476*(7359), 214–219. https://doi.org/10.1038/nature10251

60. Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., Lees, C. W., Balschun, T., Lee, J., Roberts, R., Anderson, C. A., Bis, J. C., Bumpstead, S., Ellinghaus, D., Festen, E. M., Georges, M., Green, T., Haritunians, T., Jostins, L., Latiano, A., … Parkes, M. (2010). Genomewide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics*, *42*(12), 1118–1125. https://doi.org/10.1038/ng.717

61. Medici, M., Porcu, E., Pistis, G., Teumer, A., Brown, S. J., Jensen, R. A., Rawal, R., Roef, G. L., Plantinga, T. S., Vermeulen, S. H., Lahti, J., Simmonds, M. J., Husemoen, L. L., Freathy, R. M., Shields, B. M., Pietzner, D., Nagy, R., Broer, L., Chaker, L., Korevaar, T. I., … Peeters, R. P. (2014). Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. *PLoS genetics*, *10*(2), e1004123. https://doi.org/10.1371/journal.pgen.1004123

62. Roychoudhuri, R., Hirahara, K., Mousavi, K., Clever, D., Klebanoff, C. A., Bonelli, M., Sciumè, G., Zare, H., Vahedi, G., Dema, B., Yu, Z., Liu, H., Takahashi, H., Rao, M., Muranski, P., Crompton, J. G., Punkosdy, G., Bedognetti, D., Wang, E., Hoffmann, V., … Restifo, N. P. (2013). BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature*, *498*(7455), 506–510. https://doi.org/10.1038/nature12199

63. Kim, E. H., Gasper, D. J., Lee, S. H., Plisch, E. H., Svaren, J., & Suresh, M. (2014). Bach2 regulates homeostasis of Foxp3+ regulatory T cells and protects against fatal lung disease in mice. *Journal of immunology (Baltimore, Md.: 1950)*, *192*(3), 985–995. https://doi.org/10.4049/jimmunol.1302378

64. Roychoudhuri, R., Lefebvre, F., Honda, M., Pan, L., Ji, Y., Klebanoff, C. A., Nichols, C. N., Fourati, S., Hegazy, A. N., Goulet, J. P., Gattinoni, L., Nabel, G. J., Gilliet, M., Cameron, M., Restifo, N. P., Sékaly, R. P., & Flatz, L. (2015). Transcriptional profiles reveal a stepwise developmental program of memory CD8(+) T cell differentiation. *Vaccine*, *33*(7), 914–923. https://doi.org/10.1016/j.vaccine.2014.10.007
65. Roychoudhuri, R., Clever, D., Li, P., Wakabayashi, Y., Quinn, K. M., Klebanoff, C. A., Ji, Y., Sukumar, M., Eil, R. L., Yu, Z., Spolski, R., Palmer, D. C., Pan, J. H., Patel, S. J., Macallan, D. C., Fabozzi, G., Shih, H. Y., Kanno, Y., Muto, A., Zhu, J., … Restifo, N. P. (2016). BACH2 regulates CD8(+) T cell differentiation by controlling access of AP-1 factors to enhancers. *Nature immunology*, *17*(7), 851–860. https://doi.org/10.1038/ni.3441
66. Vardaka, P., Lozano, T., Bot, C. *et al.* (2020). A cell-based bioluminescence assay reveals dose-dependent and contextual repression of AP-1-driven gene expression by BACH2. *Sci Rep* 10, 18902 https://doi.org/10.1038/s41598020-75732-z
67. Cleasby, A., Yon, J., Day, P. J., Richardson, C., Tickle, I. J., Williams, P. A., Callahan, J. F., Carr, R., Concha, N., Kerns, J. K., Qi, H., Sweitzer, T., Ward, P., & Davies, T. G. (2014). Structure of the BTB domain of Keap1 and its interaction with the triterpenoid antagonist CDDO. *PloS one*, *9*(6), e98896. https://doi.org/10.1371/journal.pone.0098896
68. Dayalan Naidu, S., Muramatsu, A., Saito, R., Asami, S., Honda, T., Hosoya, T., Itoh, K., Yamamoto, M., Suzuki, T., & Dinkova-Kostova, A. T. (2018). C151 in KEAP1 is the main cysteine sensor for the cyanoenone class of NRF2 activators, irrespective of molecular size or shape. *Scientific reports*, *8*(1), 8037. https://doi.org/10.1038/s41598-018-26269-9
69. Lynch, D. R., Farmer, J., Hauser, L., Blair, I. A., Wang, Q. Q., Mesaros, C., Snyder, N., Boesch, S., Chin, M., Delatycki, M. B., Giunti, P., Goldsberry, A., Hoyle, C., McBride, M. G., Nachbauer, W., O'Grady, M., Perlman, S., Subramony, S. H., Wilmot, G. R., Zesiewicz, T., … Meyer, C. (2018). Safety, pharmacodynamics, and potential benefit of omaveloxolone in Friedreich ataxia. *Annals of clinical and translational neurology*, *6*(1), 15–26. https://doi.org/10.1002/acn3.660
70. Ramos-Gomez, M., Dolan, P. M., Itoh, K., Yamamoto, M., & Kensler, T. W. (2003). Interactive effects of nrf2 genotype and oltipraz on benzo[a]pyreneDNA adducts and tumor yield in mice. *Carcinogenesis*, *24*(3), 461–467. https://doi.org/10.1093/carcin/24.3.461
71. Okada, K., Shoda, J., Taguchi, K., Maher, J. M., Ishizaki, K., Inoue, Y., Ohtsuki, M., Goto, N., Takeda, K., Utsunomiya, H., Oda, K., Warabi, E., Ishii, T., Osaka, K., Hyodo, I., & Yamamoto, M. (2008). Ursodeoxycholic acid stimulates Nrf2-mediated hepatocellular transport, detoxification, and antioxidative stress systems in mice. *American journal of physiology. Gastrointestinal and liver physiology*, *295*(4), G735–G747. https://doi.org/10.1152/ajpgi.90321.2008
72. Kawata, K., Kobayashi, Y., Souda, K., Kawamura, K., Sumiyoshi, S., Takahashi, Y., Noritake, H., Watanabe, S., Suehiro, T., & Nakamura, H. (2010). Enhanced hepatic Nrf2 activation after ursodeoxycholic acid treatment in patients with primary biliary cirrhosis. *Antioxidants & redox signaling*, *13*(3), 259–268. https://doi.org/10.1089/ars.2009.2903

73. Richardson, B. G., Jain, A. D., Speltz, T. E., & Moore, T. W. (2015). Nonelectrophilic modulators of the canonical Keap1/Nrf2 pathway. *Bioorganic & medicinal chemistry letters*, *25*(11), 2261–2268. https://doi.org/10.1016/j.bmcl.2015.04.019
74. Jnoff, E., Albrecht, C., Barker, J. J., Barker, O., Beaumont, E., Bromidge, S., Brookfield, F., Brooks, M., Bubert, C., Ceska, T., Corden, V., Dawson, G., Duclos, S., Fryatt, T., Genicot, C., Jigorel, E., Kwong, J., Maghames, R., Mushi, I., Pike, R., … Courade, J. P. (2014). Binding mode and structureactivity relationships around direct inhibitors of the Nrf2-Keap1 complex. *ChemMedChem*, *9*(4), 699–705. https://doi.org/10.1002/cmdc.201300525
75. Marcotte, D., Zeng, W., Hus, J. C., McKenzie, A., Hession, C., Jin, P., Bergeron, C., Lugovskoy, A., Enyedy, I., Cuervo, H., Wang, D., Atmanene, C., Roecklin, D., Vecchi, M., Vivat, V., Kraemer, J., Winkler, D., Hong, V., Chao, J., Lukashev, M., … Silvian, L. (2013). Small molecules inhibit the interaction of Nrf2 and the Keap1 Kelch domain through a non-covalent mechanism. *Bioorganic & medicinal chemistry*, *21*(14), 4011–4019. https://doi.org/10.1016/j.bmc.2013.04.019
76. Jiang, Z. Y., Lu, M. C., Xu, L. L., Yang, T. T., Xi, M. Y., Xu, X. L., Guo, X. K., Zhang, X. J., You, Q. D., & Sun, H. P. (2014). Discovery of potent Keap1Nrf2 protein-protein interaction inhibitor based on molecular binding determinants analysis. *Journal of medicinal chemistry*, *57*(6), 2736–2745.https://doi.org/10.1021/jm5000529
77. Ranjan, N., Fulcrand, G., King, A., Brown, J., Jiang, X., Leng, F., & Arya, D. P. (2014). Selective Inhibition of Bacterial Topoisomerase I by alkynylbisbenzimidazoles. *MedChemComm*, *5*(6), 816–825. https://doi.org/10.1039/C4MD00140K
78. Hooper, C., Killick, R., & Lovestone, S. (2008). The GSK3 hypothesis of Alzheimer's disease. *Journal of neurochemistry*, *104*(6), 1433–1439. https://doi.org/10.1111/j.1471-4159.2007.05194.x
79. Silva, T., Reis, J., Teixeira, J., & Borges, F. (2014). Alzheimer's disease, enzyme targets and drug discovery struggles: from natural products to drug prototypes. *Ageing research reviews*, *15*, 116–145. https://doi.org/10.1016/j.arr.2014.03.008
80. Luo J. (2009). Glycogen synthase kinase 3beta (GSK3beta) in tumorigenesis and cancer chemotherapy. *Cancer letters*, *273*(2), 194–200. https://doi.org/10.1016/j.canlet.2008.05.045
81. Lal, H., Ahmad, F., Woodgett, J., & Force, T. (2015). The GSK-3 family as therapeutic target for myocardial diseases. *Circulation research*, *116*(1), 138–149. https://doi.org/10.1161/CIRCRESAHA.116.303613
82. Lovestone, S., Boada, M., Dubois, B., Hüll, M., Rinne, J. O., Huppertz, H. J., Calero, M., Andrés, M. V., Gómez-Carrillo, B., León, T., del Ser, T., & ARGO investigators (2015). A phase II trial of tideglusib in Alzheimer's disease. *Journal of Alzheimer's disease: JAD*, *45*(1), 75–88. https://doi.org/10.3233/JAD-141959
83. Bourhill, T., Narendran, A., & Johnston, R. N. (2017). Enzastaurin: A lesson in drug development. *Critical reviews in oncology/hematology*, *112*, 72–79. https://doi.org/10.1016/j.critrevonc.2017.02.003

84. Itoh-Nakadai, A., Hikota, R., Muto, A., Kometani, K., Watanabe-Matsui, M., Sato, Y., Kobayashi, M., Nakamura, A., Miura, Y., Yano, Y., Tashiro, S., Sun, J., Ikawa, T., Ochiai, K., Kurosaki, T., & Igarashi, K. (2014). The transcription repressors Bach2 and Bach1 promote B cell development by repressing the myeloid program. *Nature immunology*, *15*(12), 1171–1180. https://doi.org/10.1038/ni.3024
85. Roychoudhuri, R., Eil, R. L., Clever, D., Klebanoff, C. A., Sukumar, M., Grant, F. M., Yu, Z., Mehta, G., Liu, H., Jin, P., Ji, Y., Palmer, D. C., Pan, J. H., Chichura, A., Crompton, J. G., Patel, S. J., Stroncek, D., Wang, E., Marincola, F. M., Okkenhaug, K., … Restifo, N. P. (2016). The transcription factor BACH2 promotes tumor immunosuppression. *The Journal of clinical investigation*, *126*(2), 599–604. https://doi.org/10.1172/JCI82884
86. Kavsan, V. M., Iershov, A. V., & Balynska, O. V. (2011). Immortalized cells and one oncogene in malignant transformation: old insights on new explanation. *BMC cell biology*, *12*, 23. https://doi.org/10.1186/1471-2121-1223
87. Thomas, P., & Smart, T. G. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. *Journal of pharmacological and toxicological methods*, *51*(3), 187–200. https://doi.org/10.1016/j.vascn.2004.08.014
88. Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical reviews in biotechnology*, *36*(6), 1110–1122. https://doi.org/10.3109/07388551.2015.1084266
89. Rio, D. C., Clark, S. G., & Tjian, R. (1985). A mammalian host-vector system that regulates expression and amplification of transfected genes by temperature induction. *Science (New York, N.Y.)*, *227*(4682), 23–28. https://doi.org/10.1126/science.2981116
90. Jang, M., Pete, E. S., & Bruheim, P. (2022). The impact of serum-free culture on HEK293 cells: From the establishment of suspension and adherent serumfree adaptation cultures to the investigation of growth and metabolic profiles. *Frontiers in bioengineering and biotechnology*, *10*, 964397. https://doi.org/10.3389/fbioe.2022.964397
91. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, *32*(10), 2319–2327. https://doi.org/10.1002/jcc.21787
92. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), 33–28. https://doi.org/10.1016/0263-7855(96)00018-5
93. V Sedan, Y., Marcu, O., Lyskov, S., & Schueler-Furman, O. (2016). Peptiderive server: derive peptide inhibitors from protein-protein interactions. *Nucleic acids research*, *44*(W1), W536–W541. https://doi.org/10.1093/nar/gkw385
94. Lyskov, S., Chou, F. C., Conchúir, S. Ó., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D., Sripakdeevong, P., Borgo, B., Havranek, J. J., Kuhlman, B., Kortemme, T., Bonneau, R., Gray, J. J., & Das, R. (2013). Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PloS one*, *8*(5), e63906. https://doi.org/10.1371/journal.pone.0063906

95. Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering*, *8*(2), 127–134. https://doi.org/10.1093/protein/8.2.127
96. Schymkowitz, J. W., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F., & Serrano, L. (2005). Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10147–10152. https://doi.org/10.1073/pnas.0501980102
97. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic acids research*, *33*(Web Server issue), W382–W388. https://doi.org/10.1093/nar/gki387
98. Maupetit, J., Derreumaux, P., & Tufféry, P. (2010). A fast method for largescale de novo peptide and miniprotein structure prediction. *Journal of computational chemistry*, *31*(4), 726–738. https://doi.org/10.1002/jcc.21365
99. Thévenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P., & Tufféry, P. (2012). PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic acids research*, *40*(Web Server issue), W288–W293. https://doi.org/10.1093/nar/gks419
100. Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., & Tufféry, P. (2016). PEP FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic acids research*, *44*(W1), W449–W454. https://doi.org/10.1093/nar/gkw329
101. Porter, K. A., Xia, B., Beglov, D., Bohnuud, T., Alam, N., Schueler-Furman, O., & Kozakov, D. (2017). ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. *Bioinformatics (Oxford, England)*, *33*(20), 3299–3301. https://doi.org/10.1093/bioinformatics/btx216
102. Honorato, R. V., Trellet, M. E., Jiménez-García, B., Schaarschmidt, J. J., Giulini, M., Reys, V., Koukos, P. I., Rodrigues, J. P. G. L. M., Karaca, E., van Zundert, G. C. P., Roel-Touris, J., van Noort, C. W., Jandová, Z., Melquiond, A. S. J., & Bonvin, A. M. J. J. (2024). The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature protocols*, 10.1038/s41596-024-01011-0. Advance online publication. https://doi.org/10.1038/s41596-024-01011-0

8 CURRICULUM VITAE