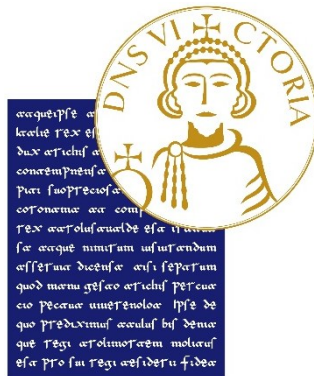


UNIVERSITY OF SANNIO



DEPARTMENT OF SCIENCES AND TECHNOLOGIES

PhD in Sciences and Technologies

for Environment and Health

XXXVIII cycle

Design and characterization of engineered proteins for biotechnological applications

Supervisor

Prof. Dr. Roberta Spadaccini

PhD Student

Andrea Bologna

PhD Coordinator

Prof. Dr. Marina Paolucci

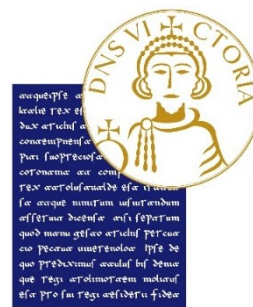
Academic Year 2024/2025



**Ministero
dell'Università
e della Ricerca**



**Universität
Marburg**



The PhD project was carried out under the supervision of Prof. Dr. Roberta Spadaccini from January 2023 to December 2025 and was supported by a doctoral fellowship awarded by the University of Sannio. The programme was structured as an inter-institutional training and research path.

As part of the doctoral programme, a 15-month research stay (January 2023 – March 2024) was conducted within the Structural Biology group of the Department of Chemical Sciences at the University of Naples Federico II, led by Prof. Dr. Delia Picone.

The PhD also included a 12-month international mobility period (April 2024 – March 2025) at Philipps-Universität Marburg (Germany), where the research activity was carried out within the Structural Biochemistry group directed by Prof. Dr. Lars-Oliver Essen.

Abstract

Protein design is nowadays a powerful tool for developing different strategies that can be used to rationally control the stability, structural dynamics, and functional behaviour of proteins of biotechnological interest. Through a biochemical, structural, and computational approach, this work explores the relationship between targeted structural modifications and emerging properties of the systems studied.

The sweet protein MNEI, a single-chain monellin, a model protein for sweetness and aggregation studies, was used to understand how topological interventions of circular permutation or point substitutions can alter the balance between stability, swapped dimer formation, and aggregation. The results show that small variations in chain connectivity are enough to change the protein's fate towards ordered fibrils, amorphous aggregates, while changes in the distribution of internal charges can lead to enhanced structural stability accompanied by a drastic reduction in aggregation.

Aggregation is therefore a process that can be modulated through targeted design choices. The second part of the thesis is focusing on PIR (Proteins with Internal Repetitions) proteins, in particular on sequence modularity and the role of internal repeats in cell wall anchoring. The analysis of single-repeat constructs made it possible to investigate their contribution to structural stability and interaction with carbohydrates, highlighting a relationship between modular organization and biological function. In summary, the work shows how minimal but strategic changes to the sequence or topology can be exploited to optimize proteins for food, industrial, and biotechnological applications.

Table of contents

1 Introduction	1
1.1 Proteins as tools for biotechnological engineering	1
1.1.1 Design approaches and structural characterisation methods	3
1.1.2 Protein dynamics and stability	6
1.1.3 Effects of mutations and topology changes on protein-protein interactions	7
1.2 Sweet proteins	8
1.2.1 Monellin and its variants	11
1.2.2 Cooperative and conformational remodelling phenomena	14
1.2.3 Protein Aggregation	15
1.2.4 Domain Swapping	18
1.2.5 Circular Permutation as a tool to modulate protein stability and conformational behaviour	21
1.3 PIR proteins for stable cell wall architectures	23
1.4 Aim of the thesis	28
2 Materials	30
2.1 Strains and Plasmids	30
2.1.1 <i>E. coli</i> strains	30
2.1.2 Plasmids	31
2.2 Enzymes	32
2.3 Chemicals	32
2.3.1 Buffers and solutions	32
2.3.2 Culture media	34
2.3.3 Inductors	34
2.3.4 Antibiotics	34
2.3.5 Antibodies	35
2.4 Computational resources and software	35
2.4.1 Servers and computing infrastructure	35
2.4.1.1 Marconi100 on CINECA HPC	35

2.4.1.2 Local server at UniMarburg -----	36
2.4.2 Software, algorithms and web tools -----	38
3 Methods -----	39
3.1 Molecular Biology-----	39
3.1.1 Circular permutation-----	39
3.1.2 Plasmid preparation and Verification -----	39
3.1.3 Sanger sequencing, verification of the DNA and protein concentration -----	40
3.1.4 Preparation of chemical competent <i>E. coli</i> cells-----	41
3.1.5 Transformation of <i>E. coli</i> -----	42
3.2 Protein biochemistry-----	43
3.2.1 Cultivation of <i>E. coli</i> -----	43
3.2.1.1 Test expression -----	43
3.2.1.2 Measuring the optical density of liquid cultures -----	44
3.2.1.3 Large-scale expression -----	45
3.2.2 Cell lysis-----	45
3.2.2.1 Mechanical disruption using glass beads-----	46
3.2.2.2 Ultrasonic cell disruption-----	46
3.2.2.3 High-pressure cell disruption (French Press)-----	47
3.2.2.4 High-pressure homogenization (Avestin® Emulsiflex-C5 System) -----	48
3.3 Protein purification -----	49
3.3.1 Centrifugation and Clarification of Crude Extracts -----	49
3.3.2 Ni-NTA Affinity chromatography -----	49
3.3.2.1 Small-scale purification using spin Ni-NTA columns --	49
3.3.2.2 Medium- to large-scale purification using Ni-NTA columns-----	51
3.3.2.3 Reverse Ni-NTA purification -----	52
3.3.3 Cation exchange chromatography -----	53
3.3.4 Size-exclusion chromatography (SEC) -----	54
3.4 Protein analytics -----	56

3.4.1 SDS-PAGE	56
3.4.2 Western blot.....	58
3.4.2.1 His-tag detection (DAB chromogenic development)---	58
3.4.2.2 Immunodetection with primary and secondary antibodies (ECL chemiluminescence).....	59
3.4.3 Native PAGE.....	60
3.4.4 UV-Vis spectrophotometry for protein concentration determination	61
3.4.5 Circular Dichroism (CD) spectroscopy	62
3.4.6 Mass Spectrometry analysis	64
3.4.7 Thioflavin T (ThT) fluorescence assay	64
3.4.8 Transmission Electron Microscopy (TEM).....	66
3.4.8.1 Grid preparation	66
3.4.8.2 Grid Analysis	67
3.4.9 Lyophilization and Long-Term Protein Preservation	67
3.5 Protein crystallography	68
3.5.1 Crystallization and diffraction data processing.....	68
3.5.2 Structure Refinement.....	70
3.7 Bioinformatics	70
3.7.1 Protein structure prediction tools	70
3.7.2 Molecular dynamics simulations	71
4 Results.....	73
4.1 Development of a MNEI Mutant.....	73
4.1.1 Design of the Mut4 (E23Q) mutant	73
4.1.2 Protein production and purification	74
4.1.3 CD spectroscopy and evaluation of the secondary stability of Mut4	76
4.1.4 Aggregation propensity assessed by ThT fluorescence ---	81
4.1.7 Structure prediction and MD simulations.....	83
4.1.7.1 RMSD analysis	85
4.1.7.2 RMSF analysis	87

4.1.7.3 Statistical comparison between pH and temperature conditions	88
4.1.8 Discussion	91
4.2 Circular permuted mutants of the sweet protein MNEI	95
4.2.1 Design of Perms	95
4.2.2 Protein production and purification	97
4.2.3 Stability and aggregation studies	98
4.2.3.1 CD spectroscopy and secondary stability evaluation -	98
4.2.3.2 ThT analysis and aggregation curves	106
4.2.3.4 SEC and TEM analysis in aggregation conditions	109
4.2.4 Crystal structure of Perm mutants	115
4.2.4.1 Topological comparison of PERM domain-swapped dimers with cystatin family members	120
4.2.5 Molecular dynamics analysis of circularly permuted mutants	123
4.2.5.1 RMSD analysis	123
4.2.5.2 RMSF analysis	124
4.2.5.3 DSSP analysis	131
4.2.5.4 PCA: calculation setup and trajectory projection	134
4.2.6 Discussion	143
4.3 PIRs	149
4.3.1 Concept and design of PIR constructs	149
4.3.2 Exploratory expression and solubility tests	151
4.3.2.1 GFP-PIR1r1-8	151
4.3.2.2 PIR1r1	156
4.3.2.3 PIR2r1	158
4.3.2.4 PIR4r1	163
4.3.2.5 PIR5r1	166
4.3.2.6 PIR sequencing	168
4.3.3 Large-scale expression and purification	169
4.3.3.1 PIR2r1	169

4.3.3.2 PIR4r1 -----	173
4.3.3.3 PIR5r1 -----	175
4.3.4 Western blot: immunological validation of single-repeat constructs -----	177
4.3.5 Crystallisation -----	181
4.3.5.1 PIR1r1 and rationale for computational analysis -----	181
4.3.5.2 Crystallisation of other PIRs -----	186
4.3.6 Computational characterisation of PIR1r1: APO molecular dynamics simulations -----	188
4.3.6.1 RMSD analysis -----	188
4.3.6.2 RMSF analysis -----	190
4.3.7 Discussion -----	192
5 Conclusions and future perspectives -----	197
6 References -----	202
7 Supplement -----	228
7.1 List of Publications -----	228
7.1.1 Substantial contributions to collaborative research not reflected in authorship -----	237
7.2 Abbreviations -----	244
7.3 List of Amino Acids -----	248
7.4 Curriculum Vitae – Academic career -----	249
8 Acknowledgements – Academic career -----	251
9 Closing statements -----	256

1 Introduction

1.1 Proteins as tools for biotechnological engineering

Proteins are one of the foundations of biotechnological engineering, thanks to their capability to perform chemical and structural functions with a level of precision that is difficult to emulate by synthetic molecules. This versatility is the result of a combination of highly selective interactions, remarkable architectural diversity and the possibility of remodelling their behaviour through even minimal modifications to their amino acid sequence. For these reasons, proteins are an ideal platform for developing catalysts, receptors, functional materials and self-organised systems in sectors ranging from food to medicine, biocatalysis and environmental protection.

The conceptual shift that made all these possible dates to the 1970s, with the introduction of recombinant DNA technology. The studies of Stanley N. Cohen and Herbert W. Boyer demonstrated that genetic fragments from different sources could be combined in vitro, cloned in *Escherichia coli* and expressed in a stable way. [1] This demonstration led to the controlled production of heterologous proteins and ushered in the era of genetic engineering. The subsequent development of recombinant human insulin was a turning point not only in terms of therapy but also culturally: it revealed that genetic information could be manipulated to obtain, on a large scale, proteins that were functionally equivalent (and in many cases superior) to their natural counterparts. [2–4]

Once this model was established, protein engineering evolved rapidly. Early attempts at rational design highlighted how even a single mutation could significantly alter structure and function; however, they also highlighted the limitations of an approach based on an incomplete understanding of the rules governing the sequence-structure-function relationship and, therefore, protein folding. [5] The introduction of directed evolution in the 1990s radically changed the picture because, by combining iterative cycles of mutagenesis and selection, it became possible to develop enzymes and proteins with new specificities, increased stability or optimised catalytic activity, overcoming constraints that purely rational design could not

address. [6–8] More recently, the integration of advanced structural methods, molecular dynamics simulations and artificial intelligence tools has made truly customised design possible, leading to the creation of de novo proteins with desired properties. [9–12]

The applications resulting from this technological advancement are extremely widespread. The versatility of protein engineering can be illustrated through the study of specific protein systems that embody complementary functional and structural paradigms. Among these, sweet-tasting proteins and fungal cell wall proteins represent two conceptually distinct but convergent examples of how protein sequence and modular organisation can be exploited to tailor macroscopic properties.

Sweet proteins, such as Monellin and its engineered variants, are traditionally investigated for their potential use as natural low-calorie sweeteners. [13] However, their relevance extends well beyond the food sector. Due to their well-defined fold, high solubility under native conditions and pronounced sensitivity to sequence perturbations, Monellin and its variants constitute a valuable system for probing protein stability, misfolding and aggregation phenomena. Protein aggregation is not only a central issue in food and bioprocessing but also a major biomedical concern, being associated with a wide range of pathological conditions. [14] Moreover, the controlled self-assembly of protein-based systems has recently attracted considerable interest for the development of functional biomaterials. In this context, engineered proteins and peptides capable of forming ordered aggregates or amyloid-like structures have been explored for applications ranging from nanostructured materials to the sequestration of heavy metals, dyes and organic pollutants from aqueous environments, highlighting the potential of protein self-assembly as a programmable tool in environmental biotechnology. [15,16]

In parallel, cell wall proteins such as the PIR family from *Saccharomyces cerevisiae* provide a complementary framework to investigate how modular protein architecture contributes to collective mechanical and functional properties at the cellular level. PIR proteins are characterised by internal

repetitive units involved in covalent attachment to β -1,3-glucan, playing a key role in cell wall integrity and rigidity. The intrinsic plasticity of the yeast cell wall, combined with the modular organisation of PIR proteins, makes this system particularly amenable to engineering strategies aimed at reshaping cell surface properties in a controlled manner. Manipulating the number, organisation or anchoring efficiency of PIR repeats offers the possibility to modulate not only cell wall mechanics but also the spatial presentation of functional moieties at the cell surface. This approach underpins emerging biotechnological strategies in which yeast cells are exploited as dynamic platforms for surface-associated functionalities, including the display of enzymatic activities that are difficult to implement intracellularly due to substrate toxicity, limited permeability or high local concentrations. [17] From this perspective, PIR-mediated cell wall engineering is relevant not only for the development of more robust industrial strains, but also for the design of self-renewing, functionally active cell-based materials with applications in biocatalysis and biotechnology. Together, these two protein systems exemplify how protein engineering can be employed to explore structure-function relationships across different scales, from molecular self-assembly to supramolecular organisation, thereby providing a versatile platform for applications that span biotechnology, materials science and cellular engineering.

1.1.1 Design approaches and structural characterisation methods

Contemporary protein design arises from the convergence of molecular biology, structural chemistry, bioinformatics and, more recently, artificial intelligence, which in modern laboratory practice operate as a closely integrated framework rather than as independent disciplines.

Protein design can therefore be better understood as an iterative process in which structural hypotheses, computational modelling and experimental validation continuously influence each other. [6,18–20]

When high-resolution structural information or reliable predictive models are available, rational design is a natural starting point. Classical structural biology techniques, such as X-ray crystallography and NMR spectroscopy, have historically defined the principles underlying the structure-function relationship [21,22], while cryo-electron microscopy has greatly expanded access to the structures of large assemblies and conformational heterogeneous systems. [23]

In recent years, deep learning-based structural prediction methods, notably AlphaFold2, AlphaFold3 and RoseTTAFold, have profoundly changed this landscape. Rather than serving only as tools for in silico mutational analysis, these predictors increasingly function as general structural frameworks supporting experimental structural biology. The predicted models often provide valuable starting points for further experimental structural and functional studies, including crystallography, model building, and cryo-EM model interpretation, as well as for the analysis of protein-protein interfaces. [24–26] However, the lack of explicit information on environmental conditions, conformational heterogeneity, and dynamic equilibria requires that predicted structures be interpreted with caution, particularly for flexible, charged, or intrinsically disordered regions. [27,28]

Alongside rational approaches, evolutionary strategies have proven essential for exploring regions of sequential space that are difficult to anticipate computationally. Site-directed mutagenesis remains indispensable for testing well-defined hypotheses, while broader approaches such as saturation or random mutagenesis allow functional regions to be interrogated without assuming in advance what the best solution is. [29] Direct evolution, introduced by Arnold & co. [6,8], has demonstrated that iterative cycles of mutation and selection can efficiently produce variants with improved or novel properties, highlighting how mutations distant from active or binding sites can exhibit significant functional effects. [27,30] Although direct evolution in the literal sense is not employed in this thesis, its conceptual framework emphasises the

importance of distributed and often non-intuitive sequence-function relationships.

Computational methods now play a central role in all stages of protein design and characterisation. Historically, explicit energy-based approaches such as Rosetta and its design modules (RosettaDesign, FastDesign) were the standard for evaluating mutational effects and identifying stabilised variants. [31,32] This paradigm has been redefined by deep learning. The release of AlphaFold2 marked a turning point in structural prediction, and AlphaFold3 further extended its applicability to protein complexes and interactions with nucleic acids and small molecules, achieving near-experimental accuracy for many systems, particularly for well-behaved domains represented in the training datasets. [24,33,34] In parallel, transformation-based platforms such as Chai-1 and ChaiDiscovery have demonstrated the feasibility of rapid, high-resolution prediction of multi-protein assemblies, enabling early-stage evaluation of new variants and scaffolds. [35] Generative approaches, including ProteinMPNN for structure-sequence design [36] and RFdiffusion for controlled de novo fold generation [9], further expand the accessible design space by enabling the creation of entirely new protein sequences with physically plausible geometries. Although these methods do not replace experimental validation, they greatly accelerate the exploratory phase of protein engineering.

Despite these advances, many functionally relevant protein properties only emerge when conformational dynamics are explicitly considered. Molecular dynamics simulations provide a time-resolved description of protein behaviour, enabling the study of stability, residue-specific mobility, transient interactions, hydration effects, and rare conformational states that are often inaccessible to static experimental techniques. [37] It is important to note that dynamic information does not come exclusively from simulations, but also from experimental approaches such as NMR spectroscopy, which directly probes conformational flexibility and exchange processes, and from the overall interpretation of crystallographic and cryo-EM data.

1.1.2 Protein dynamics and stability

Protein stability is a more nuanced concept than what emerges from a single 3D structure. Even in its native state, a protein explores a set of conformational microstates in equilibrium, and its functional properties depend on the distribution of these states rather than on a single “average” conformation. This “ensemble-based” perspective has established itself in recent decades as the interpretative framework of reference for understanding allostery, evolvability and functional adaptation of proteins. [28,38]

From a biophysical point of view, it is useful to distinguish between thermodynamic stability, linked to the difference in free energy between the native and denatured states, and kinetic stability, which depends on the height of the energy barriers that separate the native state from alternative, partially or totally unfolded states. [39] In fact, a protein can be thermodynamically unstable but kinetically “locked”, or vice versa. Point mutations and changes in conditions can therefore non-intuitively alter the balance between these two aspects, modifying the population of accessible states and their functional relevance. [27,40,41] Another crucial element is the role of the environment: pH, ionic strength, temperature and solvent composition influence the protonation of residues, the hydrogen bond network and hydrophobic interactions, reshaping the set of populated conformational states. Even local changes can have significant effects on folding cooperativity, aggregation tendency, or the appearance of marginally stable intermediates, especially in proteins with flexible regions or non-canonical architectures. [40,41]

To study these phenomena, both experimental and computational techniques are used. Spectroscopies such as CD and fluorescence provide information on global stability and folding cooperativity [42,43], while crystallography allows native and non-native states to be observed with different spatial resolutions. On the computational side, molecular dynamics simulations provide insight into residue-specific mobility, hydrogen bond

persistence, and the response of the structure to controlled perturbations, integrating and interpreting experimental data. [44–46] In the context of protein engineering, understanding how stability and dynamics combine is particularly important for systems in which mutations, topological reorganizations, or modular interactions alter the balance between native and intermediate conformational states, making stability itself a functional and design parameter.

1.1.3 Effects of mutations and topology changes on protein-protein interactions

The biological function and biotechnological applications of a protein do not depend exclusively on its static 3D structure but emerge from the set of interactions it establishes with other proteins, macromolecular surfaces, or components of the extracellular matrix. [37,47] In this context, point mutations and topological changes in the polypeptide chain are powerful tools for reshaping these interactions, substantially altering the collective behaviour of the protein system. [27]

In recent years, protein engineering has shown how apparently limited variations, such as the substitution of a single residue, the removal of a charge, or the repositioning of the N- and C-terminals, can profoundly influence the accessibility of contact surfaces, the stability of molecular complexes, and the propensity for self-association. [48] These effects are particularly evident in proteins whose fold is maintained by a delicate balance between rigid and flexible regions, or in systems where function depends on multiple, cooperative interactions rather than a single active site. [49,50]

In the food industry, a symbolic example is represented by sweet proteins, whose taste effect derives from selective interaction with the human sweet taste receptor T1R2/T1R3. [51] In these systems, point mutations or topological modifications can preserve receptor interaction and thus the perception of sweetness, while altering stability, solubility, or pH response, which are critical parameters for industrial use. [52,53] The rational design

of more resistant variants shows that it is possible to intervene on the structural stability and intermolecular interactions of the protein without necessarily altering its ability to activate the sweet taste receptor.

A second example, functionally very different but conceptually similar, is provided by the PIR (Proteins with Internal Repeats) proteins of *Saccharomyces cerevisiae*. In this case, the function does not depend on interaction with a receptor, but on the capability to establish multiple and cooperative bonds with the polysaccharide matrix of the cell wall. The repeated modules containing the conserved motif SQGDGQQA constitute units of interaction with β -1,3-glucan, and variation in the number or arrangement of these repeats directly modulates the mechanical stability and integrity of the wall. [54–56] Similarly, in PIR proteins, variations in the arrangement of repeated modules or in their organization along the polypeptide chain directly influence how the protein anchors to the glucan matrix of the cell wall.

These examples, despite belonging to distinct biological and application contexts, share a common principle: mutations and topological reorganizations act as levers to modulate protein-protein and protein-matrix interactions, influencing not only specific function but also the stability, cooperativity, and collective behaviour of proteins. Precisely this perspective makes it necessary to integrate structural, dynamic, and computational approaches in the study of engineered proteins.

These examples highlight how targeted modifications to the sequence or topology can prompt a protein to explore alternative conformational states. The following sections will therefore discuss the main cooperative phenomena that emerge from such structural reorganizations, with particular attention to aggregation processes and other conformational remodelling mechanisms.

1.2 Sweet proteins

Nowadays, eating too much sugar is seen as one of the main risk factors for obesity, type 2 diabetes, heart disease, and other metabolic disorders.

International organizations such as the WHO and FAO have long recommended reducing the intake of free sugars, also in light of the environmental impact associated with the industrial production of sucrose from cane and beet, which involves intensive use of soil, water, and energy. [57,58] This context has led to the search for alternative sweeteners that are effective, safe, and compatible with sustainability goals.

High-intensity artificial sweeteners, such as aspartame, sucralose, and acesulfame-K, have been the most immediate technological response for years: their high sweetening power and negligible caloric contribution have favoured their widespread use in “light” or “sugar-free” products. However, various findings have raised questions about their long-term impact, particularly in relation to carbohydrate metabolism, the composition of the gut microbiota, and the formation of unwanted degradation products under specific conditions of use. [59,60] Although there is no definitive agreement on all these aspects, interest in natural sweeteners with a more reassuring physiological profile is constantly growing. It is in this context that sweet proteins come into play, a heterogeneous group of small proteins of plant origin that share an extremely high sweetening power, sometimes hundreds or thousands of times greater than that of sucrose at the same concentration. [52,61] Among the first to be identified are miraculin, curculin, monellin, thaumatin, and brazzein, extracted from tropical fruits or African plants. [52,62] Their effect is mediated by the activation of the T1R2/T1R3 receptor through complex recognition surfaces, often sensitive to the protonic state of the residues and the local environment. [51,63] (Figure 1)

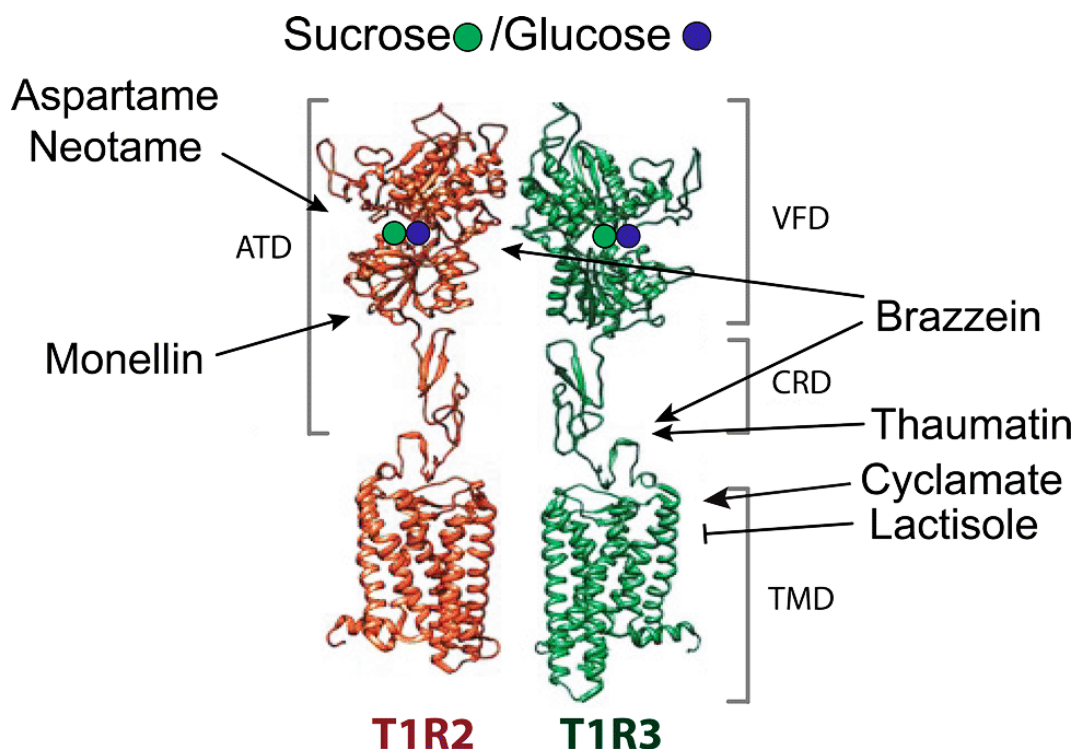


Figure 1. Schematic representation of the human sweet taste receptor heterodimer T1R2 (red) and T1R3 (green), highlighting the extracellular amino-terminal domain, composed of the Venus flytrap domain (VFD) and the cysteine-rich domain (CRD), and the seven-helical transmembrane domain (TMD). Different classes of sweet compounds interact with distinct regions of the receptor, reflecting the presence of multiple recognition sites. Figure adapted from [64].

From a biotechnological point of view, sweet proteins have some particularly interesting characteristics, such as the absence of significant caloric intake, high taste specificity, digestive degradation into amino acids, which suggests good physiological compatibility, and potential sustainability, especially when produced through recombinant expression in microbial systems. [65] Despite these advantages, their large-scale industrial application is still limited by problems of thermal stability, pH sensitivity, solubility, and production costs. To address these obstacles, several protein engineering approaches have been developed: rational mutagenesis targeting critical residues, stabilization of flexible loops, and removal of particularly labile residues. [66–68]

1.2.1 Monellin and its variants

Among the various sweet proteins identified in nature, one of the most studied from a structural and engineering point of view is monellin, a small dimeric protein isolated from the fruit *Dioscoreophyllum cumminsii*. [13] (Figure 2) The natural form, consisting of two non-covalently associated chains, has a high sweetening intensity but shows limited stability under typical food industry operating conditions, particularly at acidic pH and high temperatures. [69–71]



Figure 2. Photograph of the fruit of *Dioscoreophyllum cumminsii*, the plant source from which the monellin is isolated. Figure adapted from [72].

To overcome this limitation, the monomeric variant MNEI (SCM) was developed by connecting the two original chains with a Glycine-Phenylalanine peptide linker between residues 50-51. [73] This strategy led to the generation of a compact protein with a well-defined fold and greater resistance to denaturation than the dimeric form. The conversion to a single chain eliminates the dependence on non-covalent association and stabilizes the interface between the two original domains, making the system more suitable for recombinant production and possible biotechnological applications. [68,74,75] (Figure 3)

Monellin | Chain A

REIKGYEYQL¹⁰VYASDKLFR²⁰ADISEDYKTR³⁰GRKLLRFNGP⁴⁰VPPP

Monellin | Chain B

GEWEIIDIGP¹⁰FTQNLGKFAV²⁰DEENKIGQYG³⁰RLTFNKVIRP⁴⁰CMKKTIY
ENE⁵⁰

MNEI | SCM

GEWEIIDIGP¹⁰FTQNLGKFAV²⁰DEENKIGQYG³⁰RLTFNKVIRP⁴⁰CMKKTIY
ENE⁵⁰GFREIKGYEY⁶⁰QLYVYASDKL⁷⁰FRADISEDYK⁸⁰TRGRKLLRFN⁹⁰G
PVPPP

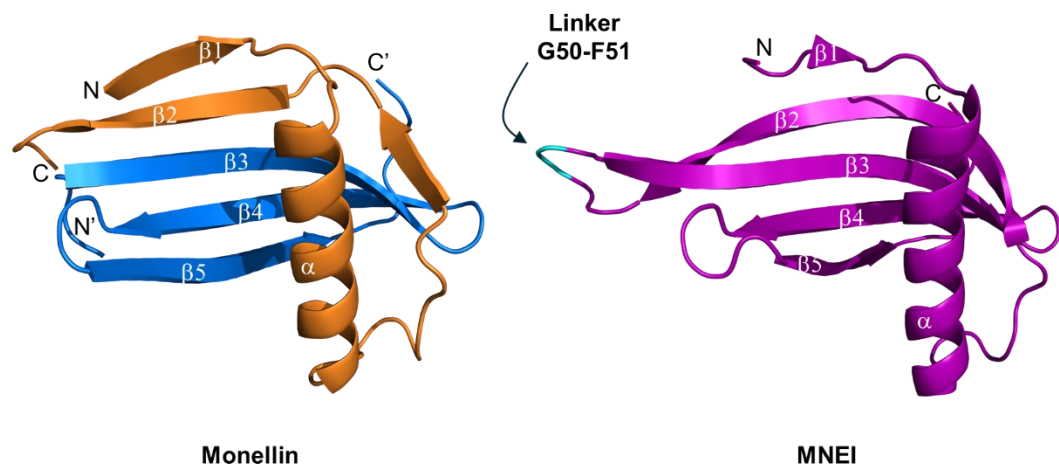


Figure 3. Comparison between native monellin and single-chain monellin MNEI (SCM). Top: amino acid sequences of chain A and chain B of monellin and MNEI. In the single-chain structure, the two original polypeptides are covalently linked via a Gly50–Phe51 linker, highlighted in cyan. Bottom: Three-dimensional structures of native monellin (left) and MNEI (right), showing the conservation of the cystatin-like α/β fold.

From a conceptual point of view, the design of MNEI represents more than just a stabilisation strategy: it is a true topological modification, capable of remodelling the distribution of intramolecular interactions without compromising those responsible for taste receptor recognition. This makes MNEI a particularly suitable system for studying how sequence and

topology variations can influence stability and conformational behaviour while maintaining primary function.

Despite improvements over native monellin, the stability of MNEI remains sensitive to the environment, particularly to acidic pH, a condition relevant to many beverages and food products. [53] This has inspired several studies aimed at identifying the residues responsible for structural lability and designing more robust variants. Engineering interventions have focused on residues that are charged or easily protonated at low pH, particularly flexible backbone regions, weak contacts that contribute to local stability, and conservative mutations capable of reinforcing the hydrophobic core without substantially altering the recognition profile. [53,76,77]

This line of investigation builds upon a substantial body of previous work aimed at improving the stability of MNEI through targeted mutations. Early studies identified individual residues involved in pH sensitivity and conformational lability, leading to the design of single-point mutants such as MNEI-Y65R, in which the introduction of a positively charged residue was shown to enhance electrostatic stabilization and resistance to acidic conditions without impairing sweetness perception. [53,76] Subsequent efforts expanded this approach to multiple-site variants, combining conservative substitutions in flexible or weakly packed regions to further reinforce the hydrophobic core and local interaction networks. More recently, advanced mutational strategies have resulted in the development of higher-order variants, including Mut9, reported in 2024, which integrates multiple stabilising substitutions and displays improved structural robustness under harsh physicochemical conditions while preserving sensory properties. [78]

Within this framework, the Mut4 variant, corresponding to the replacement of Glu23 with glutamine (E23Q), represents a paradigmatic example of how the neutralisation of protonation-sensitive residues in stability-critical regions can significantly enhance resistance to acidic environments and promote the maintenance of a compact fold. [77] The design of Mut4 and related variants therefore fits into a broader strategy in which MNEI serves

as a model system for investigating how local sequence modifications and topological constraints can be combined to tune stability, aggregation propensity and functional resilience. In this sense, MNEI and its engineered mutants exemplify how rational protein engineering can selectively modulate molecular interactions and conformational behaviour while preserving primary biological function.

It is crucial to emphasise that the significance of MNEI goes beyond its application as a sweet-tasting protein. Under destabilising conditions such as acidic pH, high temperature, or high ionic strength, both monellin and MNEI are known to undergo aggregation, forming highly ordered amyloid or amyloid-like fibrils. [53,79] This behaviour has positioned MNEI as a privileged experimental system for studying the molecular mechanisms underlying protein aggregation, allowing systematic investigation of how local stability and sequence-dependent characteristics contribute to cooperative self-assembly processes. Consequently, the extensive characterisation of the stability of MNEI and its engineered variants has historically been driven not only by technological considerations, but also by the broader goal of understanding the determinants and consequences of protein aggregation.

1.2.2 Cooperative and conformational remodelling phenomena

Many protein systems of biotechnological interest cannot be understood solely in terms of monomeric unit stability. In these cases, the observable behaviour emerges from the overall energy landscape of the protein and its ability to explore alternative conformational states, which may involve multiple regions of the same molecule or multiple molecules simultaneously. Phenomena such as aggregation, multimer association and cooperative conformational remodelling are therefore concrete expressions of the collective nature of proteins. Cooperativity is a central element of these processes: relatively small local perturbations, such as variations in pH, temperature or ionic strength, or point mutations and topological changes,

can propagate along the protein structure and trigger non-linear responses, giving rise to global states that are distinct from the native state. This behaviour is typical of systems in which the fold is maintained by a delicate balance between rigid and flexible regions, and in which the energy separation between alternative states is comparable to the available thermal energy. [28,41] In this context, aggregation and domain swapping represent two different but conceptually related ways of cooperative remodelling. In the first case, the transient exposure of normally buried surfaces can lead to the self-association of multiple molecules; in the second, more orderly conformational reorganisations allow the exchange of structural elements between identical subunits. Both phenomena can be interpreted as alternative outcomes of the exploration of the conformational landscape, made accessible by perturbations that alter the equilibrium between native and intermediate states.

1.2.3 Protein Aggregation

Protein aggregation is one of the most evident manifestations of cooperativity in protein systems. Although often associated with neurodegenerative diseases, aggregation is not intrinsically pathological but constitutes a possible response of the system when the native fold loses, even temporarily, its stability. In general terms, aggregation becomes possible when partially unfolded states expose surfaces that are shielded in the native monomer, favouring non-native contacts between molecules. [37,47,80]

From a structural point of view, aggregation can follow different pathways. In some cases, it leads to the formation of amorphous, disordered and often irreversible aggregates; in others, to ordered structures such as amyloid or amyloid-like fibrils, characterised by high supramolecular organisation and well-defined kinetics. The tendency towards one of these paths depends on the sequence, environmental conditions and kinetic stability of the monomer, as well as the presence of segments locally predisposed to aggregation, often rich in β elements or hydrophobic residues. [81,82] A

comprehensive overview of these pathways is well represented by the kinetic model described by Kulikova *et al.* in Figure 4, which distinguishes three main regions of the aggregation process: nucleation, elongation, and saturation. During nucleation, a population of monomers explores partially misfolded microstates, some of which have sufficient complementarity to form dimers and small oligomers. These nuclei are rare and unstable, but they constitute the critical point that triggers the cooperative transition. The elongation phase is characterised by accelerated growth, whereby oligomers recruit monomers and organise themselves longitudinally into more complex structures, protofibrils, giving rise to the typical sigmoid curve of the fibrillation process. Finally, saturation or the plateau phase reflects the slowing down of the process due to the depletion of free monomers and the dynamic equilibrium between growth, fragmentation and reassociation. [83]

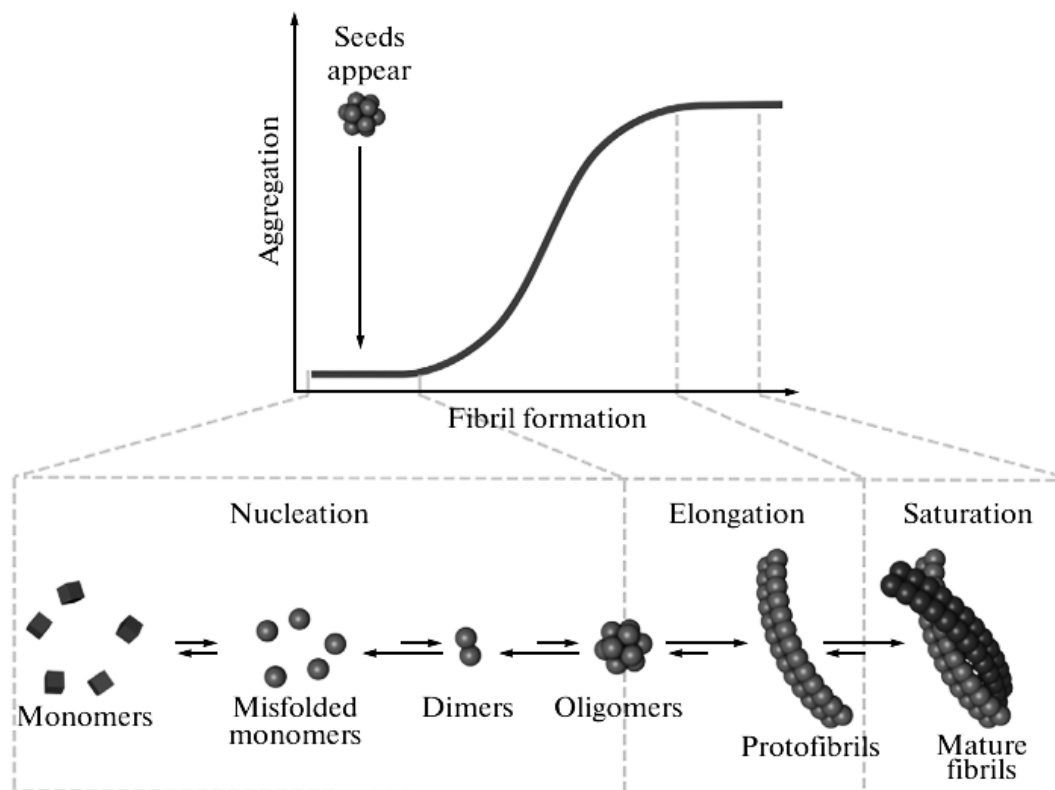


Figure 4. Schematic representation of the aggregation process described by Kulikova *et al.*, highlighting the nucleation, elongation and saturation phases and their correspondence with the characteristic sigmoidal aggregation curve. Figure adapted from [83].

This scheme is particularly useful for interpreting the behaviour of small, compact globular proteins, which are stable under physiological conditions but can undergo aggregation under stressful conditions. A notable example is provided by monellin and its monomeric variant MNEI, which, despite not being associated with any diseases, show a marked propensity for fibrillation under extreme conditions. Pioneering studies have shown that, under extreme conditions such as acidic pH, high temperatures and ionic strength, native monellin can form fibrillar aggregates that bind Congo red and exhibit the typical spectral characteristics of amyloid fibrils. [84–86]

The monomeric variant MNEI has further consolidated the use of this system as an experimental model for the study of aggregation. The fusion of the two chains into a single sequence makes the protein more compact and manageable from a biochemical point of view, while maintaining a marked sensitivity to environmental perturbations such as acidic pH, high ionic strength or increased temperature. [87] Under such conditions, MNEI can access partially unfolded states that temporarily expose regions predisposed to lateral association, following kinetics consistent with the nucleation-elongation model. [79,88] Due to these characteristics, MNEI has been established as a useful platform for analysing how point mutations, topological changes, such as circular permutation, and environmental variations influence the aggregation propensity of a protein. Studies on pH-stabilised variants, super-stable mutants, and circular permutants have made it possible to directly link alterations in the conformational landscape to the kinetics of aggregate nucleation and growth, providing a quantitative overview of the relationship between local stability and cooperative behaviour. [78,89–91]

Overall, aggregation in sweet proteins, and in particular in MNEI and its derivatives, represents a privileged testing bench for studying the relationship between native states, marginally stable intermediates and cooperative self-assembly pathways. The absence of pathophysiological implications makes these systems particularly suitable for isolating basic

biophysical mechanisms, while offering a natural starting point for the analysis of other cooperative conformational remodelling phenomena that emerge in engineered proteins, as discussed in the following sections.

1.2.4 Domain Swapping

Domain swapping is a particular protein association mechanism in which two or more identical monomers exchange equivalent structural portions, giving rise to dimers or oligomers in which each subunit preserves an overall architecture similar to that of the native monomer. Unlike amorphous or fibrillar aggregation, domain swapping represents a form of ordered self-association, in which native interactions are reorganised rather than lost. [92,93] From a mechanistic point of view, domain swapping requires the transient opening of a flexible region of the protein, commonly defined as a hinge loop, which allows a domain or structural segment to partially separate from the rest of the molecule. (Figure 5) This segment can then rearrange itself with an identical copy of the protein, completing the fold in trans rather than in cis. The result is a symmetrical oligomeric complex, stabilised by interactions that mimic those present in the monomeric state but distributed among multiple subunits. [93,94]



Figure 5. Schematic representation of domain swapping mediated by hinge loop opening, in which, in the monomeric state, the protein fold is stabilised by intramolecular interactions and a compact hinge loop. Upon transient opening and extension of the hinge loop, a structural domain can detach and reassociate with an identical subunit, leading to the formation of a domain-swapped dimer. In this configuration, the exchanged domains complete the native fold in trans, and an additional secondary interface contributes to the stabilisation of the oligomeric assembly. Figure adapted from [95].

Domain swapping can be considered an alternative manifestation of the exploration of the conformational landscape, made accessible when the equilibrium between the native state and partially open states is disturbed. Conditions such as pH changes, increased temperature, high protein concentration, or mutations that alter local flexibility can promote the opening of the hinge loop and, consequently, the formation of swapped states. [94,96]

Classical structural observations, based on X-ray crystallography or cryo-EM, continue to be a fundamental starting point for the identification of domain-swapped states. However, recent studies have emphasised that such structures should be interpreted with caution, as they reflect states selected by the crystalline context and are not necessarily representative of the dominant populations in solution. [97–99]

A more comprehensive view emerges from solution techniques such as SEC-MALS, analytical ultracentrifugation, NMR and concentration-dependent spectroscopy, which allow the evaluation of dynamic equilibria between monomeric and oligomeric forms and their dependence on environmental parameters. [100,101]

In recent years, computational analysis has contributed decisively to redefining the meaning of domain swapping, shifting the focus from the description of the final state to the characterisation of the dynamic pathways that make it accessible. Molecular dynamics simulations have shown that regions acting as hinge loops frequently coincide with the main contributors to the low-frequency collective motions of the system, identifiable by PCA or essential dynamics analysis. [102–104] This has led to domain swapping being interpreted as the result of a cooperative reorganisation of movements already encoded in the protein topology, rather than as a radical structural transition induced *ex novo*.

One biological context in which domain swapping has been extensively studied is that of the cystatin superfamily, proteins characterised by a conserved fold consisting of an α -helix resting on a five-stranded β -sheet. [105,106] In this structural context, domain swapping has been observed

across multiple members and under different destabilising conditions, supporting the view that swapped states may be accessible not only from the fully folded native pool but also from partially unfolded or molten-globule-like intermediates. [107] Within this family, human cystatin C represents one of the best characterised examples. Crystallographic studies revealed tight, twofold-symmetric domain-swapped dimers, and subsequent studies have connected the swapping-prone behaviour to amyloidogenic potential and disease-associated variants. [108–111]

Related swapping phenomena have also been described for proteins like Stefins. Stefin A, although mainly monomeric under native conditions, can form domain-swapping dimers under destabilising conditions, and NMR-based analyses have provided insight into how backbone dynamics and local flexibility contribute to the swapping transition. [112] Stefin B (cystatin B) is a further well-documented example, in which swapping occurs through the opening of a specific loop that reorganises into a new β element, symmetrically connecting the subunits. In these systems, the formation of swapped dimers has been linked to greater stability in some conditions and, in other cases, to a different propensity for amyloid aggregation. [82,108,113] In addition to these classical examples in mammals, domain-swapping assemblies have also been reported for other cystatin-like proteins, including cystatin E/M, in which conformational destabilisation can promote the formation of domain-swapping dimers with altered stability properties (for example, structural evidence in PDB entries such as 6FK0). [114]

Monellin and its monomeric variant MNEI share the same basic architecture as cystatins, although they don't have the protease-inhibiting activity typical of this family. [105] From a structural point of view, MNEI has a compact and well-defined fold which, under standard conditions, does not show a spontaneous tendency towards domain swapping. However, its close architectural relationship with domain-swapping-competent proteins makes this system particularly interesting for exploring the extent to which point

mutations or topological changes can alter the accessibility of swapped states. [106]

This is the context for the studies by Gosavi & co., who have shown how engineered variants of monellin can be induced to form domain-swapped dimers through targeted modifications to the length and composition of the loop regions. These studies have shown that the propensity for swapping is not a binary property but critically depends on the local geometry of the hinge and the balance between monomer stability and the accessibility of partially open states. [94,95] A relevant aspect of domain swapping is its possible relationship with aggregation. In some systems, the formation of swapped dimers has been interpreted as a competitive pathway with respect to amyloid fibrillation, capable of trapping aggregation intermediates in well-ordered oligomer states. [115] In other cases, however, swapping may represent a preliminary stage that facilitates the assembly of larger structures. This ambivalence emphasises that domain swapping and aggregation are not mutually exclusive phenomena, but alternative outcomes of the same energy landscape. [94,116]

In the framework outlined in this thesis, domain swapping is therefore considered a specific form of cooperative remodelling, complementary to aggregation, which emerges when mutations, environmental variations or topological changes make alternative conformational states accessible. The analysis of this phenomenon, particularly in systems structurally similar to cystatins such as MNEI and its engineered variants, highlights how controlled perturbations of protein topology can be exploited to modulate conformational equilibria without necessarily compromising ordered structural character.

1.2.5 Circular Permutation as a tool to modulate protein stability and conformational behaviour

Circular permutation is a protein engineering strategy that consists of redefining the position of the N- and C-terminals along the polypeptide

sequence, while maintaining the relative order of the amino acids unchanged.

Technically, the chain is “opened” in a new position, and the original ends are connected by a peptide linker, creating a protein with the same primary composition but a different topology. (Figure 6)

Theoretically, circular permutation doesn't change the information encoded by the sequence, but it does change how that information is translated into a 3D structure. Specifically, this approach alters the distribution of tensions along the backbone, the continuity of secondary interactions, and the connectivity between structural elements. Depending on the fold considered, these changes may have minimal effects or cause significant changes in terms of stability, flexibility, and accessibility of functional regions.

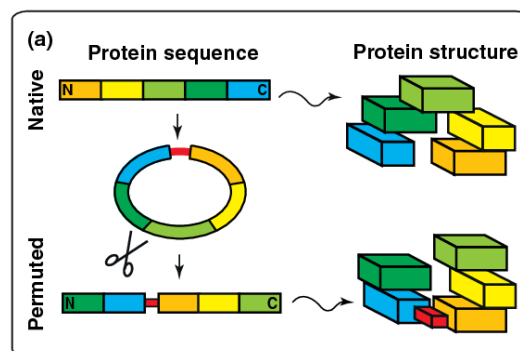


Figure 6. Conceptual illustration of circular permutation, in which the original N- and C-termini of a protein are covalently linked by a peptide linker and new termini are introduced at an alternative position along the sequence, resulting in an unchanged amino acid composition but a different chain topology. Figure adapted from [117].

For this reason, circular permutation is widely used as a tool to probe the topological robustness of a fold and to assess how much the function depends on the position of the polypeptide ends rather than on the overall architecture alone. [20,117]

In the context of protein engineering, circularly permuted mutants have found application both as exploratory tools and as functionally improved variants. In several systems, the relocation of terms has made it possible to

stabilize otherwise fragile regions, reduce pH sensitivity, or increase resistance to critical operating conditions such as high temperatures or ionic strength variations. A particularly relevant aspect of this strategy is that it does not necessarily require mutations in the hydrophobic core: the primary function can therefore be preserved while selectively intervening on stability and conformational behaviour. From the point of view of molecular interactions, circular permutation can influence the accessibility of contact surfaces, the dynamics of specific loops, and the cooperativity between structural elements that are distant in sequence but close in space. In this sense, it represents a natural complement to point mutagenesis, allowing the effects due to the chemistry of residues to be distinguished from those related to the topological organization of the polypeptide chain. [118]

Overall, mutants obtained through circular permutation are a powerful tool for investigating the contribution of topology to the structural and functional properties of engineered proteins. This approach allows us to explore how changes in backbone connectivity can predispose a protein to alternative cooperative behaviours, from the stabilization of the native fold to the emergence of oligomeric states, providing the conceptual basis for the analysis of aggregation and conformational remodelling phenomena discussed in the following sections. [118]

1.3 PIR proteins for stable cell wall architectures

PIR (Proteins with Internal Repeats) proteins are an interesting example of how sequence modularity and topological organization can determine the collective behaviour of a protein through multiple and cooperative interactions with a biological matrix, such as the polysaccharide matrix of the cell wall. In yeasts and many ascomycete fungi, PIRs are, after GPI-anchored proteins, one of the most abundant classes of protein components residing in the cell wall, where they play an essential structural role. [119,120]

The yeast cell wall is a highly organised and dynamic extracellular structure that plays a central role in maintaining cell shape, protecting against

environmental stresses and mediating interactions with the external environment. [121] In *Saccharomyces cerevisiae*, the cell wall (Figure 7) is organised as a multilayered composite matrix in which different polysaccharides and proteins contribute distinct mechanical and functional properties. [121,122] The inner layer is mainly composed of a dense network of β -1,3-glucans branched with β -1,6-glucans and interspersed with chitin, forming a rigid scaffold that bears most of the mechanical load. This inner glucan-chitin matrix is responsible for cell wall strength, resistance to osmotic pressure and overall structural integrity. [55,119,121] The outer layer of the cell wall is enriched in mannoproteins, which are primarily involved in modulating permeability, surface charge and interactions with the environment. [121,123] Most of these proteins are covalently linked to the inner glucan network through glycosylphosphatidylinositol (GPI) anchors, thereby forming a functional interface between the rigid polysaccharide core and the extracellular environment. [124,125] In contrast, PIR proteins constitute a distinct class of cell wall components that are directly linked to β -1,3-glucans via alkali-sensitive ester bonds and are distributed throughout the inner wall rather than being confined to the outer mannoprotein layer. This unique mode of anchoring places PIR proteins in an optimal position to contribute directly to the reinforcement and mechanical stabilisation of the glucan network. [56,126]

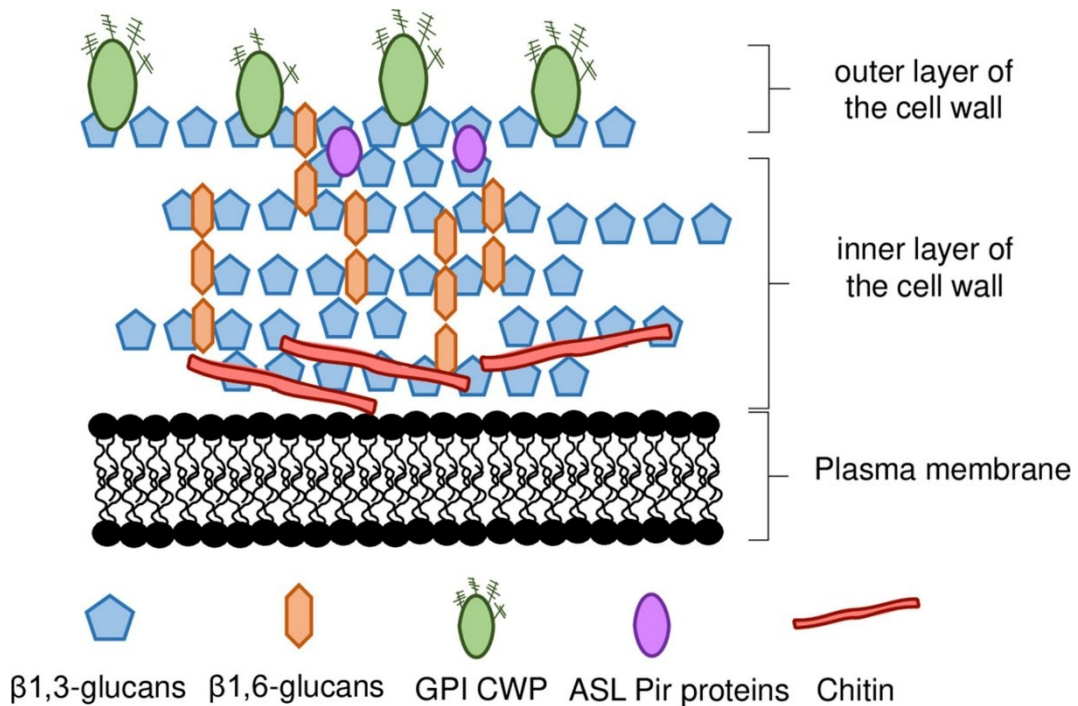


Figure 7. Schematic representation of the organisation of the *Saccharomyces cerevisiae* cell wall, structured as a multilayered polysaccharide matrix, with an inner layer mainly composed of β -1,3-glucans branched with β -1,6-glucans and chitin, and an outer layer enriched in mannoproteins. Most mannoproteins are covalently attached to the glucan network through glycosylphosphatidylinositol (GPI) anchors. In contrast, PIR proteins are linked to β -1,3-glucans via alkali-sensitive ester bonds and are distributed throughout the inner wall, positioning them as structural elements directly involved in reinforcing the glucan network and contributing to cell wall mechanical stability. Figure adapted from [127].

In *Saccharomyces cerevisiae*, the PIR family consists of five members (PIR1-PIR5), which share a key feature: the presence of highly conserved internal repeats, mainly located in the N-terminal region, each containing the consensus motif SQGDGQQA. This motif represents the fundamental functional unit of the family and is directly responsible for the covalent anchoring of PIRs to the β -1,3-glucan matrix of the wall. In particular, a glutamine within the motif is deamidated and involved in the formation of an ester bond with a hydroxyl group of the glucan, creating an extremely stable protein-polysaccharide link. [126,128,129] A central aspect of the PIR family is the variability in the number of internal repeats, which in *S. cerevisiae* ranges from a single unit (PIR4) to ten repeats (PIR2). [130] (Table 1) This heterogeneity suggests a modular organization, in which each repeat can

potentially provide an additional anchor point to glucan. From this perspective, the function of PIRs does not depend on a classic globular fold but emerges from the multiplicity of interactions established along the protein chain with the polysaccharide matrix.

Mutagenesis studies have shown that alteration of key residues, such as glutamine or aspartate, within the SQGDGQQA motif is sufficient to cause loss of wall anchoring and release of the protein into the extracellular medium, confirming that the single module constitutes the minimum functional unit. [126]

Gene locus	Gene name	Number of repeats
YKL164c	PIR1/CCW6	8
YJL159w	PIR2/CCW7/HSP150	10
YKL163w	PIR3/CCW8	8
YJL158c	PIR4/CCW5/CIS3	1 (-2)
YJL160c	PIR5	4

Table 1. List of the five PIR proteins encoded in *Saccharomyces cerevisiae*, reporting gene locus, alternative gene names and the number of conserved internal repeats containing the SQGDGQQA consensus motif. The number of repeats varies among family members, reflecting the modular organisation of PIR proteins. Table adapted from [126].

From a mechanistic point of view, PIRs are best interpreted as distributed reinforcing elements, capable of contributing to the mechanical strength of the cell wall through a network of multiple and cooperative bonds with glucan. Coherently, numerous genetic and biochemical studies have highlighted a critical role of PIRs in maintaining wall integrity under stress conditions, such as osmotic variations, treatments with agents that interfere with glycan synthesis, or thermal stress. [131–133] In this sense, the number and organization of the repeated modules constitute structural parameters that modulate the protein's ability to absorb and distribute mechanical stresses. Interest in PIRs goes beyond *S. cerevisiae*.

Comparative analyses have identified PIR or PIR-like proteins in numerous yeasts and filamentous fungi, often characterized by different architectures but sharing the presence of functionally equivalent repeating motifs. A notable example is provided by the recent study by Montaña-Silva *et al.* on *Neurospora crassa*, which showed that the PIRs of this organism have an “inverted” architecture compared to the yeast model, while maintaining the ability to stabilize the cell wall. [56] This result reinforces the idea that the function of PIRs is mainly defined by the combination of functional repetitions and cooperative interactions with glucan, rather than by a unique and strictly conserved 3D organization.

Considering these elements, PIRs constitute a particularly instructive system for analysing how point mutations, variations in the number of repeated modules, and topological modifications can influence protein-matrix interactions and, consequently, the mechanical properties of a complex biological structure. In this sense, they fit fully into the conceptual framework outlined in this section, offering a structural, non-receptor counterpart to the systems discussed above.

Cell-surface display techniques in *Saccharomyces cerevisiae* are sometimes used in industrial biocatalysis, in the production of bioethanol from biomass, in protein screening and in the construction of artificial enzyme complexes. In this context, the modular nature of PIR proteins and their integration into the cell wall also make this family a potential scaffold for biotechnological applications based on engineered microbial surfaces, where stability and mechanical robustness are key requirements. [134,135]

1.4 Aim of the thesis

The overall aim of this thesis is the design and structural, biophysical and dynamic characterisation of engineered proteins with potential relevance for biotechnological applications, with a particular focus on how point mutations and topological modifications influence stability, molecular interactions and cooperative behaviours. In particular, the work focuses on two main protein systems, selected for their structural features and biotechnological relevance to investigate controlled remodelling of molecular properties.

The first system is represented by monellin and its monomeric variant MNEI, used as model globular proteins to study the relationship between stability, conformational dynamics, and propensity to aggregate. Although monellin is historically known as a sweet-tasting protein, in this thesis, MNEI and its engineered variants are not studied in the context of sweet perception. Instead, they are exploited as well-defined protein scaffolds to study how topological rearrangements and point mutations influence the accessibility of partially unfolded states and the emergence of cooperative self-assembly processes. In particular, the circularly permuted variants of MNEI and the Mut4 (MNEI-E23Q) mutant, designed to reduce sensitivity to protonation in a critical region of the MNEI protein, are characterized to assess how local and global perturbations affect structural stability, molecular flexibility and aggregation behaviour. This approach is motivated by the growing interest in protein aggregation and amyloid-like self-assembly as programmable mechanisms for the design of functional biomaterials, where controlled fibril formation can be exploited to generate ordered and mechanically robust architectures. [136,137] The second system investigated in this thesis concerns PIR (Proteins with Internal Repeats) proteins of *Saccharomyces cerevisiae*, a family of structural cell wall proteins characterised by sequence modularity and covalent anchoring to the glucan matrix. The interest in PIR proteins arises from their role as distributed reinforcing elements that contribute to the mechanical stability and integrity of the yeast cell wall, a property of fundamental relevance for cell physiology and for industrial biotechnology. This study aims to understand how the number and

organization of internal repeats in these proteins can modulate protein-matrix interactions in cell wall. By analysing engineered constructs containing a reduced number of PIR repeats, the work aims to assess whether a single modular unit is sufficient to preserve the main structural and functional features attributed to PIR proteins.

2 Materials

2.1 Strains and Plasmids

2.1.1 *E. coli* strains

Strain	Genotype	Reference
TOP10	F ⁻ mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80lacZ Δ M15 Δ lacX74 recA1 deoR araD139 Δ (ara-leu)7697 galU galk rpsL(Str ^R) endA1 nupG	Invitrogen
DH5α	F ⁻ ϕ 80lacZ Δ M15 Δ (lacZYA- argF)U169 recA1 endA1 hsdR17(rK ⁻ mK ⁺) phoA supE44 λ - thi-1 gyrA96 relA1	Life Technologies
BL21 (DE3)	F ⁻ ompT gal dcm lon hsdSB(rB ⁻ mB ⁻) λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) [malB+]K-12(λ S)	Novagen / NEB
BL21-Gold (DE3)	F ⁻ ompT hsdSB(rB ⁻ mB ⁻) gal dcm Δ lon λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) endA Hte [malB+]K-12(λ S)	Agilent
SHuffle T7	F ⁻ ompT hsdSB(rB ⁻ mB ⁻) gal dcm lacY1 Δ trxB Δ gor (pTn7t-T7 lysY, lacI, T7 RNA polymerase) KanR	NEB
ArcticExpress (DE3)	F ⁻ ompT hsdSB(rB ⁻ mB ⁻) gal dcm lon λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) endA Hte [cpn10, cpn60, CamR]	Agilent

Table 2. *E. coli* strains used in this work.

2.1.2 Plasmids

Strain	Genotype	Source
MNEI	AmpR pET22b(+) MNEI	Structural Biology Lab, Prof. D. Picone, Dept. of Chemical Sciences, Univ. of Naples “Federico II”
Mut4	MNEI_E23Q	Structural Biology Lab, Prof. D. Picone, Dept. of Chemical Sciences, Univ. of Naples “Federico II”
Perm1	KanR pET-26b(+) perm1	Biocat
Perm2	KanR pET-26b(+) perm2	Biocat
Perm3	KanR pET-26b(+) perm3	Biocat
GFP- PIR1r1	KanR pET-28a(+) Pir1_r1	Structural Biochemistry Lab, Prof. L.-O. Essen, Dept. of Chemistry, Philipps-Univ. Marburg (Germany)
GFP- PIR1r1-8	KanR pET-28a(+) Pir1_r1-8	Structural Biochemistry Lab, Prof. L.-O. Essen, Dept. of Chemistry, Philipps-Univ. Marburg (Germany)
PIR2r1	KanR pET-28a(+) Pir2_1_HSP150/YJL159W	Twist Bioscience
PIR4r1	KanR pET-28a(+) Pir4_1	Twist Bioscience
PIR5r1	KanR pET-28a(+) Pir5_1	Twist Bioscience

Table 3. Plasmids used in this work.

2.2 Enzymes

Enzyme	Reference
Lysozyme	Sigma-Aldrich
DNase I from bovine pancreas	Roche

Table 4. Enzymes used in this work.

2.3 Chemicals

All chemicals and reagents used in this study were of analytical grade and were purchased from Affymetrix, Biozym Scientific GmbH, Carl Roth (Karlsruhe, Germany), Carlo Erba Reagents (Milan, Italy), Fisher Scientific, Sigma-Aldrich/Merck (Darmstadt, Germany), VWR International, and, unless otherwise specified. All culture media, buffers, and solutions were prepared using double-distilled water (ddH₂O) and adjusted to the desired final volume. When required, pH adjustments were made using HCl or NaOH, and all solutions were filtered or sterilized according to their intended use.

2.3.1 Buffers and solutions

Buffer	Recipe
1× SDS buffer or Laemmli sample buffer	62.5 mM TRIS, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.01% (w/v) bromophenol blue, pH 6.8
5× Native loading buffer	1 mg methyl green, 50% glycerol, 1.5 M KOAc, pH 4.3
Blocking buffer	TBS buffer+ 3% BSA
Coomassie Brilliant Blue R-250	0.32% in 40% EtOH and 8% Glacial acetic acid
Competent cells buffer	100 mM CaCl ₂ solution and

	100 mM CaCl ₂ solution + 15 % (v/v) glycerol
Desalting solution	100 mM acetic acid (CH ₃ COOH)
Destaining solution	5% Glacial acetic acid, 28% Isopropanol, 67% Water
ECL-solution	100 mM TRIS, 2.5 mM 3-Aminophthalhydrazide (Luminol), 0.45 mM 4-hydroxycinnamic acid (p-coumaric acid), 5.4 mM 30 % Hydrogen peroxide (H ₂ O ₂), pH 9.2
Imidazole cleaning-elution solution	500 mM Imidazole
Native running buffer	0.35 M β-alanine, 0.14 M acetic acid, pH 4.3
Ni-NTA elution buffer	250 mM Imidazole + PIR optimized buffer solution
Perm buffer solution	50mM NaOAc pH 5.5
Perm cleaning solution	2 M NaCl + 50 mM Perm buffer solution
Perm elution buffer	0, 50, 100, 200 mM NaCl + 50 mM Perm buffer solution
PIR buffer solution	50 mM TRIS, 200 mM NaCl, pH 7
PIR optimized buffer solution	50 mM TRIS, 300 mM NaCl, pH 8
Ponceau S	5% v/v Glacial acetic acid, 0.1% wt Ponceau red dye
TBS buffer	20 mM Tris, 150 mM NaCl, pH 7.4
TBST buffer	TBS buffer + 0.1% Tween-20
Transfer buffer	25 mM TRIS, 192 mM glycine, 20% methanol, pH 8.3

Table 5. Buffers and solutions used in this work.

2.3.2 Culture media

Culture medium	Composition	Reference
Lysogeny Broth (Miller)	0,5% Yeast extract 1% Tryptone 1% NaCl	Sigma-Aldrich
Terrific Broth	1,2% Tryptone 2,4% Yeast extract 1,254% K ₂ HPO ₄ 0,231% KH ₂ PO ₄ 0,4% Glycerol (v/v)	Carl Roth

Table 6. Media used in this work.

All culture media used in this study were sterilized in an autoclave (121 °C at 3 bar) to render them sterile, used immediately, or stored at room temperature. Solid media were obtained by adding agar to the corresponding nutrient broths to a final concentration of 2% (w/v) and, after sterilization, were stored at 4 °C.

2.3.3 Inductors

Inductor	Concentration	Reference
Lactose	5 mM	Sigma-Aldrich
IPTG 0.1 mM	0.1 mM – 0.4 mM	Sigma-Aldrich

Table 7. Protein expression inductors used in this work.

2.3.4 Antibiotics

Antibiotic	Concentration	Origin
Ampicillin	100 µg/mL	Sigma-Aldrich
Kanamycin	50 µg/mL	Carl Roth

Table 8. Antibiotics used in this work.

2.3.5 Antibodies

Antibody	Host species	Application (dilution)	Source
Anti-His tag	Mouse monoclonal	WB (1:3000)	Merck
Anti-PIR2	Rabbit polyclonal	WB (1:1000)	Molecular Genetics Lab, Prof. H.-U. Mösch, Dept. of Biology, Philipps-Univ. Marburg (Germany)
Goat anti-rabbit IgG (HRP-conjugated)	Goat	WB (1:10000)	Sigma-Aldrich

Table 9. Antibodies used in this work.

2.4 Computational resources and software

In this study, high-performance computing (HPC) infrastructures and local servers were used to perform molecular dynamics simulations. The use of two complementary platforms allowed us to work with high computing power at different times.

2.4.1 Servers and computing infrastructure

Two facilities are employed, as described in the following paragraphs.

2.4.1.1 Marconi100 on CINECA HPC

Some of the MD simulations concerning circularly permuted proteins were conducted on the Marconi100 cluster, managed by the CINECA consortium (Bologna, Italy), as part of the HP10C9MHVF computing project with ISCRA access (IsCa0_PermMD, duration 9 months, 6000 standard hours, and 1 TB of workspace). Marconi100 is one of Italy's leading Tier-0 HPC

infrastructures, designed to support highly parallelized scientific applications. The system is based on IBM POWER9 architecture with NVIDIA V100 GPUs and Mellanox InfiniBand dual-rail high-speed networking, with a peak power of approximately 29 PFlops.

The operating environment is a Linux system with SLURM job scheduler and software modules dedicated to parallel and accelerated GPU computing, ensuring compatibility and high performance with leading molecular simulation packages.

2.4.1.2 Local server at UniMarburg

A big part of the simulation and analysis activities was conducted on the computing infrastructure of Prof. Lars-Oliver Essen (Department of Chemistry, Philipps-Universität Marburg, Germany), dedicated exclusively to the research group's activities.

The system, called *ak64cluster*, is managed using SLURM scheduling software and has been configured to flexibly support medium- and long-term molecular simulations. Unlike shared HPC resources, this server offered unlimited access in terms of time and number of jobs, ensuring a highly stable and available environment for experimentation, optimization, and continuous simulation execution.

The cluster consists of three main nodes: *ak64server*, *ak64base*, and *ak64work1*. These are interconnected in a local area network (internal LAN) and managed by a control node (*ak64server.chemie.uni-marburg.de*), which also acts as a login node and SLURM scheduler.

Node	CPU	Physical cores/threads	Memory (RAM)	GPU
ak64server	2	64 cores / 128 threads	248 GB	2 (Turing)
ak64base	1	12 cores / 24 threads	116 GB	1
ak64work1	2	10 cores / 20 threads	124 GB	2

Table 10. Hardware specifications of the *ak64cluster* nodes used in this work.

The GPUs are Turing-type, corresponding to NVIDIA Turing architecture (RTX 2080 Ti or equivalent). The operating system installed is a Linux Ubuntu 24.04 LTS distribution or equivalent Debian-based distribution, with kernel 4.x, compatible with CUDA and OpenCL libraries for parallel computing. In addition, the cluster has 4 partitions configured for different purposes, specifically the *normal* partition with 2 nodes and up to 16 CPUs/node for a maximum execution time of 24 hours, the *base* partition with a single node and up to 16 CPUs/node with unlimited time for testing activities, the *long* partition, also with unlimited duration, with 2 nodes and up to 16 CPUs/node, used for long-term simulations, and finally the *huge* partition with a limit of 48 CPUs/node and time limited to 24 hours dedicated to intensive memory use during minimization operations prior to long-term simulations.

The infrastructure was then employed to prepare and validate the systems with the building of solvation boxes, charge control, and preliminary minimization, in order to perform short-term simulations in an NVT/NPT environment for stability and input parameter testing, as well as to perform long-duration simulations and proceed with post-simulation analysis in Jupyter Notebook on the trajectories (methodology described in detail in section 3.7.2). Continuous access and the opportunity to run jobs without time limits made this platform extremely efficient for exploratory trials and complex analyses that demanded extended integration times or long post-processing procedures. The operating environment is based on Linux, with a stable installation of SLURM 19.x, CUDA toolkit, and AmberTools/AMBER 24 compiled locally. This configuration allowed complete control on software dependencies and versions, ensuring coherence between local tests and HPC productions.

2.4.2 Software, algorithms and web tools

Name	Version	Source / Reference
AlphaFold 3	2024 release	[33]
AlphaFold2 via Colab	Colab notebook implementation	[24,138]
Amber	v 24 / AmberTools 24	[139]
Chai Discovery	2025 beta	[35]
ChromLab Software	Standard / Security Edition (for Bio-Rad NGC system)	[140]
Coot	v 0.9.8	[141]
Expasy Bioinformatics Portal	Online suite	[142]
H++ Server	v 3.2	[143]
OriginPro	v 8.0	[144]
Phenix	v 1.21.1-5286	[145,146]
PyMOL Molecular Graphics System	v 3.0.0	[147–149]
Sequence Alignment (VectorBuilder)	Online tool	[150]
SnapGene Viewer	v. 8.2.0	[151]
UCSF Chimera	v 1.17.3	[152]
VMD	v 2.0.0a7	[153]

Table 11. Summary of the software packages, algorithms and web-based tools used throughout this work for computational modelling, structural analysis and data processing, with version information and corresponding references.

3 Methods

3.1 Molecular Biology

3.1.1 Circular permutation

Circular permutation represents an advanced molecular biology approach designed to modify the sequential topology of a protein without altering its amino acid composition. The technique involves the covalent linkage of the native N- and C-termini through a short, flexible peptide linker, while introducing novel termini at a defined internal position within the original polypeptide chain. Although the amino acid content remains unchanged, the linear order of residues is rearranged, generating a protein variant that exhibits an alternative topological connectivity of the same structural framework.

At the genetic level, circular permutation is commonly achieved through recombinant DNA methodologies. In this process, the DNA segments encoding the original termini are joined via an oligonucleotide linker, typically comprising 3 to 10 amino acids, often enriched in glycine and serine residues to ensure conformational flexibility, while new start and stop codons are tactically introduced at the desired internal position of the protein. While the sequence is subjected to this amino acid reorganisation, the elements of secondary structure remain unchanged, with only the order of placement changing.

3.1.2 Plasmid preparation and Verification

Plasmid DNA was extracted from cultures of *E. coli* cloning strains (TOP10 or DH5 α ; see Section 2.1.1) using the QIAprep Spin Miniprep Kit (Qiagen, Germany) following the manufacturer's standard protocol. The maintenance of selective pressure throughout bacterial growth was ensured by supplementing the LB medium with the appropriate antibiotic (see Section 2.3.4; depending on the plasmid backbone). Overnight cultures were prepared by inoculating a single colony into 3-5 mL of selective medium and incubating at 37 °C under agitation (180-200 rpm) for about 12-16 hours.

Cells were harvested by centrifugation and subjected to alkaline lysis according to the Qiagen procedure, which is based on the method described by Birnboim and Doly (1979). [154] Briefly, bacterial pellets were resuspended in a buffer containing RNase A to degrade RNA, lysed in an alkaline SDS-containing solution to denature cellular components, and neutralized with a potassium acetate-based buffer to promote plasmid renaturation and precipitate genomic DNA and proteins. The cleared lysate was applied to a silica membrane column, allowing selective binding of plasmid DNA, followed by washing to remove residual contaminants and ethanol. (Figure 8) Plasmid DNA was finally eluted in 30-50 μL of elution buffer (10 mM Tris/HCl, pH 8.5) or sterile deionized water. Purified plasmids were stored at $-20\text{ }^{\circ}\text{C}$ until further use in downstream procedures such as restriction digestion, sequencing, or bacterial transformation.

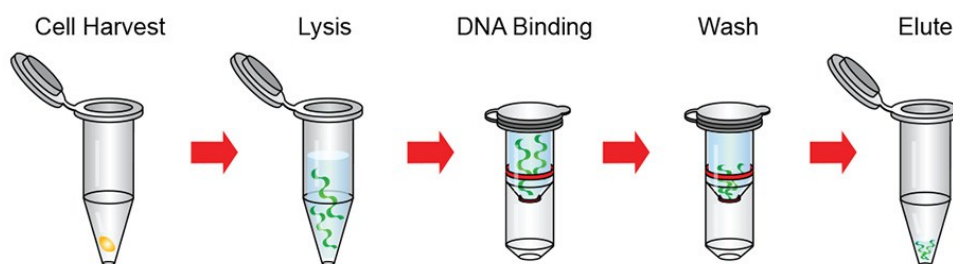


Figure 8. Schematic representation of the plasmid DNA extraction workflow based on alkaline lysis and silica membrane purification, including cell harvesting, lysis, selective binding of plasmid DNA to the column, washing and final elution. Figure adapted from [155].

3.1.3 Sanger sequencing, verification of the DNA and protein concentration

The correctness of the recombinant plasmid constructs was confirmed through Sanger sequencing outsourced to Microsynth Seqlab GmbH (Göttingen, Germany). Sequencing reactions were prepared according to the provider's submission requirements, using 12 μL of purified plasmid DNA at the recommended concentration (50-100 $\text{ng}/\mu\text{L}$). The obtained DNA

sequences were directly supplied in digital format and analysed using SnapGene for visualization and annotation.

The region of interest, corresponding to the engineered construct between specific restriction sites (e.g., BamHI, EcoRI, or equivalent depending on the plasmid), was extracted and translated into the corresponding amino acid sequence using the ExPASy Translate Tool (SIB Swiss Institute of Bioinformatics). [142] The translated sequence was then compared to the expected construct through multiple sequence alignment using Clustal Omega (EMBL-EBI) to verify the correct reading frame, sequence integrity, and absence of mutations.

In addition, the theoretical physicochemical parameters of the confirmed protein sequence, such as the molecular weight, extinction coefficient, and isoelectric point, were calculated using ProtParam (ExPASy). [142] The concentration and purity of the plasmid preparations were subsequently assessed spectrophotometrically using a NanoDrop 2000c (Thermo Fisher Scientific), recording absorbance values at 260 and 280 nm to determine nucleic acid concentration and the A_{260}/A_{280} ratio as an indicator of sample purity.

3.1.4 Preparation of chemical competent *E. coli* cells

Chemically competent *Escherichia coli* cells were prepared according to the calcium chloride method originally described by Mandel and Higa (1970) and subsequently optimized by Hanahan (1983). [156,157] This approach exploits transient changes in the bacterial cell envelope induced by divalent cations at low temperature, facilitating DNA uptake during transformation.

A single *E. coli* colony was inoculated into 5 mL of LB medium and grown overnight at 37 °C with agitation (180-200 rpm). The following day, 1 mL of the overnight culture was used to inoculate 50 mL of fresh LB medium, which was incubated under the same conditions until the culture reached an optical density at 600 nm (OD_{600}) of 0.5-0.6, corresponding to the mid-exponential growth phase.

Cells were harvested by centrifugation at 4000 g for 10 minutes at 4 °C, and the pellet was gently resuspended in 15 mL of ice-cold 100 mM CaCl₂ solution (operating on ice). The suspension was incubated on ice for 30-60 minutes to promote membrane conditioning. After a second centrifugation under the same conditions, the supernatant was discarded, and the cells were resuspended in 2 mL of cold 100 mM CaCl₂ containing 15 % (v/v) glycerol to enhance stability during freezing.

Aliquots of 50 µL were dispensed into pre-chilled microtubes, flash-frozen in liquid nitrogen, and stored at -80 °C until use. The transformation efficiency of each batch was periodically assessed using a standard plasmid control to ensure consistent performance.

3.1.5 Transformation of *E. coli*

Transformation of chemically competent *E. coli* cells was performed using the strains described in Section 2.1.1, mainly BL21-Gold (DE3) and SHuffle T7, selected according to the experimental purpose and protein expression requirements.

For each transformation, an aliquot of 50 µL of competent cells was thawed on ice and mixed gently with 1-5 ng of purified plasmid DNA. The mixture was incubated on ice for 30 minutes to allow DNA adsorption to the cell surface, followed by a brief heat shock at 42 °C for 90 seconds to promote DNA uptake. Immediately afterward, cells were returned to ice for 2 minutes to stabilize the membrane transition.

Subsequently, 950 µL of pre-warmed LB medium (without antibiotics) were added to the transformation mix, and the cells were incubated at 37 °C with shaking (180 rpm) for 1 hour to allow expression of antibiotic resistance genes. Following recovery, 100 µL of the culture was plated on LB agar supplemented with the corresponding antibiotic for selection (see Section 2.3.4), while the remaining cells were pelleted by brief centrifugation (6000 g, 2 minutes) resuspended in 100 µL of fresh LB medium and plated onto the same selective medium to maximize colony yield. Plates were incubated overnight at 37 °C, or at room temperature when slower growth was needed.

3.2 Protein biochemistry

3.2.1 Cultivation of *E. coli*

3.2.1.1 Test expression

For preliminary expression assays, *E. coli* expression strains (see section 2.1.1) were streaked on LB agar plates supplemented with the appropriate antibiotic and incubated overnight at 37 °C. A single isolated colony was then inoculated into 3 mL of liquid LB medium, containing the antibiotic at a 1:1000 dilution from the corresponding 1000× stock solution, and incubated overnight at 37 °C with shaking (180 rpm).

The following day, 0.5 mL of the overnight preculture was used to inoculate 50 mL of fresh LB or TB medium contained in autoclaved 250 mL baffled flasks. Antibiotics were added at the same final concentrations as in the preculture step. Cultures were grown at 37 °C with orbital shaking (180 rpm) until reaching an optical density (OD₆₀₀), typically corresponding to mid-logarithmic phase. When the optimal optical density was reached, the cultures were immediately transferred onto ice to reduce metabolic activity while the incubator temperature was lowered to the selected expression temperature. Protein expression was then induced by adding the appropriate inductor (see Section 2.3.3), and the cultures were subsequently incubated at 180 rpm under the selected expression conditions and monitored until the OD₆₀₀ reached approximately 0.3-0.4, corresponding to the late induction phase. Then, bacterial cells were harvested by centrifugation at 4000 rpm for 20 minutes at 4 °C, dividing the volume into two 50 mL Falcon tubes. The supernatant was discarded, and the cell pellets were resuspended using the lysis buffer. Specifically, 1 mL was used for each 50 mL Falcon tube, which was then transferred to 2 mL tubes with screw caps and O-rings pre-filled to one-third of their volume with sterile 0.5 mm glass beads. The samples were immediately frozen in liquid nitrogen and then stored at -4°C until mechanical disruption by glass bead homogenization using the FastPrep system, as described in detail in Section 3.2.2.1.

3.2.1.2 Measuring the optical density of liquid cultures

The growth of *Escherichia coli* cultures was monitored spectrophotometrically by measuring the optical density at 600 nm (OD₆₀₀) using a calibrated spectrophotometer. Prior to each measurement, the instrument was blanked against the corresponding sterile culture medium to eliminate background absorbance. Measurements were performed in disposable polystyrene cuvettes with a 1 cm optical path length, using aseptically withdrawn culture aliquots to prevent contamination.

During the cultivation phase, OD₆₀₀ readings were recorded at regular intervals to monitor bacterial growth kinetics and to determine the optimal point for induction. Induction was typically initiated during the exponential phase, corresponding to an OD₆₀₀ of approximately 0.5-0.6, as determined empirically for each expression system. For dense cultures (OD₆₀₀ > 0.8), samples were appropriately diluted with sterile double-distilled water (ddH₂O) to ensure that absorbance readings remained within the linear detection range of the spectrophotometer; final OD₆₀₀ values were calculated by multiplying the measured absorbance by the respective dilution factor.

Following induction, bacterial growth was continuously monitored to assess cell proliferation dynamics and the progression of recombinant protein expression over time. Post-induction optical density measurements (OD₆₀₀) were performed using 1:10 diluted samples to maintain absorbance values within the linear detection range of the spectrophotometer. Recorded absorbance values were subsequently multiplied by the corresponding dilution factor to obtain the actual OD₆₀₀.

Monitoring was continued until cultures reached an OD₆₀₀ of approximately 0.3-0.4, which was indicative of the post-induction growth phase. These measurements provided insight into potential growth arrest, recovery, or overgrowth phenomena associated with the different experimental conditions, thereby enabling the evaluation of inducer-dependent effects on cellular physiology and protein yield. All OD₆₀₀ determinations were

performed in triplicate to ensure the accuracy, consistency, and reproducibility of the recorded growth profiles.

3.2.1.3 Large-scale expression

Large-scale expression cultures were prepared according to the same procedure described for the preliminary expression assays (see Section 3.2.1.1), with adjustments in total culture volume and flask capacity. In detail, for large-scale cultures starting from 2 L of sterile LB or TB medium, each supplemented with the appropriate antibiotic (1:1000 dilution from a 1000× stock solution) and in 5 L flasks, 40 mL of overnight culture was inoculated. The cultures were incubated at 120 rpm, induced and harvested under the same conditions as described for small-scale testing (see Section 3.2.1.1).

3.2.2 Cell lysis

Bacterial cells were disrupted by mechanical means using one of several methods, depending on the protein construct, project requirements, and available instrumentation: bead-beating with glass beads (FastPrep® system, MP Biomedicals), ultrasonic sonication, high-pressure homogenization (Avestin® Emulsiflex-C5), or French press disintegration (Thermo Scientific™ French Press Cell Disrupter). All lysis procedures were performed at 4 °C or on ice to prevent protein denaturation and minimize proteolytic degradation.

Before lysis, in the case of large-scale production, the cell pellets obtained after centrifugation were resuspended in an appropriate volume of lysis buffer, adjusted according to the optical density (OD₆₀₀) of the culture at the time of harvesting. The resuspension volume was calculated to normalize the biomass to approximately 10 OD units per millilitre of buffer.

This approach ensured a consistent cell density and homogeneous viscosity across preparations, facilitating reproducible lysis efficiency regardless of the original culture volume.

Following resuspension, the samples were supplemented with lysozyme (to weaken the peptidoglycan layer), phenylmethylsulfonyl fluoride (PMSF) (as a serine protease inhibitor), and DNase I (to reduce sample viscosity by digesting released nucleic acids). The suspensions were gently mixed on ice for 10-15 minutes before proceeding to the selected mechanical disruption protocol.

3.2.2.1 Mechanical disruption using glass beads

Cell disruption by bead milling was performed using the FastPrep® homogenisation system (MP Biomedicals). The bacterial pellets frozen in 2 mL screw-cap tubes, as described in section 3.2.1.1, were left on ice for 5 minutes after the addition of enzymes. Lysis was then performed using three to five cycles at 6.0-6.5 m/s for about 30 seconds each, with cooling intervals of at least 5 minutes on ice between cycles to prevent heat-induced denaturation of the target proteins. The duration and number of cycles were optimised based on the expression system and sample viscosity. After homogenisation, the tubes were centrifuged (15000 g for 30 minutes at 4 °C) to allow sedimentation of beads and coarse debris, and the resulting supernatant was carefully transferred to an Eppendorf tube after filtration.

3.2.2.2 Ultrasonic cell disruption

Cell pellets were disrupted by ultrasonic homogenization using a Bandelin Sonopuls HD 3200 sonicator equipped with a MS 73 microtip probe. Prior to sonication, the cell pellet obtained after centrifugation of the bacterial culture was resuspended in buffer, to achieve a homogeneous cell suspension suitable for protein extraction.

The sonication process was carried out in an ice bath to prevent thermal denaturation of proteins. Pulsed sonication was applied for a total of 20 minutes, with alternating cycles of 30 seconds ON and 30 seconds OFF to ensure efficient cell disruption while minimizing heat accumulation. The amplitude of the sonicator was maintained at 60% (about 60-70 W output power). After each sonication session, the probe tip and connecting parts

were immediately rinsed with distilled water to remove residual cellular material, followed by immersion in 70% ethanol for several minutes to ensure complete decontamination. The probe was then gently wiped with a lint-free tissue and allowed to air-dry. To prevent cross-contamination between samples, the microtip was occasionally sonicated in distilled water for 30-60 seconds to dislodge any residual particles adhering to the surface. The cleaned probe was finally stored in a dust-free container at room temperature. Prior to subsequent use, the sonicator amplitude setting and probe integrity were verified to ensure consistent energy delivery and prevent tip erosion or metal particle release into the samples.

3.2.2.3 High-pressure cell disruption (French Press)

For certain preparations, bacterial cells were disrupted using a Thermo Scientific™ French Press Cell Disrupter. The instrument employs a stainless-steel pressure cell in which cell suspensions are forced through a narrow orifice under high pressure, generating shear forces sufficient to rupture the bacterial membranes.

Before the operation, the pressure cell and piston assembly were lubricated with high-vacuum grease and pre-cooled by immersion in an ice bath to maintain low temperature throughout the process. The cell suspension (typically 20-40 mL) was loaded into the pressure chamber, and the system was operated at an internal pressure of approximately 110-140 MPa (1100-1400 bar). The homogenate was collected in pre-chilled 50 mL Falcon tubes. Each sample was passed through the press three times, or until a visible change in colour and viscosity of the suspension indicated efficient cell disruption. Between cycles, the piston and chamber were cooled and cleaned to ensure consistent performance and prevent cross-contamination. After use, the pressure cell and piston were immediately rinsed with distilled water to remove residual cell debris and buffer salts. The system was then flushed with 70% ethanol to ensure complete sanitation, followed by a final rinse with deionized water to remove traces of solvent. All components were thoroughly dried with lint-free tissues, and the piston

was lightly lubricated with high-vacuum grease before reassembly to prevent oxidation and ensure smooth operation during subsequent use. The pressure cell was stored disassembled at 4 °C, covered with Parafilm to avoid dust contamination.

3.2.2.4 High-pressure homogenization (Avestin® Emulsiflex-C5 System)

In large-scale preparations, bacterial cell disruption was carried out using an Avestin® Emulsiflex-C5 high-pressure homogenizer (Avestin Inc., Canada). (Figure 9)

The system operates by forcing the cell suspension through a narrow interaction chamber under high pressure, generating intense shear and cavitation forces that efficiently disrupt the *E. coli* cell envelope. All operations were conducted with appropriate protective equipment, as the device operates under extremely high pressure and may eject liquids if mishandled. Before and after the use, the entire fluidic circuit was sanitized by flushing with 70% (v/v) ethanol, followed by extensive rinsing with sterile double-distilled water until all ethanol residues were removed.

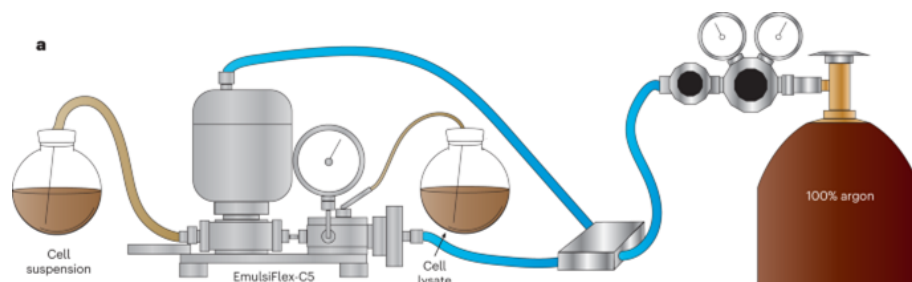


Figure 9. Schematic representation of high-pressure cell disruption using the Emulsiflex-C5 homogenization system, in which a bacterial suspension is forced through a narrow interaction chamber under high pressure, generating shear and cavitation forces that lead to efficient cell lysis. Figure adapted from [158].

The system was then pre-cooled by circulating ice-cold buffer to prevent overheating and to ensure the hydraulic circuit never ran dry. The bacterial suspension was loaded into the stainless-steel feed tank and the outlet tube

was initially redirected into the inlet funnel to allow internal circulation and homogenization of the sample. Cell lysis was performed at a constant pressure of compressed argon at 70-100 MPa (10,000-15,000 psi), in accordance with the manufacturer's recommendations for *E. coli*. Each sample underwent multiple complete passes through the homogenizer for approximately 20 minutes to ensure complete disruption and the absence of visible clumps. The process was conducted at 4°C, with the lysate collected in refrigerated tubes to minimize heat-induced protein denaturation. The compressed argon supply was then shut off and the tank sealed for maintenance.

3.3 Protein purification

3.3.1 Centrifugation and Clarification of Crude Extracts

Centrifugation steps were performed to separate cellular debris and clarify protein extracts prior to downstream purification. Cell pellets obtained from large-scale expression cultures were collected by centrifugation using a Beckman Coulter Avanti™ J-25S XP centrifuge equipped with a JA-14 rotor, operating at 8000 g for 30 minutes at 4 °C. The resulting cellular pellet was retained for subsequent lysis procedures, whereas the supernatant was discarded.

Following cell disruption (see Section 3.2.2), the crude lysates were clarified by centrifugation at 18000 g for 30 minutes at 4 °C using a JA-25.50 fixed-angle rotor. After centrifugation, the pellet containing insoluble material was discarded, while the clarified supernatant containing the soluble protein fraction was collected and filtered through 0.45 µm filters before the chromatographic purification, as described in the following sections.

3.3.2 Ni-NTA Affinity chromatography

3.3.2.1 Small-scale purification using spin Ni-NTA columns

Small-scale purification of His-tagged proteins from test expression assays was carried out using Ni-NTA Spin Columns (Qiagen) operated in a

benchtop centrifuge according to the manufacturer's recommendations, with minor procedural adjustments. (Figure 10)

Before loading the sample, each column was centrifuged once at 700 g for 2 minutes to remove the storage buffer, then equilibrated with equilibration buffer (without imidazole) to establish binding conditions. Clarified lysates were applied in 600 μ L aliquots and centrifuged, allowing gravity-assisted binding of the His-tagged proteins to the nickel-charged resin. The flow-through fraction (FT) was collected for subsequent analysis.

The columns were then sequentially washed to remove non-specifically bound proteins using two distinct wash buffers:

- Wash 1: 20 mM imidazole in equilibration buffer.
- Wash 2: 35 mM imidazole in equilibration buffer.

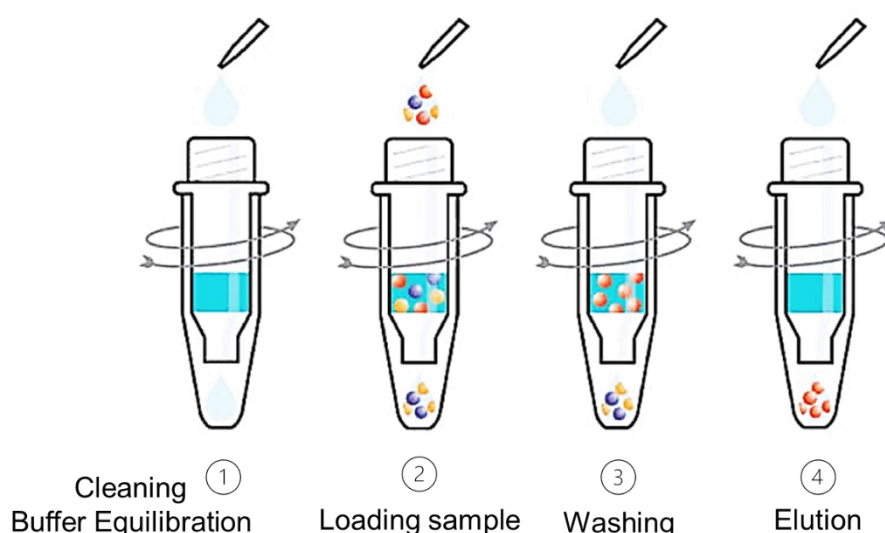


Figure 10. Schematic representation of small-scale purification of His-tagged proteins using Ni-NTA spin columns, illustrating sequential steps of column equilibration, sample loading and binding, washing at increasing imidazole concentrations, and final elution of the target protein.

Figure adapted from [159].

Each wash was followed by centrifugation at 700 g for 2 minutes, and the eluates were collected for SDS-PAGE analysis.

Target proteins were finally eluted using elution buffer containing 500 mM imidazole, and the eluate was collected by centrifugation under the same

conditions. Aliquots from all purification steps (flow-through, washes, and eluate) were stored at -20 °C and analysed by SDS-PAGE to assess purity and expression efficiency.

3.3.2.2 Medium- to large-scale purification using Ni-NTA columns

For preparative purification, 1 or 5 mL Protino® Ni-NTA columns were employed, exploiting immobilized metal affinity chromatography (IMAC) for the selective purification of His-tagged proteins, operated through a peristaltic pump (Pumpdrive 5201, Heidolph) at controlled flow rates of 1 or 5 mL/min, respectively. (Figure 11)

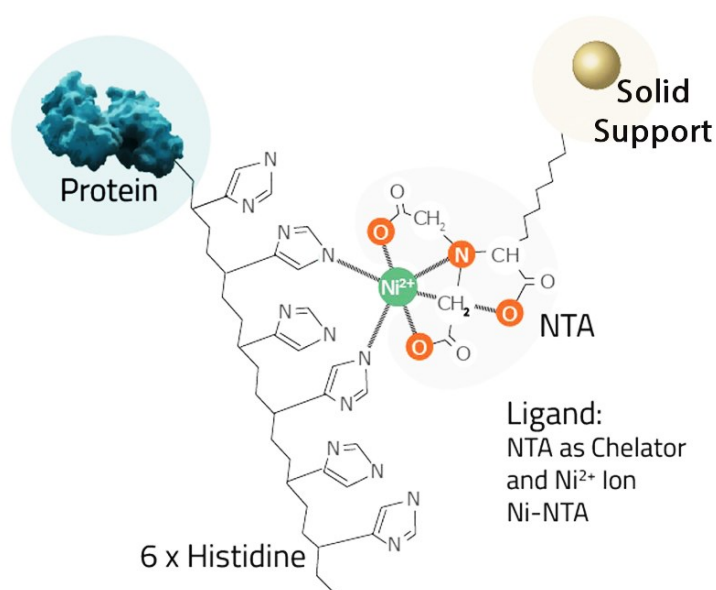


Figure 11. Schematic representation of Ni-NTA affinity chromatography for the purification of His-tagged proteins. The nickel ion is chelated by nitrilotriacetic acid (NTA) groups immobilised on the resin and coordinates the imidazole side chains of histidine residues within the polyhistidine tag. Bound proteins are retained on the column through metal-ligand coordination and are selectively eluted by competition with free imidazole. Figure adapted from [160].

Columns were first washed with 10 column volumes (CV) of deionized water, followed by 10 CV of equilibration buffer without imidazole. Clarified

lysates were then loaded onto the columns and recirculated at reduced flow rate for approximately one hour to ensure efficient binding of His-tagged proteins. After binding, the columns were connected to the NGC Chromatography System (Bio-Rad), which had been pre-equilibrated with two inlet lines, one containing equilibration buffer and the other an elution buffer supplemented with 500 mM imidazole.

The purification sequence proceeded as follows:

- Wash 1: 8 CV of equilibration buffer (no imidazole).
- Wash 2: 5 CV of buffer containing 20 mM imidazole.
- Wash 3: 5 CV of buffer containing 35 mM imidazole.
- Elution: buffer containing 250 mM imidazole until baseline absorbance returned to zero.
- Column cleaning: 5-10 CV of buffer containing 500 mM imidazole.

All fractions were collected, and aliquots were analysed by SDS-PAGE. The eluted protein fractions (with 250 mM imidazole) were pooled and diluted with equilibration buffer to a final volume of 50 mL to reduce the imidazole concentration prior to concentration.

Protein solutions were concentrated using Merck Millipore Amicon™ centrifugal filters (appropriate molecular weight cutoff) to approximately 10 mL and then refilled with imidazole-free buffer to a total of 50 mL. For His-tag removal, TEV protease (internally purified, 2 mg/mL stock) was added at a 1:50 protease-to-protein ratio (w/w), and the mixture was incubated overnight at 4 °C under gentle agitation (70 rpm) using a CAT M. Zipperer RM 5 roller mixer.

3.3.2.3 Reverse Ni-NTA purification

After TEV protease cleavage, the recombinant proteins lost the N-terminal His₆-tag and consequently their affinity for nickel ions immobilized on Ni-NTA matrices. To separate the cleaved, tag-free proteins from the remaining His-tagged species (uncleaved protein, TEV protease, and free His-tag peptide), a reverse Ni-NTA affinity chromatography step was performed.

Before loading, the digested protein samples were centrifuged at 15,000 g for 15 minutes at 4 °C to remove possible aggregates or precipitated material. The clarified supernatant was then applied to a Ni-NTA column previously washed with 10 column volumes (CV) of deionized water and equilibrated with an equal volume of binding buffer (without imidazole).

During this step, His-tagged species and residual TEV protease bound to the nickel resin, whereas the cleaved, untagged target proteins eluted directly in the flow-through fraction. The flow-through was collected, and aliquots from all fractions (flow-through, wash, and elution) were analysed by SDS-PAGE to check the efficiency of tag removal and recovery of the target protein. The collected flow-through containing the purified, tag-free protein was subsequently concentrated using Merck Millipore Amicon™ centrifugal filters (with the appropriate molecular weight cutoff) to approximately $\leq 1\%$ of the column bed volume used for subsequent size-exclusion chromatography (SEC). This concentration ensured optimal sample homogeneity and resolution during gel filtration analysis (see Section 3.3.4).

3.3.3 Cation exchange chromatography

Further purification of the target proteins was achieved by cation-exchange chromatography using a Macro-Prep® High S column (5 mL, Bio-Rad) connected to the BioLogic LP low-pressure chromatography system (Bio-Rad). This step allowed separation of protein isoforms and removal of residual contaminants based on differences in net surface charge under controlled pH conditions.

Before loading, the column was equilibrated with 10 CV of equilibration buffer (pH adjusted below the protein's isoelectric point to ensure positive net charge). The clarified and filtered protein solution was then loaded onto the column at a constant flow rate of 1 mL/min.

The bound proteins were eluted using 10 CV of NaCl stepwise gradient elution buffer at a constant flow rate of 3 mL/min, with final salt concentrations of 0, 50, 100, and 200 mM, followed by a final high-salt wash

with 2 M NaCl to remove tightly bound contaminants and regenerate the column.

Elution was monitored spectrophotometrically at 280 nm, and fractions corresponding to distinct elution steps were collected manually.

Aliquots from specific fraction were analysed by SDS-PAGE to identify those containing the target protein. Selected fractions were pooled and, when necessary, desalted and concentrated using Amicon™ centrifugal filters (Merck Millipore) prior to subsequent size-exclusion chromatography (see Section 3.3.4).

3.3.4 Size-exclusion chromatography (SEC)

Size-exclusion chromatography (SEC) was performed as the final purification and polishing step to achieve high sample homogeneity, remove aggregates, and exchange the protein into the desired experimental buffer. In SEC, proteins are separated according to their hydrodynamic radius, with larger species eluting earlier due to exclusion from the porous stationary phase, while smaller molecules penetrate the pores and therefore elute at higher volumes. (Figure 12)

Depending on the scale and purpose of the purification, different column formats were employed: HiLoad™ Superdex 75 pg (26/60) column (Cytiva) (bed volume $\approx$ 320 mL) and Superdex 75 16/60 column (Cytiva) (bed volume $\approx$ 120 mL) for preparative-scale purifications. When required, desalting was performed using a Sephadex™ G-25 desalting column equilibrated with 100 mM acetic acid (CH₃COOH) prepared from a commercial 16.7 M glacial acetic acid stock solution. This step ensured complete removal of high-salt components or buffer residues prior to subsequent analyses.

All SEC runs were performed on an NGC chromatography system (Bio-Rad) operated at 4 °C to ensure protein stability. The system was equipped with a UV-Vis detector (set at 280 nm), a conductivity sensor, and an autosampler for automated fraction collection. Columns were equilibrated with at least 1.5 column volumes (CV) of SEC buffer (see Section 2.3.1)

before sample injection to achieve baseline stabilization. All buffers were freshly prepared, filtered through 0.22 μm membranes, and degassed under vacuum to avoid bubble formation during chromatographic runs.

Protein samples were clarified by centrifugation at 15000 g for 10 minutes and filtered (0.45 μm) immediately before loading to prevent clogging or pressure buildup. Each sample was concentrated to $\leq 1\%$ of the total bed volume (typically 1-5 mL depending on column format) using Amicon™ centrifugal filters (Merck Millipore) with the appropriate molecular weight cutoff.

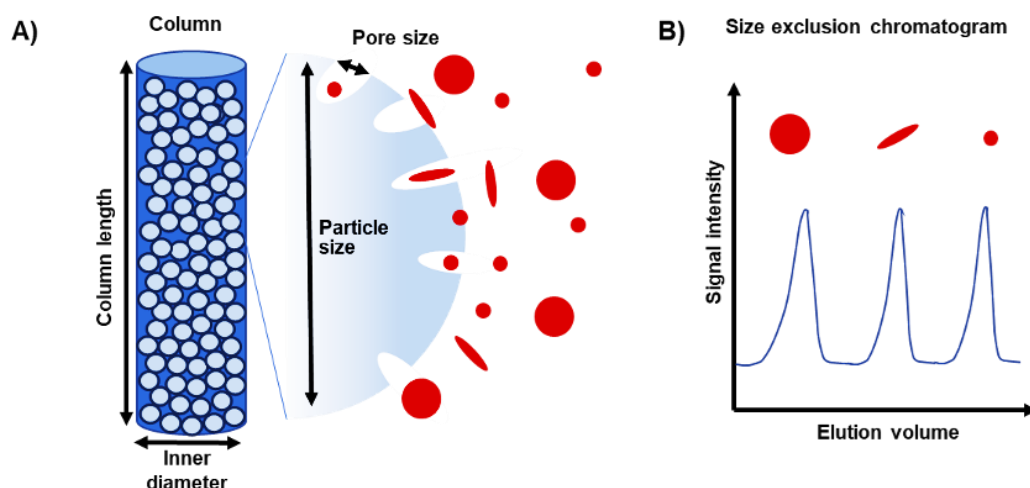


Figure 12. Schematic representation of the separation mechanism in size-exclusion chromatography. (A) Separation of molecules based on hydrodynamic size because of differential access to the porous stationary phase, with larger species eluting earlier than smaller ones. (B) Representative SEC elution profile, in which distinct peaks correspond to molecular species of different apparent sizes. Figure adapted from [161].

Samples were injected into the system using a 2 or 5 mL injection loop, and elution was performed isocratically at a constant flow rate of 0.5 mL/min (Superdex 16/60) or 1.0 mL/min (Superdex 26/60). The elution profile was monitored by UV absorbance at 280 nm, and chromatograms were recorded in real time using ChromLab software (Bio-Rad). Fractions were collected automatically every 2 mL using a programmable fraction collector. Peaks corresponding to the monomeric form of the target protein were identified based on elution volume and compared with the expected

calibration range of the column. Calibration was performed periodically using gel filtration molecular weight standards (Cytiva) to verify column performance and estimate apparent molecular mass.

Collected fractions were analysed by SDS-PAGE to verify the presence and purity of the target protein. Fractions corresponding to the main peak were pooled and concentrated to the desired volume using Amicon™ filters, then either used immediately for downstream biochemical or structural studies, or flash-frozen in liquid nitrogen and stored at -80 °C.

3.4 Protein analytics

3.4.1 SDS-PAGE

Following each purification step, aliquots of the collected fractions were analysed by SDS-PAGE to evaluate protein expression, solubility, purification efficiency and approximate molecular mass. This technique allows separation of denatured proteins according to their apparent molecular weight under the influence of an electric field, with resolution determined by the acrylamide concentration of the polyacrylamide matrix. The method was performed according to the standard protocol of Laemmli (1970), with slight modifications to gel composition depending on the protein size range to be resolved. [162]

Resolving (lower) gels of various acrylamide concentrations (12% and 15%) were prepared according to the formulations summarized in Table 12, with a total gel volume of 6 mL. Each gel was polymerized using 10% ammonium persulfate (APS) and TEMED as radical initiators. The stacking (upper) gel (6%) was prepared using Tris-HCl 0.5 M pH 6.8 as buffering system and polymerized with 10% APS and TEMED, with a total volume of 3 mL to ensure uniform sample loading. The specific compositions for each gel percentage are reported below:

	<i>Lower</i>		<i>Upper</i>
Gel percentage	15%	12%	6%
H₂O – Helix	2.13 mL	2.58 mL	1.74 mL
Acrylamide (40%)	2.25 mL	1.80 mL	450 µL
Tris-HCl 1.5 M pH 8.8	1.50 mL	1.50 mL	-
Tris-HCl 0.5 M pH 6.8	-	-	750 µL
SDS (10%)	60 µL	60 µL	30 µL
TEMED	6 µL	6 µL	3 µL
APS (10%)	60 µL	60 µL	60 µL

Table 12. Composition of resolving and stacking gel for SDS-PAGE.

Protein samples were mixed with Laemmli loading buffer (see Section 2.3.1), then boiled for 5 minutes at 95 °C to ensure complete denaturation. Typically, 10-20 µg of total protein was loaded per lane, alongside a prestained molecular weight marker (Thermo Scientific™ PageRuler™). Electrophoresis was carried out in Tris-glycine-SDS running and the gels were run at a constant current of 65 mA and a voltage of 300 V for approximately 30 minutes, until the dye front reached the bottom of the gel. After electrophoresis, gels were stained with Coomassie Brilliant Blue R-250 (see Section 2.3.1) for 30-45 minutes. To accelerate the staining process, the gels were briefly heated in a microwave oven for approximately 30 seconds and subsequently incubated on an orbital shaker to ensure uniform dye distribution. Decolouration was carried out in the same solvent mixture, with a brief microwave heating (approx. 45 seconds) followed by gentle agitation on a shaker. The destaining solution (see Section 2.3.1) was replaced several times, and, when applicable, recycled through an activated charcoal filtration system to reduce solvent waste. Gels were considered fully destained when protein bands were clearly visible against a transparent

background. Stained gels were documented using a Gel Doc™ (Bio-Rad) imaging.

3.4.2 Western blot

Protein transfer and immunodetection were performed using standard Western blotting procedures with adaptations optimized in the laboratory. The technique was used to check protein expression, assess the presence of the His-tag, and confirm target specificity through antibody recognition. Following SDS-PAGE, proteins were electrotransferred onto nitrocellulose membranes using the Trans-Blot® Turbo™ Transfer System (Bio-Rad). The transfer “sandwich” was assembled by sequentially placing six pre-soaked blotting paper sheets at the bottom of a transfer cassette, followed by the nitrocellulose membrane, the SDS-PAGE gel, and another stack of six blotting paper sheets, all equilibrated in transfer buffer (see Section 2.3.1). Each layer was carefully rolled with a roller to remove air bubbles that could interfere with the transfer efficiency. The cassette was then sealed and inserted into the Trans-Blot® Turbo™ apparatus. Protein transfer was performed using two sequential transfer cycles, each at 20 V for 30 minutes at room temperature. After transfer, the membrane was sometimes incubated in 0.1% Ponceau S staining solution for 1 hour to verify the efficiency and uniformity of protein transfer. The stained membrane was thoroughly rinsed with deionized water, visualized, and documented using a Gel Doc™ imaging system (Bio-Rad) equipped with a red filter suitable for Ponceau detection. Depending on the primary antibody used, two distinct immunodetection protocols were applied. All incubation steps were performed in Falcon tubes on a CAT M. Zipperer RM 5 roller mixer set to 70 rpm to ensure uniform mixing at room temperature.

3.4.2.1 His-tag detection (DAB chromogenic development)

Membrane was first washed twice with 20 mL of TBS buffer (see Section 2.3.1) for 15 minutes each, followed by blocking in 15 mL of blocking buffer (see Section 2.3.1) for 1 hour.

After blocking, the membrane was washed twice with 20 mL of TBST (see Section 2.3.1) for 15 minutes each and then incubated for 1 hour with the anti-His primary antibody (10 mL, recovered after use).

Subsequently, the membrane was washed once in 20 mL of TBS for 15 minutes, followed by an additional wash in TBST under gentle agitation on an orbital shaker. For signal development, membrane was incubated with 10 mg of DAB (3,3'-diaminobenzidine) dissolved in 20 mL of TBST, to which a small volume of 30% H₂O₂ was added immediately before incubation to start the chromogenic reaction.

The reaction was monitored visually until brown bands corresponding to the His-tagged protein became visible. The membrane was then rinsed completely with distilled water to stop the reaction and imaged using the Gel Doc™ system.

The total time of the procedure was approximately 5 hours.

3.4.2.2 Immunodetection with primary and secondary antibodies (ECL chemiluminescence)

In cases where two antibodies were used, membranes were first washed in distilled water after Ponceau staining and imaged to document total protein loading.

The membranes were then washed three times for 10 minutes each in 20 mL of TBST, followed by blocking for 30 minutes in 15 mL of TBST supplemented with 5% non-fat dry milk.

After blocking, membranes were washed again three times for 10 minutes each in TBST and then incubated for 1 hour with the primary antibody (Antibody A, 20 mL, see Section 2.3.5). The antibody solution was recovered for reuse.

After incubation, membranes were washed three times with 10 mL TBST for 10 minutes each, followed by incubation for 1 hour with the secondary antibody (Antibody B, 20 mL, see Section 2.3.5) under the same conditions. After three final washes (10 mL TBST, 10 minutes each), the ECL (Enhanced Chemiluminescence, see Section 2.3.1) detection reagent was

prepared and applied to the membrane in near-dark conditions. The membrane was gently rolled to ensure even reagent distribution and incubated for approximately 1 minute.

Chemiluminescent signals were detected using a ChemoStar imaging system (Intas), and the resulting bands were analysed for specificity and relative signal intensity. The total time of the procedure was approximately 2 days.

3.4.3 Native PAGE

Native-PAGE was performed to analyse the monomeric and dimeric state of purified proteins under non-denaturing conditions. This method allows the separation of proteins according to their native conformation and net charge at the selected pH, thus preserving their biological activity. For the analysis of basic proteins (Perm), electrophoresis was conducted under acidic conditions using the fast native electrophoresis protocol optimized in the lab. The procedure was based on a β -alanine/AcOH buffer system at pH 4.3, which ensures the migration of basic proteins toward the cathode after reversal of the electrode polarity. Resolving (lower) gels were prepared in volumes of 5 mL at different acrylamide concentrations (10% and 12%) depending on the molecular size of the target protein, as detailed in Table 12. The gels were composed of H₂O-Helix, 40% acrylamide, 1.5 M potassium acetate (KOAc) buffer pH 4.3, 50% glycerol, 10% ammonium persulfate (APS), and TEMED as a polymerization catalyst. The stacking (upper) gel (4%) was prepared with KOAc 0.5 M, pH 5.2 to ensure optimal focusing of protein bands. Polymerization was allowed to proceed for 2 hours at room temperature, and gels were stored overnight at 4 °C in a humid chamber to maintain hydration until use.

	Lower		Upper
Gel percentage	10%	12%	4%
H ₂ O – Helix	765 µL	435 µL	3.12 mL
Acrylamide (40%)	1.67 mL	2.00 mL	2.00 mL
KOAc 1.5 M pH 4.3	1.28 mL	1.28 mL	-
KOAc 0.5 M pH 5.2	-	-	870 µL
Glycerol (50%)	1.15 mL	1.15 mL	-
TEMED	15 µL	15 µL	10 µL
APS (10%)	120 µL	120 µL	60 µL

Table 12. Composition of resolving and stacking gel for Native PAGE.

Before loading the samples, the instrument electrodes were inverted (positive and negative poles exchanged) to accommodate the reversed migration direction typical of acidic native PAGE systems. A pre-run of 30 minutes at a constant current of 35 mA was performed to stabilize the electric field and equilibrate the gel. Then, samples were loaded into the wells, and electrophoresis was carried out for 150 minutes at a constant current of 35 mA. The entire procedure was conducted at 4 °C to preserve protein conformation and minimize diffusion. Upon end, the gels were removed and directly stained with Coomassie Brilliant Blue R-250 for 1 hour, followed by destaining in the same solution until well-defined bands appeared. Stained gels were imaged using a Gel Doc™ (Bio-Rad) system, and the migration pattern was compared to known standards or to control samples analysed in parallel by SDS-PAGE.

3.4.4 UV-Vis spectrophotometry for protein concentration determination

Protein concentrations were determined using a NanoDrop™ OneC spectrophotometer (Thermo Scientific), which allows microvolume

absorbance measurements without the need for cuvettes. For each sample, 1-2 μL of purified protein solution were deposited on the pedestal, and absorbance spectra were recorded between 200 and 350 nm. The protein concentration was calculated automatically by the instrument software based on the absorbance at 280 nm (A_{280}) and the theoretical molar extinction coefficient (ϵ) and molecular weight (MW) of each protein, previously computed using the ProtParam tool (ExPASy). The Beer-Lambert law was applied as:

$$c = \frac{A_{280}}{\epsilon \times l}$$

where c is the molar concentration (mol/L), A_{280} is the measured absorbance, ϵ is the molar extinction coefficient ((1/M)/cm), and l is the optical path length (1 mm for NanoDrop measurements).

Samples with absorbance values exceeding 1.5 were diluted appropriately in the corresponding buffer to ensure measurements within the linear range of the detector. All measurements were performed at room temperature (22-25 °C), and the buffer spectrum was used as baseline reference.

3.4.5 Circular Dichroism (CD) spectroscopy

Circular dichroism (CD) spectroscopy was employed to evaluate the secondary structure content and thermal stability of the purified proteins under native conditions. Measurements were performed using a Jasco J-1500 spectropolarimeter (JASCO Corporation, Milan, Italy) equipped with a Peltier temperature control system. Quartz cuvettes with a 0.1 cm path length were used, and all measurements were carried out at 25 °C.

Before the analysis, lyophilized protein samples were dissolved in 5 mM HCl, using minimal quantities of powder collected with a micro spatula and suspended in 300 μL of solvent. Protein concentration was first determined by UV-Vis spectrophotometry (see Section 3.4.4). The concentration was adjusted to yield an absorbance between 0.4 and 0.8 at 280 nm, in

accordance with the Lambert-Beer law, ensuring a final working concentration of approximately 0.1 mg/mL.

CD analyses were performed on samples with a final volume of 250 μ L without or with salt (150 mM NaCl), with a buffer (20 mM NaPO₄) adjusted to the desired pH.

Before acquisition, samples were centrifuged at 12000 g for 5 minutes to remove potential aggregates.

Far-UV CD spectra were collected in the 190-260 nm range, averaging three consecutive scans per sample. Instrument settings were scanning speed 50 nm/min, bandwidth 2.0 nm, data pitch 1.0 nm, response time 2 s, and sensitivity 50 mdeg. Spectra were baseline-corrected using the corresponding buffer and smoothed using the manufacturer's software. The ellipticity values were converted to mean residue ellipticity $[\theta]$ (deg·cm²/dmol) according to the equation:

$$[\theta] = \frac{\theta_{obs} \times 100 \times M}{c \times l}$$

where θ_{obs} is the observed ellipticity (mdeg), M is the mean residue molecular weight, c is the protein concentration (mg/mL), and l the optical path length (cm).

The number of amino acid residues required for the calculation of mean residue ellipticity was obtained from UniProtKB/Swiss-Prot (ExPASy Bioinformatics Portal). Normalized far-UV CD spectra were subsequently analysed using the BeStSel web server to estimate secondary structure content. [163]

Thermal unfolding experiments were performed by monitoring the CD signal at 215 nm as a function of temperature, over the 20-100 °C range, with a 1.0 °C/min heating rate. Each denaturation curve represented the average of three independent scans, baseline-corrected and fitted to a sigmoidal model to determine the apparent melting temperature (T_m), corresponding to the midpoint of the unfolding transition.

3.4.6 Mass Spectrometry analysis

Mass spectrometry (MS) analysis was employed for the qualitative identification of the purified recombinant protein and to verify its correspondence with the expected amino acid sequence.

A distinct protein band corresponding to the target molecular weight was excised from a Coomassie Brilliant Blue-stained SDS-PAGE gel using a sterile scalpel under clean conditions. The gel slice was transferred into a 1.5 mL microcentrifuge tube (Eppendorf), sealed, and labelled with the sample code. Subsequently, a *Bioanalysis Request Form* was completed and submitted together with the sample to the Core Facility for Mass Spectrometry and Elemental Analysis at the Philipps-Universität Marburg, Department of Chemistry (Synmikro, Marburg, Germany). The requested service was qualitative protein identification via peptide mass fingerprinting and tandem mass spectrometry (MS/MS).

According to the facility's protocol, proteins embedded in the gel matrix were enzymatically digested in situ (typically with trypsin) following DTT reduction and alkylation, and the resulting peptides were analysed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Peptide fragmentation spectra were processed through a database search (e.g., UniProtKB) to identify protein sequences matching the theoretical digest of the submitted construct. The resulting report included a list of identified peptides, their mass-to-charge ratios (m/z), and the corresponding sequence coverage percentage.

3.4.7 Thioflavin T (ThT) fluorescence assay

Fluorescence measurements were performed using a Shimadzu RF-5301PC spectrofluorometer equipped with a thermostatted cell holder connected to a circulating water bath to maintain a constant temperature of 20 °C. Samples were prepared in 20 mM sodium phosphate buffer, pH 6.8, containing 0.1 mg/mL of protein. Where required, 150 mM NaCl was added to the samples. Thioflavin T was added from a concentrated stock solution

to achieve a protein-to-dye molar ratio of 1:3. Samples were gently mixed and incubated under controlled conditions prior to fluorescence acquisition. Fluorescence emission spectra were recorded at different incubation times using a 1 cm path-length quartz cuvette. The excitation wavelength (λ_{ex}) was set to 440 nm, and emission spectra were collected in the 460-550 nm range, with $\lambda_{\text{em}} = 485$ nm as the characteristic emission maximum. Both excitation and emission slit widths were set to 5 nm, and the integration time was 1 s per data point. Fluorescence intensity values were monitored over time to detect any increase indicative of ThT binding to β -structured aggregates. [164,165] (Figure 13) Control samples containing only ThT in buffer were used to subtract background fluorescence. Fluorescence spectra were corrected for buffer background and normalized to the maximum intensity of ThT-free samples. Each measurement was repeated in triplicate, and averaged spectra were reported.

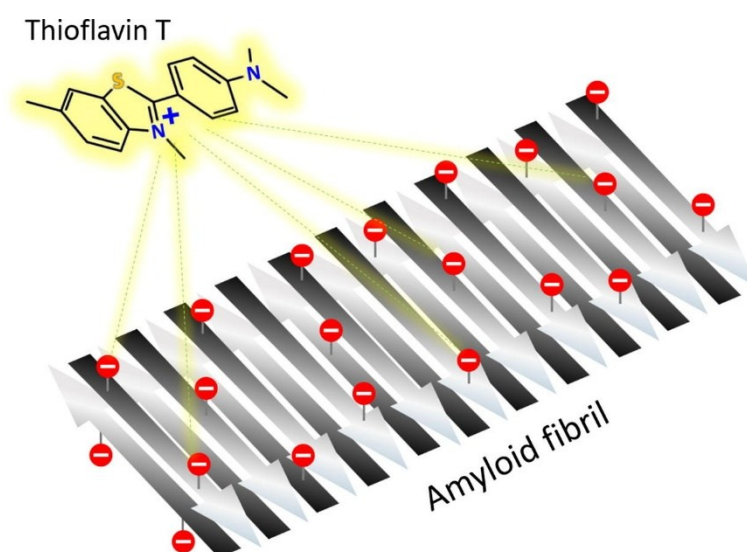


Figure 13. Schematic representation of the interaction between Thioflavin T (ThT) and amyloid fibrils. Upon binding to the repetitive cross- β architecture of amyloid aggregates, ThT molecules intercalate along the β -sheet grooves, leading to restriction of intramolecular rotations and a marked increase in fluorescence emission. This fluorescence enhancement provides a sensitive readout for the presence and accumulation of β -structured aggregates. Figure adapted from [165].

3.4.8 Transmission Electron Microscopy (TEM)

TEM was employed to visualize the morphology of protein aggregates and supramolecular assemblies formed by proteins.

3.4.8.1 Grid preparation

Samples were prepared for TEM imaging following a negative-staining protocol optimized for protein samples. For each variant, a 3 μL aliquot of the appropriately diluted protein solution was deposited onto a carbon-coated copper grid (400 mesh, Formvar film). (Figure 14) The sample was allowed to adsorb for 2 minutes at room temperature to promote attachment of the protein particles to the carbon surface. Excess liquid was carefully removed using filter paper to avoid sample deformation. Immediately afterward, 3 μL of 2% (w/v) uranyl acetate were applied to the grid and allowed to incubate for 4 minutes to achieve negative staining. The excess stain was blotted off gently, and the grids were left to air-dry for several hours under dust-free conditions before imaging. All preparation steps were performed at room temperature using clean tweezers and filter paper to minimize contamination or mechanical damage to the film surface.

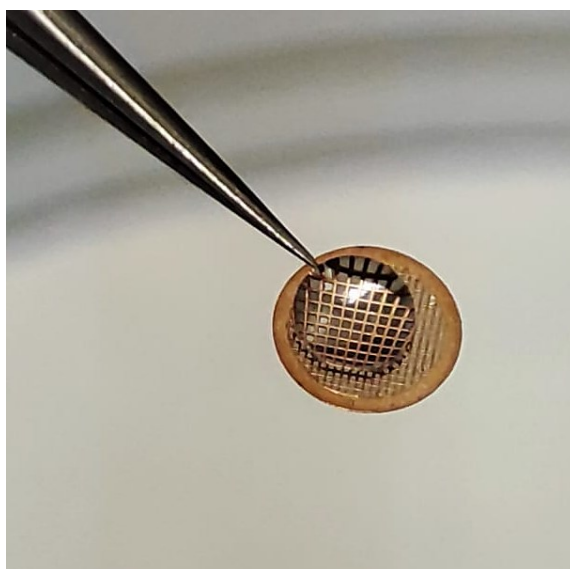


Figure 14. Close-up view of a carbon-coated copper TEM grid after sample deposition. The image highlights the millimetric size of the grid relative to the tweezers and the

micrometric mesh structure supporting the thin carbon/Formvar film. The presence of a liquid layer on the grid surface enhances the visibility of the underlying mesh pattern.

3.4.8.2 Grid Analysis

Electron micrographs were recorded using a FEI Tecnai G2 transmission electron microscope operated at an acceleration voltage of 120 kV. Images were acquired using a digital CCD camera under low-dose conditions to minimize radiation damage. Multiple regions of each grid were examined to ensure sample homogeneity and reproducibility. The obtained micrographs were analysed to assess aggregate morphology and structural organization.

3.4.9 Lyophilization and Long-Term Protein Preservation

Lyophilization was employed as the final step for long-term preservation of purified proteins, ensuring sample stability during extended storage and facilitating subsequent reconstitution for biochemical and structural analyses. After purification, protein samples were aliquoted into 50 mL Falcon tubes, each filled approximately to half of their total volume to allow adequate surface exposure during sublimation. Each tube was sealed by applying a layer of aluminium foil, covered with a layer of Parafilm in place of the original cap. The Parafilm surface was then perforated using a sterile needle to produce multiple small holes, allowing efficient moisture removal and preventing pressure buildup during the lyophilization process. Samples were first frozen at -20 °C overnight to ensure gradual solidification, followed by rapid freezing at -80 °C for at least 6 hours before lyophilization. This two-step freezing process was used to promote uniform ice crystal formation, improving drying efficiency and preserving protein structure. Lyophilization was carried out using a Lio 5P freeze-dryer (5Pascal, Italy). The pre-frozen Falcon tubes were placed in the vacuum chamber, and the lyophilization cycle was maintained until complete sublimation of the frozen solvent was achieved, as visually confirmed by the appearance of dry, powder-like protein residues on the tube walls. Upon finishing point, the tubes were immediately sealed with their original caps under dry conditions

to prevent rehydration and stored at -20 °C for long-term preservation. Lyophilized protein samples were reconstituted in the appropriate buffer before use in following biochemical or spectroscopic assays.

3.5 Protein crystallography

Crystallization trials were carried out at the MarxXtal Crystallization Facility (Philipps-Universität Marburg, Germany), under the supervision of the Structural Biochemistry Laboratory of Prof. Lars-Oliver Essen. Before the crystallization, the purified protein was sterile-filtered and maintained at 4 °C until use, to preserve structural integrity and minimize aggregation. [166,167]

Successful crystallization critically depends on the use of highly pure and monodisperse protein samples. For this reason, proteins used in this study were obtained from size-exclusion chromatography fractions corresponding to the predominant solution state, thereby reducing the presence of oligomeric heterogeneity. Protein concentration was adjusted empirically in a range of 2-50 mg/mL suitable for crystallization screening, avoiding visible precipitation or turbidity.

Crystallization trials were carried out using vapor diffusion techniques, employing both sitting-drop and hanging-drop configurations depending on the experimental stage. In both approaches, protein and precipitant solutions were mixed and equilibrated against a larger reservoir containing the precipitant solution. This setup promotes gradual vapor diffusion, leading to an increase in protein and precipitant concentration within the drop and enabling the transition from undersaturated to supersaturated conditions required for nucleation and crystal growth. [166,167]

3.5.1 Crystallization and diffraction data processing

Initial crystallization screening was performed by the vapor diffusion method using a Honeybee 963™ liquid-handling robot (Digilab) and MRC 2-well crystallization plates (Swissci). To maximize the probability of identifying suitable crystallization conditions, eight commercial screening kits were

tested: JB Basic HT, JB PEG/Salt, Classics, JCSG+, Cryos, PACT, MBC I, and Morpheus I-III, covering a wide range of pH values, precipitants, and additive conditions. Each screen consisted of 96 empirically optimized formulations. Protein and reservoir solutions were mixed in equal volumes and equilibrated against 50 μ L of precipitant solution per well. The plates were incubated at 8 °C and 20 °C in a Rock Imager™ automated imaging system (Formulatrix), which continuously monitored crystal growth under controlled humidity and temperature. Conditions yielding crystalline material, including microcrystals or phase separation, were prioritized for further inspection and optimization. For selected conditions, crystals were harvested directly from the drops using mounted CryoLoops (Hampton Research). When required, ligand incorporation was performed by soaking, by transferring pre-formed crystals into crystallization solutions supplemented with glucose. This allowed diffusion of the ligand into the crystal lattice while minimizing perturbation of crystal order. Before data collection at the European Synchrotron Radiation Facility (ESRF, Grenoble, France), crystals were cryoprotected either by direct transfer into a cryoprotectant solution or by exploiting the presence of cryoprotective agents already contained in the mother liquor. Crystals were then mounted in nylon loops and flash-frozen in liquid nitrogen.

Diffraction data were processed using the XDS software package [168], starting from the automatically generated input file (XDS.INP) provided by the ESRF beamline. The workflow included indexing of diffraction images, integration of reflection intensities, and scaling of the dataset. Data quality was assessed based on standard crystallographic indicators, including resolution limits, completeness, redundancy, and merging statistics.

The processed data were imported into CCP4i2 for graphical analysis and data quality assessment. [169] The phase problem was solved by molecular replacement (MR) using a homologous structure as a search model. This approach enabled generation of initial phase estimates necessary for model building and refinement.

3.5.2 Structure Refinement

Model refinement was carried out using Phenix.refine within the PHENIX package. [145] Iterative cycles of coordinate, occupancy, and B-factor refinement were performed, alternating with manual model inspection in Coot. [141] Phenix.refine automatically adjusts model parameters to optimize stereochemical geometry and minimize the difference between observed and calculated structure factor amplitudes.

Final coordinates were exported in PDB format and visualized in PyMOL (Schrödinger LLC) for figure preparation. All structural representations reported in this thesis were generated using PyMOL.

3.7 Bioinformatics

Computational analyses were conducted to complement the experimental results and to gain insight into the structural and dynamic properties of proteins. Structure prediction, protonation-state estimation, and molecular dynamics (MD) simulations were performed using a combination of state-of-the-art algorithms and high-performance computing resources, as described below.

3.7.1 Protein structure prediction tools

The 3D structures of the studied proteins were predicted using AlphaFold2, AlphaFold3, and the Chai-Discovery modelling framework. AlphaFold2 and AlphaFold3 predictions were performed through the ColabFold implementation [24,34,138], which integrates the AlphaFold neural network model with an accelerated MMseqs2-based multiple sequence alignment (MSA) generation pipeline. The structural confidence of each predicted model was evaluated based on the predicted Local Distance Difference Test (pLDDT) and Predicted Aligned Error (PAE) metrics provided by the algorithm. In parallel, Chai Discovery was employed to validate and compare the predicted conformations, focusing on surface charge distribution, solvent accessibility, and the arrangement of secondary-structure motifs. [35] The most reliable conformers, based on confidence

metrics and structural stability, were selected for subsequent molecular dynamics (MD) simulations.

3.7.2 Molecular dynamics simulations

The 3D protein structures obtained from the prediction tools described in Section 2.8.2 and 3.7.1 were used as starting models for all-atom molecular dynamics (MD) simulations. All simulations were performed on monomeric protein structures using the AMBER 24 software suite on remote high-performance computing clusters (see Section 2.8.1).

Before assembling the system, protein structures were pre-processed using the H++ web server to estimate residue pKa values and assign protonation states consistent with the target pH conditions. [143] After automated residue renumbering, minor manual adjustments were applied to ensure compatibility with AMBER force-field conventions. System preparation was carried out using AMBER's tleap, assembling the protein, cofactors, and ligands. The ff14SB or ff19SB force fields were used for protein residues, GAFF for small-molecule ligands, and custom parameter files were loaded for non-standard residues when required. [44,45,170] Each system was solvated in an explicit water box using either the TIP3P or OPC water model [171,172], extending at least 10 Å from the solute in all directions, and electrically neutralised by addition of Na⁺ and Cl⁻ ions. Energy minimisation was performed in multiple stages to progressively relax the system. An initial minimisation of 5000 steps (500 steps of steepest descent followed by conjugate gradient) was conducted with positional restraints applied to the protein and cofactors. This was followed by successive minimisation steps with harmonic restraints applied to α -carbon atoms, gradually reduced from 50 to 5 kcal/mol·Å, and concluded with a final unrestrained minimisation of 5000 steps.

Equilibration was performed using a 0.5 fs timestep and consisted of four phases: (i) 100 ps of NVT heating from 0 K to the target temperature (300-450 K); (ii) 350 ps of NVT equilibration with positional restraints on α -carbons to stabilise the protein structure; (iii) 700 ps of NPT equilibration

using a Monte Carlo barostat to reach a target pressure of 1 atm; and (iv) an extended 18 ns NPT equilibration using a 2 fs timestep to prepare the system for production runs.

Production MD simulations were carried out under NPT conditions for 200 ns per replicate with a 2 fs timestep. Temperature was maintained at the target temperature (usually 300 or 350, 400, 450, 500 K) using a Berendsen thermostat (ntt = 1) with a relaxation time of 2 ps, while pressure was controlled at 1 atm using a Monte Carlo barostat. [173,174] Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method with a 10 Å cutoff for non-bonded interactions. [175] Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm, enabling the use of a 2 fs timestep. [176] The iwrap = 1 option was enabled to maintain molecules within the periodic box during trajectory analysis. Coordinates were saved every 50 ps and restart files every 100 ps. For each condition, three independent simulations of 200 ns or more were performed to enhance statistical sampling and reduce uncertainty in interaction frequency estimates. Trajectory post-processing and analysis were conducted in Jupyter Notebook environments using CPPTRAJ and PYTRAJ [177], with NGLView [178] employed for visual inspection and VMD for PCA analysis. [153] Structural stability was assessed through backbone RMSD and RMSF analyses, hydrogen bond persistence, and secondary-structure evolution, the latter quantified using the DSSP algorithm implemented in AMBER.

4 Results

4.1 Development of a MNEI Mutant

4.1.1 Design of the Mut4 (E23Q) mutant

The design of the Mut4 mutant is based on a rational strategy developed by Leone and Picone to stabilize MNEI, which represents the conceptual starting point of this study. In this approach, attention is focused on identification of ionizable residues buried in the hydrophobic core of the protein that are considered penalizing, since their presence can contribute to a marked pH-dependent instability and reduce the conformational robustness of the native state.

The Glu23 residue emerged as particularly relevant, as it is located near the α -helix of MNEI and is inserted into a hydrophobic cavity, where the negative charge is not involved in adjacent electrostatic or ionic contacts. Previous studies on MNEI conducted by Liu *et al.* have shown that replacing the Glu23 residue with alanine (E23A) leads to a marked increase in the thermal stability of the protein, which can resist 30 minutes at 85 °C, making it one of the most efficient constructs in terms of improving heat resistance. [179] However, through molecular dynamics simulations conducted by MNEI, Leone and Picone demonstrated that Glu23, despite being buried in the structure, contributes to making the protein more unstable. Based on this evidence, they proposed the E23Q mutation, replacing glutamic acid with glutamine, thus removing the negative charge and maintaining lateral dimensions and hydrogen bonding capabilities comparable to those of the original residue. [77] The aim of the authors was to improve pH-dependent stability while preserving as much as possible the three-dimensional architecture and functional properties of the protein. The Mut4 mutant was then expressed, purified, and experimentally characterized by the same authors, who demonstrated its improved stability without significant alterations to the functionality linked to the recognition of the sweet taste receptor. Mut4 therefore represents a system that has already been extensively validated from a structural and functional point of view.

In the context of this work, Mut4 was not subjected to a new design, but was adopted as a reference system to investigate aspects that remained relatively unexplored. In particular, attention was focused on its behaviour in solution and its propensity to aggregate under stress conditions. For this reason, the use of a mutant designed using a rational approach and already characterized in the literature allows for a more targeted analysis of the relationship between conformational stability and aggregation phenomena.

4.1.2 Protein production and purification

The expression and purification of Mut4 (E23Q) were carried out using a procedure partly derived from the protocols reported in the literature on MNEI and its mutants, but with targeted modifications in the expression phase to promote correct folding. In particular, after inoculation of the recombinant strain and preliminary growth at 37 °C, protein synthesis was induced by adding lactose (5 mM). Immediately after induction, the temperature was reduced to 25 °C by incubating overnight. This change was introduced with the aim of slowing down the rate of protein synthesis and reducing the formation of insoluble inclusions, potentially relevant for Mut4, given the nature of the mutation and its local impact on folding. To monitor induction efficiency, aliquots of the culture were taken immediately before the addition of the inducer ("NI" samples, non-induced) and at the end of overnight incubation ("I" samples, induced). These samples were analysed by SDS-PAGE, together with wild-type MNEI expressed under the same conditions and used as a positive reference. As shown in the Figure 15, all three mutants exhibit a band of comparable intensity to that of MNEI, located in the region between 10 and 15 kDa, in accordance with the theoretical molecular mass of approximately 11 kDa. The absence of additional bands of different weights in the induced samples also indicates that expression occurs specifically and without evidence of significant degradation.

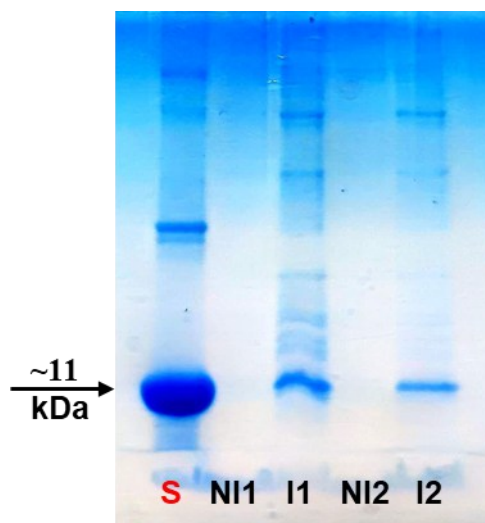


Figure 15. SDS-PAGE analysis of Mut4 (MNEI-E23Q) expression after lactose induction in two independent batches. A purified aliquot of MNEI with a weight of ~11 kDa was used as a reference instead of a commercial marker. The lanes show the soluble fraction (S), the non-induced samples before the addition of lactose (NI1, NI2), and the induced samples after overnight expression at 25 °C (I1, I2). The protein is present exclusively in the induced samples, indicating efficient and specific expression without obvious degradation phenomena.

After cell lysis by ultrasonic sonication (see Section 3.2.2.2) and clarification of the lysate by centrifugation (see Section 3.3.1), Mut4 was purified by cation exchange chromatography under the same operating conditions optimised for MNEI. Although the chromatographic profile obtained is comparable to that of the native protein, Mut4 showed a slightly different elution pattern: while MNEI typically elutes at a gradient of 100 mM NaCl in Perm elution buffer (see Section 2.3.1), Mut4 has a main peak at the same concentration but with a secondary component between 100 mM and 150 mM, as shown by the elution profile and confirmed by SDS-PAGE analysis of the corresponding fractions. (Figure 16)

This behaviour could reflect a moderate alteration in surface charge distribution due to the replacement of Glu23 with glutamine, although further structural analysis would be necessary for a definitive interpretation. Following purification, the sample was desalted to remove the high-salinity buffer in preparation for subsequent functional experiments. In this case, analytical size exclusion chromatography (SEC) was not performed to

evaluate the oligomeric state, as this was not one of the experimental objectives; desalination does not replace this characterisation but only allows the buffer to be changed. The final yield of the process was higher than that obtained for wild-type MNEI, with an average production of approximately 17 mg/L of culture.

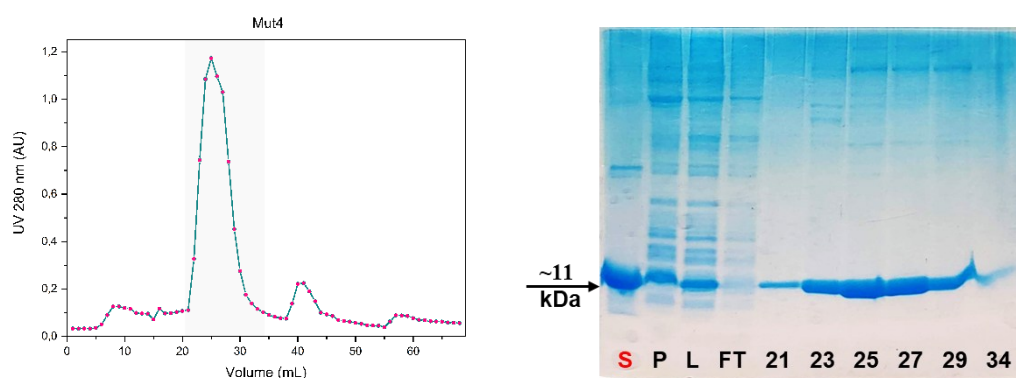


Figure 16. Cation exchange chromatography of Mut4 (MNEI-E23Q) following cell lysis and clarification, showing the UV absorbance profile at 280 nm (left) and SDS-PAGE analysis of selected fractions (right). The main elution peak is centred around the same salt concentration as wild-type MNEI, with an additional minor component eluting at slightly higher ionic strength. The shaded region in the chromatogram indicates the elution window corresponding to the fractions analysed by SDS-PAGE. The gel confirms the presence of Mut4 as a predominant band at ~11 kDa in the elution fractions, consistent with the expected molecular weight and indicating effective purification.

4.1.3 CD spectroscopy and evaluation of the secondary stability of Mut4

The secondary structure of Mut4 was first analysed using far-UV CD spectroscopy at pH 2.5, 5.1 and 6.8, in the absence and presence of 150 mM NaCl. (Figure 17) The spectra, in molar ellipticity, show that under all conditions the protein assumes a predominantly β -sheet conformation, which is characterised by a distinct minimum in the 215-216 nm region. The variations between the different pH values are mainly limited to the amplitude of the minimum. In fact, at pH 5.1, the spectrum has a slightly reduced intensity compared to the conditions at pH 2.5 and 6.8, while the position of the minimum (~215 nm) and the overall shape of the curves

remain substantially superimposable. The addition of NaCl does not induce qualitative changes in the spectral profile, indicating that neither the pH in the range considered nor the moderate ionic strength cause appreciable reorganisations of the secondary structure at room temperature, *i.e.* at 25 °C before the thermal denaturation profile. This observation is confirmed by the deconvolution analysis of the spectra (Table 13) conducted in the 195-250 nm range, which returns identical values for α -helix, β -sheet, turn and random coil content for all pH and ionic strength conditions. Although the absolute numerical identity between the conditions should be interpreted in light of the limited resolution of the deconvolution methods, overall, the data support the idea that Mut4 maintains the same rich β -folded structure across the entire pH range tested, both in the absence and presence of NaCl.

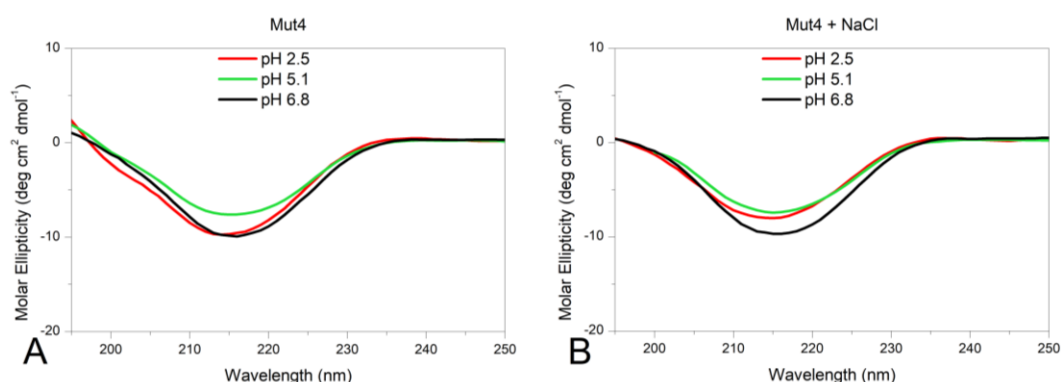


Figure 17. Far-UV circular dichroism spectra of Mut4 recorded at pH 2.5, 5.1 and 6.8 in the absence (A) and presence (B) of 150 mM NaCl at 25 °C. All spectra display a pronounced minimum at ~215-216 nm, indicative of a predominantly β -sheet secondary structure. Variations among pH conditions are limited to changes in ellipticity amplitude, with no appreciable shifts in spectral shape or minimum position. The addition of NaCl does not induce qualitative changes in the CD profiles, indicating preservation of the secondary structure across the tested pH and ionic strength conditions.

The structural stability of Mut4 was evaluated by monitoring the variation in the CD signal at 215 nm as a function of temperature, in the range 20-90 °C, at the three pH values considered, and converting the data into denatured fraction (*fu*). (Figure 18)

In the absence of NaCl, the thermal curves show a cooperative transition between ~60 and ~90 °C for all pH values; the melting temperature, estimated as the point of maximum slope of the curve, is 71 ± 0.5 °C at pH 2.5, 79 ± 0.5 °C at pH 5.1 and 82 ± 0.5 °C at pH 6.8. (Table 13) Under saline conditions, the Mut4 thermal curve maintains a cooperative transition similar to that observed in the absence of NaCl but overall shifted towards higher temperatures. (Figure 18-B)

Mut4				Mut4 + NaCl		
	pH 2.5	pH 5.1	pH.6.8	pH 2.5	pH 5.1	pH.6.8
α-helix (%)	2.2	2.2	2.2	2.2	2.2	2.2
β-sheet (%)	38.5	38.5	38.5	38.5	38.5	38.5
Turn	15.0	15	15.0	15.0	15.0	15.0
Random coil (%)	44.3	44.3	44.3	44.3	44.3	44.3
T_m (°C)	71 ± 0.5	79 ± 0.5	82 ± 0.5	71 ± 0.5	75 ± 0.5	87 ± 0.5

Table 13. Secondary structure content of Mut4 estimated by deconvolution of far-UV CD spectra in the 195-250 nm range using BeStSel. [163] The analysis returns identical percentages of α -helix, β -sheet, turns and random coil for all pH conditions, both in the absence and presence of NaCl. Thermal midpoint values (T_m) are reported for each condition, showing pH- and salt-dependent differences in thermal stability while preserving the same overall secondary structure composition.

At pH 2.5, the T_m remains substantially unchanged compared to the salt-free sample, while at pH 5.1 and, more markedly, at pH 6.8, an increase in melting temperature is observed. In particular, at pH 6.8, the apparent T_m of Mut4 in the presence of NaCl reaches ~87 °C, which is higher than that measured in the absence of salt. Again, the curves do not show a perfectly defined post-transition plateau within the instrumental limit, especially at pH 6.8, suggesting that the T_m values should be considered relative estimates

rather than absolute thermodynamic parameters. This trend indicates that the fully denatured population is not fully defined in the range explored and/or that secondary processes, such as aggregation, occur at high temperatures.

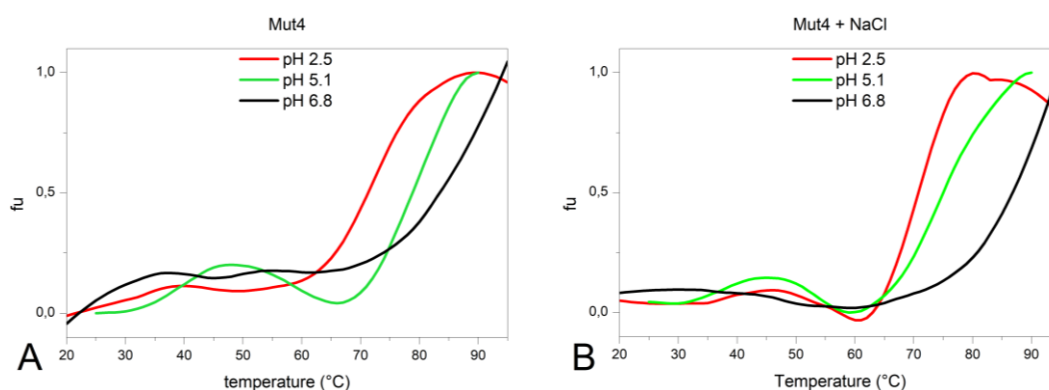


Figure 18. Thermal denaturation of Mut4 monitored by far-UV CD at 215 nm and expressed as fraction unfolded (fu) as a function of temperature (20-90 °C). Measurements were performed at pH 2.5, 5.1 and 6.8 in the absence (A) and presence (B) of 150 mM NaCl. All curves display cooperative unfolding transitions, with melting temperatures dependent on pH and ionic strength. In the absence of NaCl, T_m increases with pH, whereas under saline conditions the transitions are shifted towards higher temperatures, particularly at pH 5.1 and 6.8. The lack of a well-defined post-transition plateau indicates that the reported T_m values should be considered apparent estimates suitable for relative comparison between conditions rather than absolute thermodynamic parameters.

Consequently, the reported T_m values should be considered apparent estimates, useful for relative comparison between conditions rather than as rigorous thermodynamic parameters. Despite this limitation, these data indicate that ionic strength has a limited effect on the stability of Mut4 at strongly acidic pH (red curves in Figure 18), while at pH 5.1 (green curves in Figure 18) and especially at pH 6.8 (black curves in Figure 18) it contributes to a moderate stabilisation of the mutant, a behaviour consistent with the idea that, at pH values close to neutrality, the modulation of surface charges and the consequent balancing of electrostatic interactions may contribute to the conformational stability of the protein, as already observed for MNEI and similar systems. [77,179]

Lastly, the reversibility of the unfolding process was evaluated by comparing the CD spectra recorded before and after the thermal denaturation cycle under all pH and ionic strength conditions at 150 mM NaCl (Figure 19).

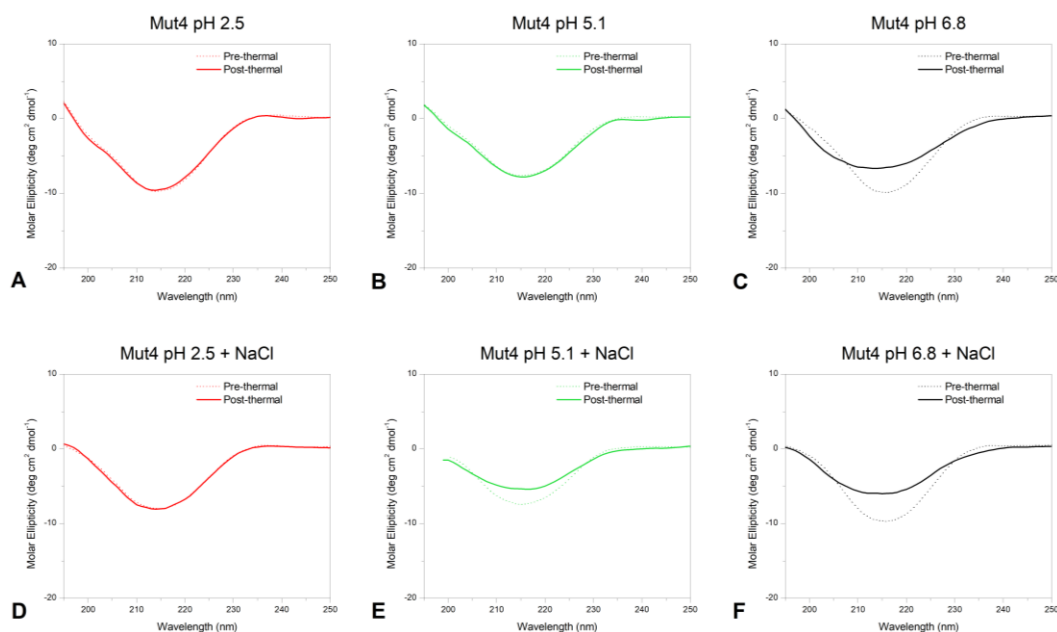


Figure 19. Far-UV CD spectra of Mut4 recorded before (pre-thermal) and after (post-thermal) the thermal denaturation cycle at pH 2.5, 5.1 and 6.8, in the absence (A-C) and presence (D-F) of 150 mM NaCl. At pH 2.5, the pre- and post-thermal spectra are largely superimposable under both ionic conditions, indicating a predominantly reversible unfolding process. At pH 5.1, reversibility remains substantial in the absence of NaCl, while the presence of salt leads to a modest reduction in ellipticity after heating. At pH 6.8, more pronounced differences between pre- and post-thermal spectra are observed, with reduced negative ellipticity and altered spectral profiles, consistent with partial irreversibility of unfolding and the formation of non-refolding species, particularly under saline conditions.

At pH 2.5, both in the absence and presence of NaCl, the post-thermal spectra are substantially superimposable on those recorded before heating, indicating that Mut4 is able to recover its secondary structure almost completely upon return to the initial conditions and that the process is predominantly reversible on the time scale of the experiment. At pH 5.1 in the absence of salt, the differences between pre- and post-thermal spectra remain limited, suggesting a good refolding capacity; in the presence of NaCl, however, a slight reduction in the intensity of the minimum is

observed, compatible with a partial loss of ordered structure or with the formation of species unable to refold. The situation changes more markedly at pH 6.8, where the post-thermal spectra show a more pronounced decrease in negative ellipticity and an alteration in the profile compared to the pre-thermal spectra, both with and without NaCl. (Figure 19)

These data indicate that, at pH close to neutrality, the unfolding process of Mut4 has a significantly irreversible component, likely associated with the formation of aggregates or incorrectly refolded conformations that are unable to return to their original structural state. Altogether, CD analysis suggests that Mut4 maintains a well-defined, mainly β -folded structure over a wide pH range, but that its stability and folding reversibility are significantly influenced by both pH and ionic strength, with particular vulnerability to conditions close to neutrality in the presence of NaCl.

4.1.4 Aggregation propensity assessed by ThT fluorescence

The propensity of Mut4 to generate β -rich aggregate species was assessed by monitoring ThT fluorescence under heat stress conditions at pH 2.5 and 6.8, in the absence and presence of 150 mM NaCl. (Figure 20)

The pH 5.1 was not considered because, based on CD spectra and stability analyses, it showed no significant differences compared to the profile observed at acidic pH.

At pH 2.5, the chosen temperature (65 °C) corresponds to approximately 6 °C below the T_m of Mut4 and allows for the exploration of a condition of partial denaturation. Under these conditions, however, the increase in fluorescence remains limited and lacks sigmoidal kinetics, suggesting the absence of a structured aggregation process. The addition of NaCl causes a moderate increase in the signal, but the trend remains essentially the same; this increase seems to be attributable to changes in the local environment of ThT induced by ionic strength, rather than to the emergence of protein structures typically recognised by the probe.

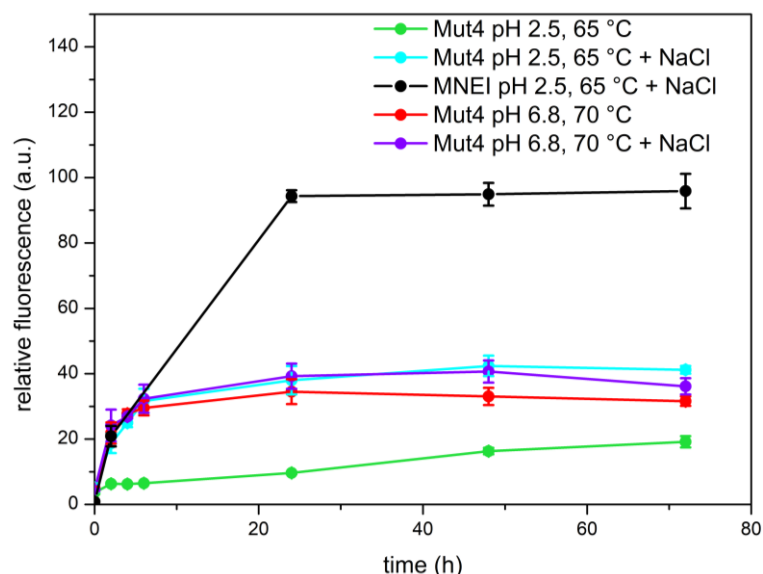


Figure 20. ThT fluorescence values were rescaled by dividing the absolute data by a constant factor (6000) for the sole purpose of facilitating comparison between graphs. Scaling does not alter the kinetic trend or the interpretation of the results.

To contextualize these observations, ThT fluorescence was also monitored for wild-type MNEI under the same acidic and high-temperature conditions in the presence of NaCl. In this case, a rapid and pronounced increase in fluorescence intensity is observed, reaching a plateau within the first 24 h, indicative of a significantly higher propensity of MNEI to access aggregation-prone states under thermal stress. This comparison confirms that the experimental conditions employed are suitable to promote aggregation and highlights the markedly reduced aggregation tendency of Mut4.

At pH 6.8, the experimental temperature (70 °C) is well below the T_m of Mut4, which reaches values of up to ~87 °C in the presence of NaCl. Under these conditions, the protein is predominantly folded and the ThT fluorescence signal remains limited and attributable to marginal interactions of the ThT probe with the protein. Also in this case, neither lag phases nor growth phases attributable to a kinetic trend attributable to an aggregation process are observed.

Transmission electron microscopy (TEM) was performed on samples incubated for 48 hours (Figure 21) under the most aggregation-prone

condition investigated (pH 2.5, 65 °C, 150 mM NaCl), corresponding to the plateau region of the ThT fluorescence profile. TEM images reveal the absence of fibrils, protofibrils, or amorphous aggregates, showing only sparse, poorly defined particulate material compatible with background or non-specific deposition.

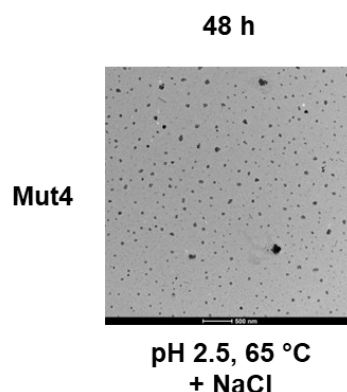


Figure 21. Transmission electron microscopy (TEM) image of Mut4 after 48 h incubation at pH 2.5 and 65 °C in the presence of 150 mM NaCl. These conditions correspond to the plateau region of the ThT fluorescence profile and represent the most aggregation-prone condition investigated. The micrograph shows the absence of fibrillar structures and does not reveal amorphous aggregates. Only sparse, poorly defined particulate material is observed, likely attributable to background staining or non-specific deposition. Scale bar: 500 nm.

The data set indicates that Mut4 does not have a significant tendency to access misfolded aggregation pathways under the conditions explored, even when subjected to temperatures close to its T_m , conditions that promote aggregation in MNEI. This behaviour is consistent with the good preservation of the secondary structure observed by CD and with the almost complete recovery of the spectra after the thermal cycle.

4.1.7 Structure prediction and MD simulations

Since Mut4 differs from MNEI (PDB ID: 1FA3) exclusively by the E23Q substitution, a 3D model was generated by introducing the mutation directly into the primary sequence of MNEI and the new sequence was provided as

input to AlphaFold2. In this context, the goal was not to “discover” an unknown fold, but rather to verify that the mutation did not substantially alter the topology of the protein and to obtain an energetically consistent model to use as a starting point for molecular dynamics simulations.

The three standard panels returned by AlphaFold2 confirm exactly this scenario. Among the five models generated, *Mut4_b091a_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb* was selected, as it had the best pLDDT profile, the lowest PAE dispersion, and the topology most similar to that of 1FA3. (Figure 22)

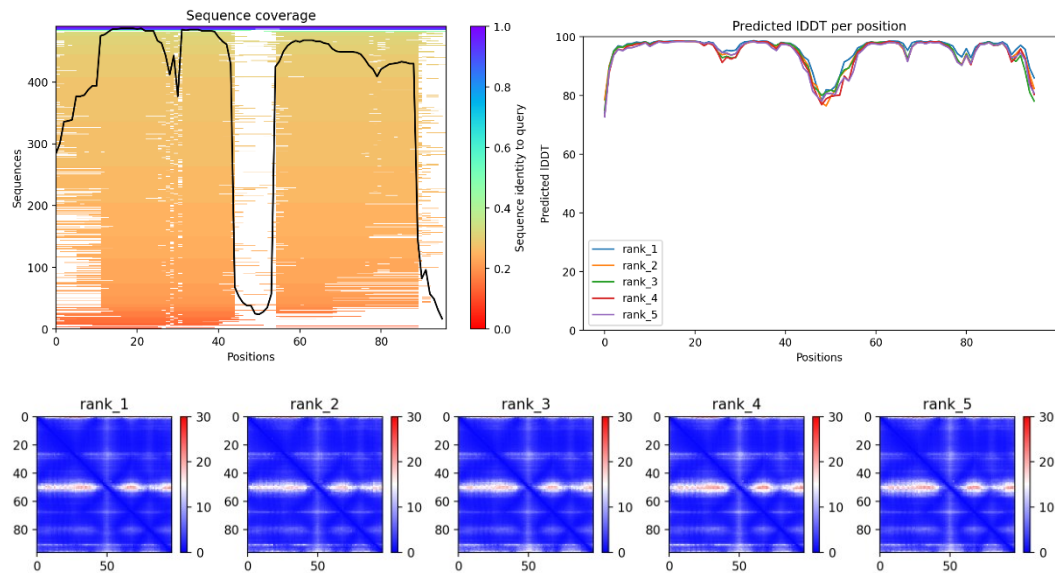


Figure 22. AlphaFold2 output for the Mut4 (E23Q) sequence generated starting from the wild-type MNEI primary structure. Top left: multiple sequence alignment (MSA) coverage across the sequence, showing high sequence depth over most positions and a local reduction around the central region. Top right: per-residue predicted local distance difference test (pLDDT) scores for the five ranked models, indicating uniformly high confidence over the structured regions and local confidence drops limited to flexible segments. Bottom row: predicted aligned error (PAE) maps for the five models (rank_1 to rank_5), showing low inter-residue error across the core of the protein and absence of alternative domain arrangements. Together, these metrics indicate that the E23Q substitution does not alter the global fold relative to the MNEI reference structure (PDB 1FA3), supporting the use of the selected model as a structurally consistent starting point for molecular dynamics simulations.

The model was considered suitable as an initial structure for MD simulations, given that the E23Q mutation does not introduce fold variations but only a local modification of the residue chemistry.

For visual comparison, the AlphaFold-predicted structure of Mut4 was superimposed onto the crystallographic structure of wild-type MNEI, confirming an almost perfect structural overlap and supporting the conclusion that the E23Q substitution does not induce detectable global conformational rearrangements. (Figure 23)

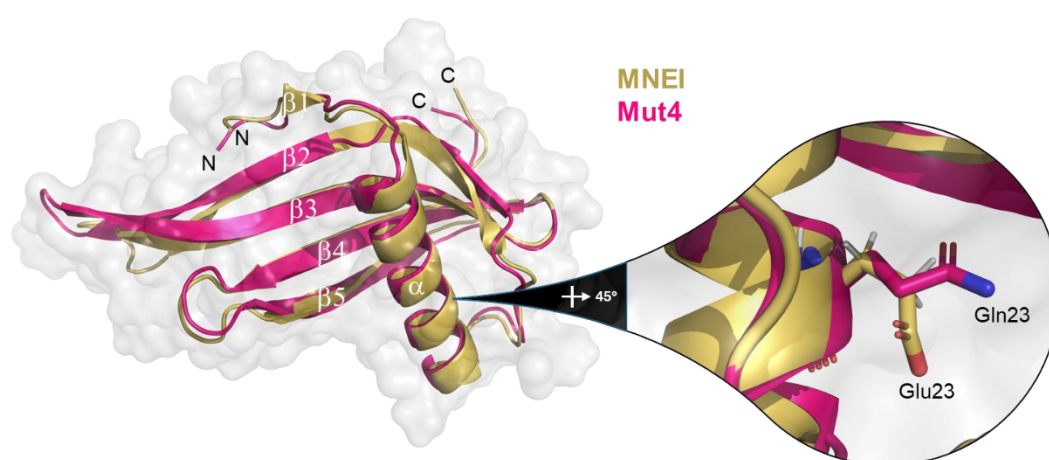


Figure 23. Superposition of the AlphaFold-predicted structure of Mut4 (hotpink) and the crystallographic structure of wild-type MNEI (yelloworange; PDB ID: 1FA3). The two structures show an almost complete overlap, indicating that the E23Q substitution does not induce detectable global conformational changes and that the overall fold of the protein is preserved. In the bubble, magnification of the residues involved in the substitution and their positions in the buried region behind the helix, where the substituent residue adopts a slightly different orientation compared to the substituted residue, but retains the side-chain size and hydrogen-bonding capability.

4.1.7.1 RMSD analysis

The overall stability of Mut4 was assessed by monitoring the RMSD of the backbone relative to the initial structure predicted by AF2, in MD simulations at pH 2.5 and 6.8 conducted at 300, 350, 400, and 450 K (Figure 24). Three independent replicates were performed for each condition, with trajectories of 200 ns. At 300 K, both systems show a rapid initial settling phase followed by stabilisation at comparable average values (pH 2.5: 2.40 ± 0.37 Å; pH

6.8: 2.36 ± 0.43 Å). A similar behaviour is observed at 350 K (pH 2.5: 2.41 ± 0.45 Å; pH 6.8: 2.51 ± 0.38 Å), indicating that Mut4 maintains a compact fold in this range and that pH does not appreciably alter its overall stability. As the temperature increases to 400 K, the average RMSD increases moderately (pH 2.5: 3.36 ± 0.62 Å; pH 6.8: 3.38 ± 0.61 Å), reflecting the expected increase in thermal flexibility but without evidence of profound structural reorganisations.

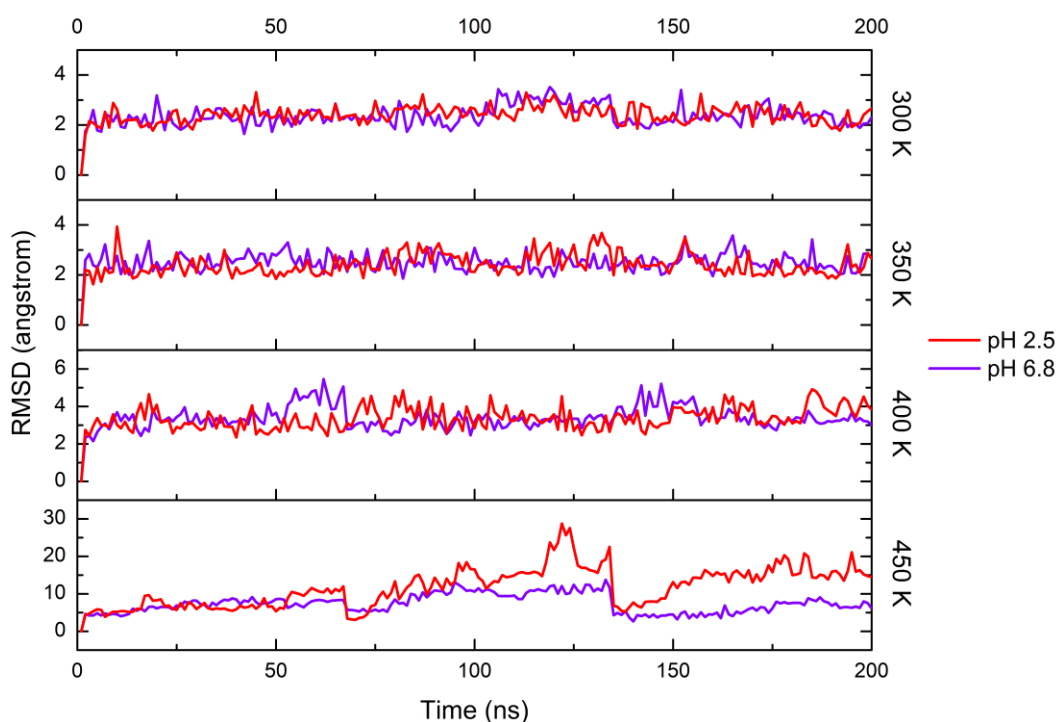


Figure 24. Temporal evolution of the RMSD backbone of Mut4 at pH 2.5 (red) and 6.8 (purple) during molecular dynamics simulations at 300, 350, 400 and 450 K. Stability is the same for the two systems at 300 and 350 K, while at 400 K the deviations increase as the temperature goes up. At 450K, deviations in an acidic environment are more pronounced than those in a neutral one.

The behaviour diverges significantly at 450 K. At pH 2.5, the RMSD increases rapidly to very high average values (11.70 ± 5.23 Å) with large and irregular oscillations, indicative of extensive backbone reorientation and an extended unfolding process. At pH 6.8, on the contrary, the average RMSD remains substantially lower (7.38 ± 2.49 Å) and with less chaotic

fluctuations, suggesting a more gradual and partially ordered structural disassembly. These differences indicate that Mut4, despite losing structure at 450 K in both conditions, is less prone to complete collapse at pH 6.8 than at acidic pH. (Figure 24)

4.1.7.2 RMSF analysis

To investigate how pH and temperature influence the local dynamics of Mut4, the RMSF of C α along the sequence was calculated for each simulated condition. (Figure 25) At 300 K, the profiles at pH 2.5 and 6.8 are almost identical: limited fluctuations for most of the chain (< 1.5 Å) and peaks located in the terminal regions and in specific surface loops, as also observed for wild-type MNEI. [53]

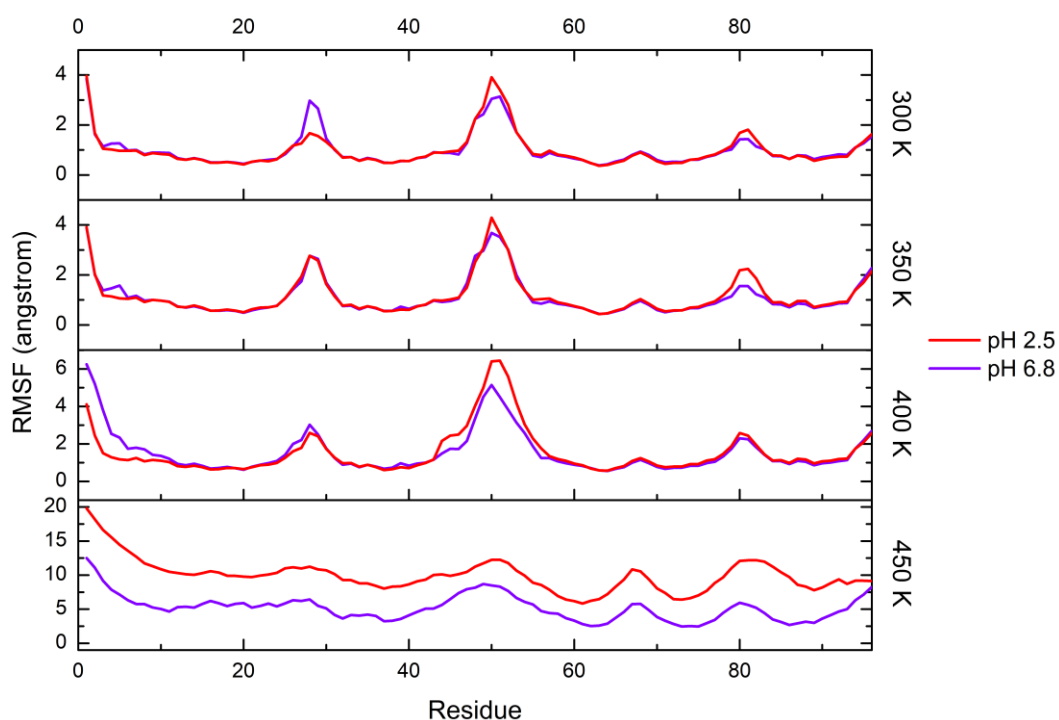


Figure 25. Residue-wise RMSF profiles of C α atoms of Mut4 at pH 2.5 and 6.8 during molecular dynamics simulations at 300, 350, 400 and 450 K, with the addition of MNEI at pH 2.5 as a reference for comparison. At pH 2.5 are reported MNEI (black) and Mut4 (red), at pH 6.8 only Mut4 (purple). Despite the increase in temperature, the RMSF profiles are superimposable up to 400 K. At 450 K, fluctuations in an acidic environment are indicative of a more pronounced collapse than those in a neutral environment.

In particular, three particularly flexible hot-spots can be identified: the first corresponds to the loop between α H- β 2, which contains the E23Q substitution, the second one corresponds to the β 2- β 3 loop, where the Gly50-Phe51 dipeptide was inserted in MNEI, and the third corresponds to the loop between β 4- β 5, which is known to be a critical hinge region in MNEI and plays a role in early unfolding during aggregation. This behaviour remains essentially unchanged at 350 and 400 K. Although the increase in temperature slightly amplifies the fluctuations, but does not change the positioning of the main flexibility peaks. Gln23 remains significantly less flexible than the surrounding loops, with no appreciable pH-dependent differences up to 400 K. This observation indicates that the E23Q substitution does not introduce local destabilisation. At 450 K, the RMSF profile changes dramatically, but with clear pH dependence. At pH 2.5, fluctuations increase dramatically along the entire sequence, especially in the loops and at the ends, indicating extensive structural collapse. At pH 6.8, the increase in fluctuations is still evident but less marked than at acidic pH. (Figure 25)

4.1.7.3 Statistical comparison between pH and temperature conditions

To quantitatively assess the extent to which pH and temperature influence the conformational response of Mut4, the mean values, standard deviations and 95% confidence intervals (CI95) of the RMSD and RMSF, calculated respectively on the backbone and C α of the residues, were analysed jointly. This approach allows truly significant variations to be distinguished from intrinsic fluctuations in molecular dynamics and enables a direct comparison between conditions at pH 2.5 and pH 6.8 across the entire temperature range explored.

In the 300-400 K temperature range, the average RMSD values are extremely similar at both pH levels, with absolute differences always less than 0.1 Å (Δ RMSD: -0.04 Å at 300 K; +0.10 Å at 350 K; +0.02 Å at 400 K).

The confidence intervals of the two systems also show a large overlap, indicating that any deviations fall within the statistical variability of the trajectory. (Table 14)

Backbone RMSD (Å)						
Temperature	pH	Mean	SD	CI95 low	CI95 high	Δ RMSD (Å)
300 K	2.5	2.40	0.37	2.35	2.45	- 0.04
	6.8	2.36	0.43	2.30	2.42	
350 K	2.5	2.41	0.45	2.34	2.47	+ 0.10
	6.8	2.51	0.38	2.46	2.56	
400 K	2.5	3.36	0.62	3.27	3.45	+ 0.02
	6.8	3.38	0.61	3.30	3.47	
450 K	2.5	11.70	5.23	10.97	12.42	- 4.32
	6.8	7.38	2.49	7.04	7.73	

Table 14. Statistical summary of backbone RMSD values of Mut4 calculated from molecular dynamics simulations at pH 2.5 and pH 6.8 across the explored temperature range (300-450 K). For each condition, the mean RMSD, standard deviation (SD) and 95% confidence interval (CI95) are reported. The Δ RMSD values represent the difference between pH 6.8 and pH 2.5 at the same temperature, allowing direct comparison of the global conformational response under different protonation conditions and identification of temperature-dependent divergence beyond intrinsic fluctuations.

RMSF analyses also confirm this substantial dynamic equivalence, with the average Δ RMSF between pH always ≤ 0.03 Å up to 400 K, and the analysis of confidence intervals (95%) on RMSF values shows no appreciable differences between pH 2.5 and pH 6.8. (Table 15) In other words, within this temperature range, Mut4 maintains strictly conserved conformational behaviour at both pH values, with no signs of selective destabilisation or appreciable alterations in the more flexible loop regions. The situation

changes markedly at 450 K, a temperature at which both systems show clear signs of unfolding, but with different intensities depending on the pH. At pH 2.5, the average RMSD reaches 11.70 ± 5.23 Å (CI95 low 10.97, high 12.42 Å), while at pH 6.8 a significantly lower value is observed, equal to 7.38 ± 2.49 Å (CI95 low 7.04, high 7.73 Å). (Table 14) The Δ RMSD between conditions therefore reaches -4.32 Å, indicating that the neutral system tends to maintain a less distorted structure despite extreme heating.

RMSF per residue (Å)						
Temperature	pH	Mean	SD	CI95 low	CI95 high	Δ RMSF (Å)
300 K	2.5	0.98	0.68	0.84	1.12	+ 0.01
	6.8	0.99	0.65	0.86	1.12	
350 K	2.5	1.15	0.76	1.00	1.31	- 0.03
	6.8	1.12	0.73	0.97	1.26	
400 K	2.5	1.54	1.19	1.30	1.78	+ 0.00
	6.8	1.54	1.12	1.32	1.77	
450 K	2.5	9.94	2.37	9.47	10.42	- 4.76
	6.8	5.18	1.88	4.81	5.56	

Table 15. Statistical summary of per-residue RMSF values of Mut4 calculated on C α atoms at pH 2.5 and pH 6.8 across the explored temperature range (300-450 K). For each condition, mean RMSF, standard deviation (SD) and 95% confidence interval (CI95) are reported. The Δ RMSF values quantify the difference between pH 6.8 and pH 2.5 at each temperature, enabling assessment of pH-dependent modulation of local flexibility and discrimination between genuine dynamical effects and thermal noise.

This result is corroborated by the behaviour of the RMSF, where at 450 K the fluctuations per residue increase dramatically in both conditions, but with a significantly greater average amplitude at pH 2.5 (9.94 ± 2.37 Å, CI95 low 9.47, high 10.42 Å) compared to pH 6.8 (5.18 ± 1.88 Å, CI95 low 4.81, high

5.56 Å), with a Δ RMSF of approximately -4.76 Å. (Table 15) The fact that at pH 6.8 the increase in flexibility is much more limited than in the acidic environment indicates that, even at 450 K, the protein manages to maintain significant portions of its original architecture.

In other words, the loss of compactness remains less drastic, and some interactions characteristic of the native fold continue to exert a stabilising effect.

Overall, statistical analyses indicate that Mut4 exhibits virtually identical stability and dynamics at the two pH values up to 400 K, while differences emerge only at 450 K, where denaturation becomes dominant. In fact, in this extreme condition, the system at pH 6.8 maintains significantly lower structural fluctuations and deviations, suggesting a more marked partial resistance to unfolding compared to the acidic pH. Although this effect occurs exclusively in a computational rather than physiological regime, it provides interesting insights into the stability landscape of the protein, consistent with the greater compactness observed under less protonated conditions.

4.1.8 Discussion

Within the aims of this thesis, Mut4 was investigated to evaluate how the removal of a buried negative charge influences thermal stability, molecular flexibility, and access to aggregation pathways under different pH and ionic strength conditions.

The mutant was expressed after low-temperature induction and purified by cation exchange chromatography, which was slightly less uniform than for the MNEI protein. In fact, as a consequence of the redistribution of surface charges caused by the neutralisation of Glu23, the surface of Mut4 is slightly different from that of MNEI and the electrostatic interaction with the cation exchange resin is stronger, requiring a slightly higher ionic strength for elution and consequently producing a slightly more heterogeneous chromatographic profile than MNEI. Nevertheless, a good yield of soluble protein is obtained. Circular dichroism analysis of this mutant was

performed at three different pH values, pH 2.5, 5.1 and 6.8, both in the presence and absence of salt. Under all pH and salinity conditions, Mut4 maintains a very stable structure rich in β -sheets. Analysis of the thermal denaturation curves reveals an interesting finding already observed in MNEI: as the pH increases, the T_m increases progressively. In fact, in the absence of NaCl, the T_m of Mut4 increases gradually when moving from an acidic environment ($\sim 71^\circ\text{C}$) to a neutral one ($\sim 82^\circ\text{C}$). However, in the presence of salt, although the T_m remains unchanged at acidic pH, this effect becomes even more evident, especially at pH 6.8, where the Mut4 protein reaches a very high T_m ($\sim 87^\circ\text{C}$). Despite its thermal stability, the unfolding process tends to become less and less reversible as the environment approaches neutrality from acidic, and this effect is more evident in the presence of salt. Probably during denaturation, the presence of salt and a pH tending towards neutral allow the exploration of alternative states to the more stable native folds that prevent renaturation.

Nevertheless, ThT fluorescence and TEM electron microscopy data were used to assess whether the irreversibility observed after thermal denaturation profiles monitored by CD was associated with the formation of β -rich aggregates.

The protein shows a very low tendency to aggregate, both in acidic and neutral environments, even when subjected to thermal stress close to the melting temperature, regardless of the presence of salt. Furthermore, the fact that no sigmoidal ThT kinetics are observed and that the intensity of the signals obtained for the mutant are significantly lower than those obtained for MNEI under the same conditions may explain why there are no fibrils or amorphous aggregates in the TEM images. These results reinforce the idea that the E23Q mutation significantly reduces accessibility to aggregation pathways normally explored by MNEI under stress conditions.

As no crystallographic structure is available, computational analyses on a predicted structure provide further insight into this mutant. Simulations show that between 300 and 400 K, there are no significant differences at pH 2.5 and 6.8, both in terms of global RMSD and local RMSF. This result is not

only qualitative, but is supported by statistical analysis of the trajectories. In fact, in the temperature ranges considered, the average differences between the two pH values remain below 0.1 Å for the RMSD and a few hundredths of an Å for the RMSF, with 95% confidence intervals that largely overlap.

Furthermore, glutamine at position 23 shows low and non-anomalous fluctuations, in line with the idea that the effect of the mutation is not to create a more flexible point, but rather to neutralise the negative charge that can destabilise the structure. At 450 K, the simulations provide an additional interpretative element, because Mut4 collapses dramatically at both pH conditions values, although the effect is less pronounced at pH 6.8, consistent with the greater thermal stability observed experimentally with CD analyses and with the statistics analysed. In particular, the Δ RMSF values confirm that the protein in neutral conditions manages to maintain significant portions of its original architecture compared to the acidic condition.

In the case of Y65R, a variant of MNEI described by Pica *et al.*, the introduction of a positive charge makes the protein more sensitive to protonation variations and promotes, especially at acidic pH, the formation of fibrillar aggregates with ordered morphology, while at pH 6.8 the protein, while maintaining good stability, is able to form amorphous aggregates. [180] Similarly, Mut9, the super-stable mutant reported by Lucignano *et al.*, which includes E23A among its substitutions, does not prevent it from opening near the helix and accessing aggregation pathways at both acidic and neutral pH, regardless of the presence of NaCl, despite its marked conformational and thermal stability greater than that of MNEI. [78] Mut4 behaves differently because the substitution does not introduce new interactions or additional charges, but rather eliminates the source of perturbation associated with the presence of a buried ionisable residue. The consequence is not only an increase in stability, but above all the loss of pH dependence and accessibility to aggregation pathways.

The neutralisation of residue 23 may have blocked the local opening of the α -helix in Mut4, which, in MNEI and other mutants, acts as a trigger point for access to aggregation pathways. Therefore, Mut4 is an effective example of a rational mutation capable not only of increasing thermal robustness and stabilising the fold through a reorganisation of the dynamics of the α -helix and its interactions, but also of substantially reducing access to the aggregation pathways observed in MNEI and its variants.

4.2 Circular permuted mutants of the sweet protein MNEI

4.2.1 Design of Perms

The conformational stability of sweet proteins, and particularly MNEI, is known to be sensitive to extreme physical and chemical conditions such as high temperatures, pH variations and intense pressures, which are frequently encountered in industrial food processing. [53,181]

In this context, rational engineering of the protein structure is a crucial strategy for obtaining more robust variants without altering their sensory properties. Alongside classical approaches based on the substitution of specific residues [76,90,182,183], circular permutation has established itself as an alternative method for modulating stability, dynamics and folding pathways without introducing point mutations in the sequence. [117]

Circular permutation requires the covalent connection of the native ends (N- and C-terminal) by means of a linker peptide and the simultaneous introduction of new ends within the polypeptide chain through the controlled breaking of a pre-existing peptide bond. This topological reorganisation alters the linear order of residues, while maintaining the amino acid composition unchanged and the overall 3D structure preserved, and can significantly modify the energy landscape of folding, stabilising local states or reducing susceptibility to partial unfolding events. [117,184]

In the case of MNEI, a recombinant monomer obtained by combining the two native chains using a short Gly-Phe linker, the presence of the artificial ring constitutes a structural element susceptible to thermal disturbances and represents a natural candidate for the introduction of circular permutation. [185,186] Based on these considerations and supported by predictive analyses obtained with the CPred tool [187], we developed three circularly permuted variants, named Perm1, Perm2 and Perm3. [91] For all three variants, the new opening site was located between residues Gly50 and Phe51, in the loop connecting the β 2 and β 3 strands. (Figure 26)

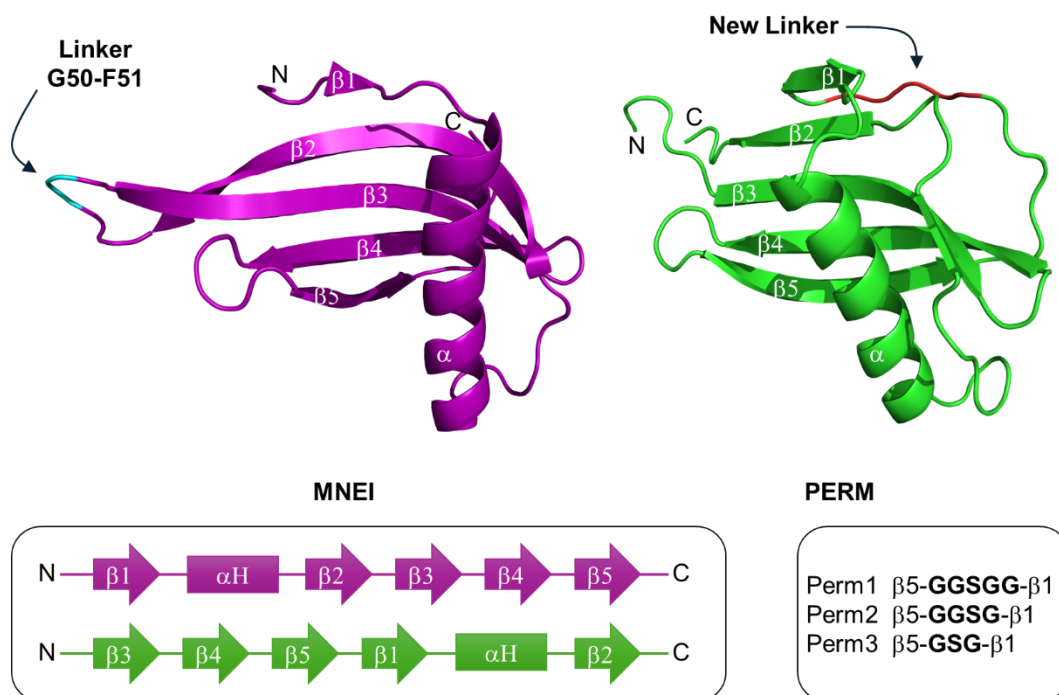


Figure 26. 3D representation of MNEI (top left, PDB ID: 1FA3) and a circularly permuted mutant modeled with AlphaFold2 (top right). In the permuted construct, the original N and C terminals are covalently linked via a linker (red), while new terminals are generated between Gly50 and Phe51 in the loop connecting $\beta 2$ and $\beta 3$ (cyan linker in MNEI), resulting in a rearrangement of the linear sequence without altering the overall cystatin-like α/β fold. The bottom left shows the sequential arrangement of the secondary structure elements in MNEI and in the circularly permuted mutants, highlighting the topological reorganisation underlying the permutation strategy. The bottom right shows a detail of the linker content.

This region has relatively high mobility in the structures reported in the literature and is less involved in maintaining the hydrophobic core of the fold, characteristics that make it a permissive site for the generation of new peptide ends without disturbing the overall topology. [53] The original ends of MNEI were instead connected by three different linker peptides consisting exclusively of glycine and serine residues (GGSGG, GGSG and GSG, respectively). The use of linkers formed by Gly and Ser, amino acids with low steric occupancy and marked conformational flexibility, is widely documented as favourable to the correct folding of permuted proteins, as it minimises local tensions and reduces the risk of unfavourable entropic constraints. At the same time, modulating the length of the linker allows different trade-offs between flexibility and topological stabilisation to be

explored: a longer linker can facilitate the correct positioning of the new ends during folding, while a shorter one can provide greater rigidity and potentially increase thermal stability. [188,189] For this reason, development has focused exclusively on Perm1-3 mutants derived from the Gly50-Phe51 break.

4.2.2 Protein production and purification

The three circularly permuted MNEI mutants were produced by heterologous expression in *Escherichia coli* BL21(DE3), a strain commonly used for the overexpression of MNEI and its mutants, as widely reported in the literature. [76,78,180,190,191] All procedures for expression, induction, cell harvesting, and lysis with a sonicator are described in detail in Methods (see Section 3.2). This workflow enabled reproducible recovery of soluble recombinant protein for all three variants, providing a reliable basis for subsequent purification (see Section 3.3) and comparative analysis of expression yield and oligomeric behavior.

The soluble fraction was initially subjected to cation exchange chromatography (see Section 3.3.3), allowing the target protein to be concentrated and most of the high molecular weight or highly charged impurities to be removed. Purification was then completed first with a desalting step on a Sephadex™ G-25 column, followed by size exclusion chromatography (see Section 3.3.4), used both to obtain the highly pure monomeric form and to analyse the oligomeric state of the variants under native conditions. In all three cases, excellent separation between the recombinant protein and the lysate components was achieved, with final yields in the order of ~30 mg per litre of culture, significantly higher than those typically reported for wild-type MNEI (~12 mg/L in our conditions). This increase in yield suggests that the permuted mutants, despite topological reorganisation, maintain good solubility and do not present marked problems of degradation or misfolding during expression.

However, a marked diversity in the behaviour of the three mutants emerged: Perm1 eluted as a well-defined single monomeric species, indicating that

the longer linker does not significantly alter the compactness or conformational stability of the protein. (Figure 27) Perm2 showed predominantly the monomeric form, accompanied by a minor but reproducible fraction of dimer. Perm3 showed the most complex profile, characterised by comparable amounts of monomer and dimer, with the additional appearance of a small population of high molecular weight species.

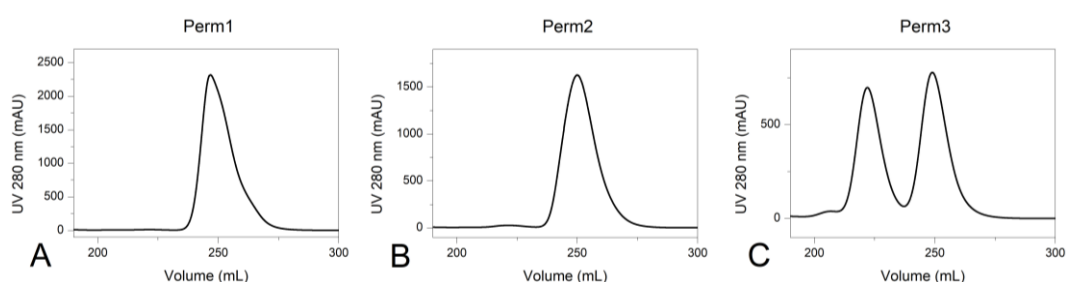


Figure 27. Size-exclusion chromatography profiles of the circularly permuted variants Perm1 (A), Perm2 (B) and Perm3 (C) recorded on a Superdex 75 26/600 column. Separations were performed in 50 mM sodium acetate buffer at pH 5.5 supplemented with 150 mM NaCl, allowing comparison of the elution behaviour of the three constructs under identical chromatographic conditions.

4.2.3 Stability and aggregation studies

4.2.3.1 CD spectroscopy and secondary stability evaluation

The impact of circular permutation on the secondary structure of the three mutants was evaluated by circular dichroism spectroscopy under strongly acidic (pH 2.5) and almost neutral (pH 6.8) conditions, chosen respectively as typical conditions for the induction of MNEI fibrillation and as a more physiological structural reference. (Figure 28) [53]

At pH 2.5, all three mutants show spectra with common characteristics, namely a main minimum in the 211-213 nm region, indicative of the dominant presence of β elements, and a less pronounced shoulder/minimum around 201 nm. However, significant differences emerge in the position and intensity of the minima. Perm1 and Perm2 have almost overlapping profiles, with two minima centred at ~201 and ~212 nm, respectively, perfectly consistent with a structurally well-organised state.

Perm3, on the other hand, shows a slight blue shift of the minima (204 and 213 nm), accompanied by a reduction in the overall signal intensity, and the low wavelength region appears less defined, suggesting a greater disordered component, a plausible phenomenon given the greater propensity for oligomerisation observed for this mutant. (Figure 28)

The quantitative deconvolution of the spectra performed with BeStSel [192] and reported in Table 16, carried out in the range 195-250 nm, confirms this picture: the β content is almost identical for all mutants (44-48%), while the α -helix content ranges from ~23% in Perm1 to ~11% in Perm2, down to much lower values in Perm3 (~6%), which also shows the highest random coil content (approximately 47% compared to ~30-35% in Perm1 and Perm2).

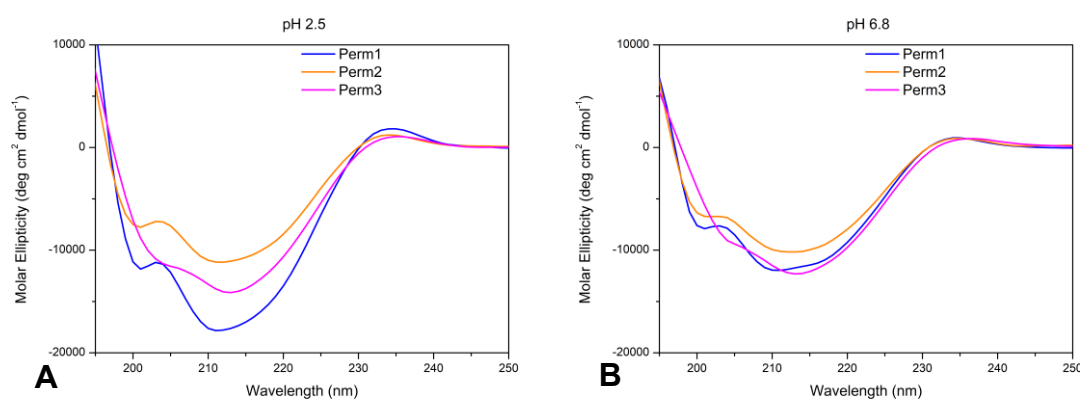


Figure 28. CD spectra (A) and thermal denaturation profiles (B) of Perm1 (blue line), Perm2 (orange line) and Perm3 (pink line) in 20 mM sodium phosphate buffer at pH 2.5 in the absence of 150 mM NaCl.

Under almost neutral conditions (pH 6.8), the behaviour of the three mutants remains broadly unchanged compared to pH 2.5, with a main minimum between 211 and 214 nm and a second minimum at 202-205 nm. In fact, Perm1 shows minima at 211 and 202 nm and maintains a secondary structure distribution similar to that observed at pH 2.5, with ~15% α -helix, ~47% β -sheet and a moderate amount of random coil (~32%). Perm2, on the other hand, with minima at 213 and 202 nm, shows a slight further reduction in α -helix (~11%) and β -sheet (~45%) but an increase in

disordered content (~38%), while Perm3, as in the acidic condition, combines the lowest helical content (~7%) with the lowest β -sheet content (~44%) and the highest random coil fraction (~45%) among the three mutants. (Table 16)

The structural stability of circularly permuted mutants was evaluated by monitoring the variation of the CD signal at 215 nm as a function of temperature, in the range 20-90 °C, both at pH 2.5 and pH 6.8. (Figure 29) In Table 16, for each pH, in addition to the percentage values of the deconvolutions performed on each pre-thermal spectrum, the estimated melting temperatures (T_m) for each protein are reported.

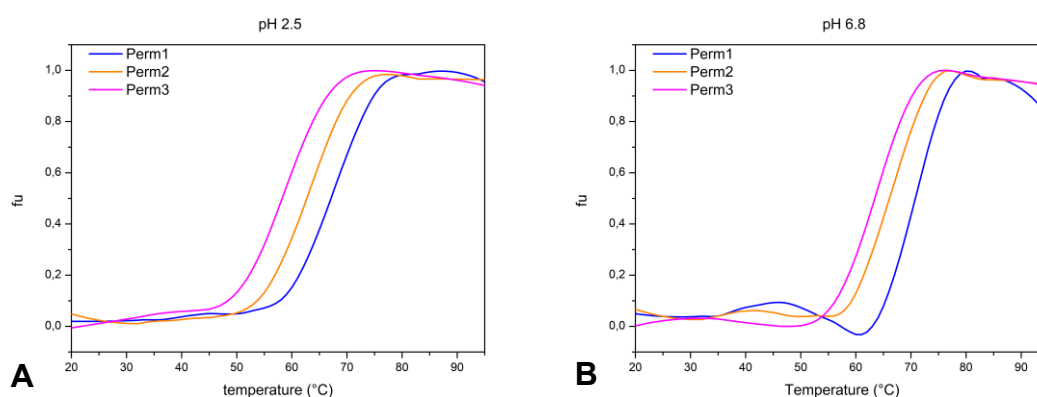


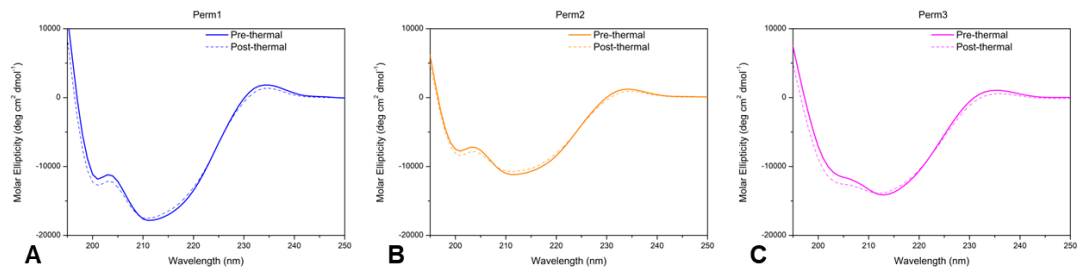
Figure 29. The fraction of unfolded protein (fu), derived from the CD signal at 215 nm, is reported as a function of temperature for the circularly permuted variants Perm1, Perm2 and Perm3 at pH 2.5 (A) and pH 6.8 (B), in the absence of 150 mM NaCl. All constructs display a cooperative thermal transition consistent with a global unfolding process. Differences among the permutants are reflected in the position and slope of the transitions, indicating distinct thermal stabilities associated with the different circular permutation patterns. Comparison between the two pH conditions reveals an overall shift of the transitions towards higher temperatures at pH 6.8, suggesting a stabilising effect of near-neutral pH on the folded state.

In fact, it can be observed that, at pH 2.5, the melting curves show cooperative transitions for all mutants, with melting temperatures (T_m) of ~67 °C for Perm1, ~63 °C for Perm2 and ~58 °C for Perm3. All values are lower than those of MNEI measured under the same conditions [90], but Perm1 remains the variant most similar to the parent protein, while Perm3

is the least stable. In addition, to evaluate the reversibility of the protein, spectra were recorded after denaturation, and it is interesting to note that the proteins showed characteristics of reversibility, recovering their initial secondary structure content. (Figure 29)

At pH 6.8, as expected for an environment more favourable to overall folding stability, the Tms are higher: ~71 °C for Perm1, ~66 °C for Perm2 and ~64 °C for Perm3. Again, none of the variants exceed the stability of MNEI at this pH [90], but Perm1 comes very close to the wild type, with a penalty of a few degrees, while Perm2 and Perm3 are more destabilised. (Figure 29) The spectra recorded after thermal denaturation, on the other hand, show an inability to refold. (Figure 30)

pH 2.5



pH 6.8

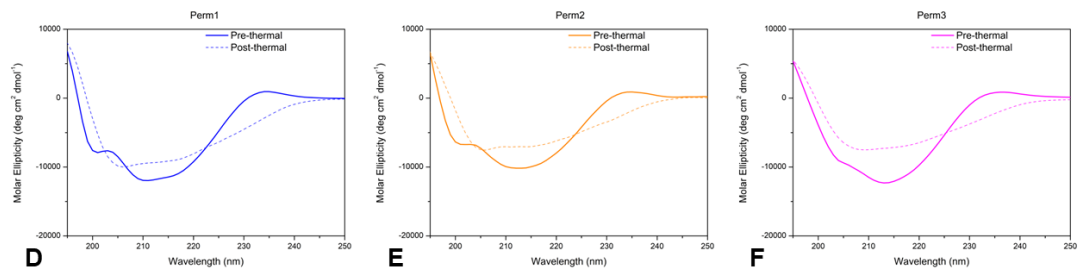


Figure 30. Far-UV CD spectra recorded before (continuous lines) and after (dashed lines) the thermal denaturation cycle are shown for Perm1, Perm2 and Perm3 at pH 2.5 (C-E) and pH 6.8 (D-F), in the absence of 150 mM NaCl. At pH 2.5, the pre- and post-thermal spectra are largely superimposable for all permutants, indicating that unfolding is predominantly reversible under acidic conditions. In contrast, at pH 6.8 the post-thermal spectra show a reduction in negative ellipticity and alterations in spectral shape, pointing to a significant irreversible component of the unfolding process, likely associated with incomplete refolding or the formation of non-native species.

pH 2.5				pH 6.8		
	Perm1	Perm2	Perm3	Perm1	Perm2	Perm3
α-helix (%)	23	11.2	6.2	15	10.6	6.9
β-sheet (%)	46.4	47.8	44.7	46.7	44.5	43.8
Turn	0	5.9	1.8	6.0	6.9	4.3
Random coil (%)	30.6	35.1	47.2	32.4	37.9	45.1
T_m (°C)	67 ± 0.5	63 ± 0.5	58 ± 0.5	71 ± 0.5	66 ± 0.5	64 ± 0.5

Table 16. Secondary structure content of Perm1, Perm2 and Perm3 at pH 2.5 and pH 6.8, in the absence of 150 mM NaCl, estimated by deconvolution of far-UV CD spectra with BeStSel [192], together with the corresponding melting temperatures (T_m) derived from thermal denaturation curves. All variants exhibit a predominantly β -sheet-rich architecture, with relatively minor contributions from α -helical elements. The melting temperatures increase systematically at pH 6.8 compared to pH 2.5, consistent with enhanced thermal stability under near-neutral conditions and with a modulation of stability imposed by the circular permutation.

Although these circularly permuted mutants were designed to obtain better alternatives to MNEI, experiments conducted in our laboratory and not reported here have shown that they do not reveal flavour like the sweet protein MNEI. Therefore, in addition to the interest related to their sweetness, in order to assess whether circular permutation could in some way influence the tendency of mutants to form fibrillar species as in MNEI, preliminary tests were carried out to study the thermal stability of Perm mutants in the same way as MNEI, known as a model protein for the study of aggregation, when exposed to extreme conditions of temperature, around T_m, acidic pH and ionic strength, it is capable of giving rise to amyloid fibrils, similar to those found in many brain diseases, which can be used for biotechnological applications. [88,180,193–195] Therefore, the results obtained at pH 2.5 and in the presence of 150 mM NaCl show that the CD spectra of Perm1 and Perm2 have the same profile with two minima at ~200

nm and ~212 nm, while in the spectra of Perm3 the minima are shifted towards blue at ~204 and ~214 nm, respectively. (Figure 31)

This behaviour is supported by the quantitative estimation of the secondary structure content of the three mutants in the Table 17, which shows that while Perm1 and Perm2 have a similar secondary structure content (*i.e.* almost the same content of α -helix, β -sheet and random coil) with ~13-17% α -helix, ~46-47% β -sheet and ~31-32% random coil, Perm3, while maintaining a similar content of β -sheet (~47%), shows a reduced content of α -helix of ~4% and a greater amount of random coil of ~46%. This phenomenon is confirmed at pH 6.8, where the deconvolution of the spectra shows behaviour similar to that observed at acidic pH. (Table 17)

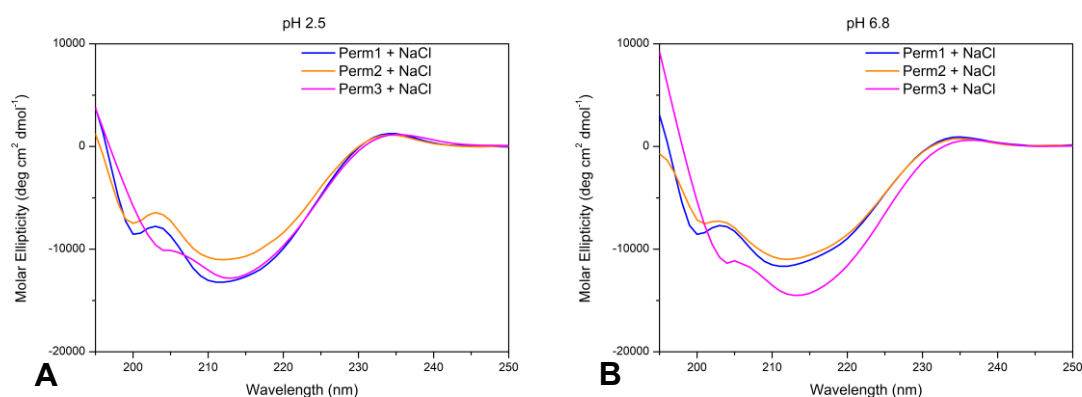


Figure 31. CD spectra (A) and thermal denaturation profiles (B) of Perm1 (blue line), Perm2 (orange line) and Perm3 (pink line) in 20 mM sodium phosphate buffer at pH 2.5 in the presence of 150 mM NaCl.

However, since thermal stability analysis is a fundamental and preliminary step in defining the experimental conditions useful for evaluating the tendency of mutants to form protein fibrils, the thermal denaturation profile was evaluated under the same conditions as in the absence of NaCl. (Figure 32)

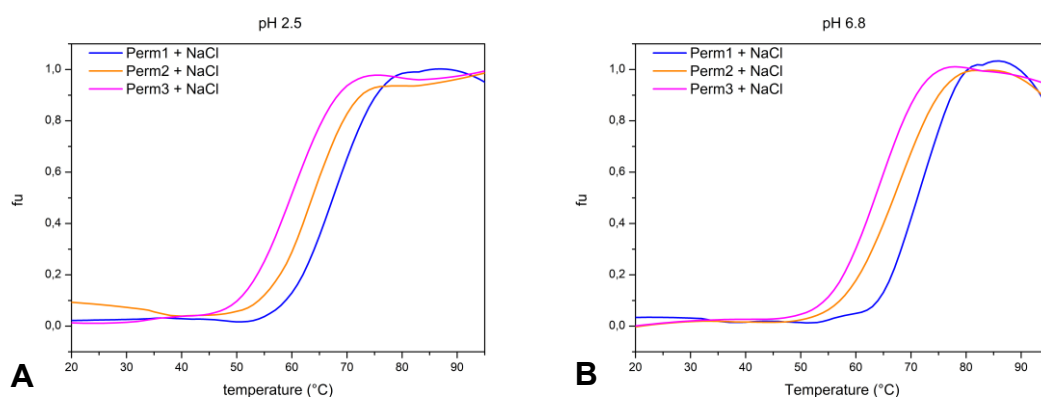
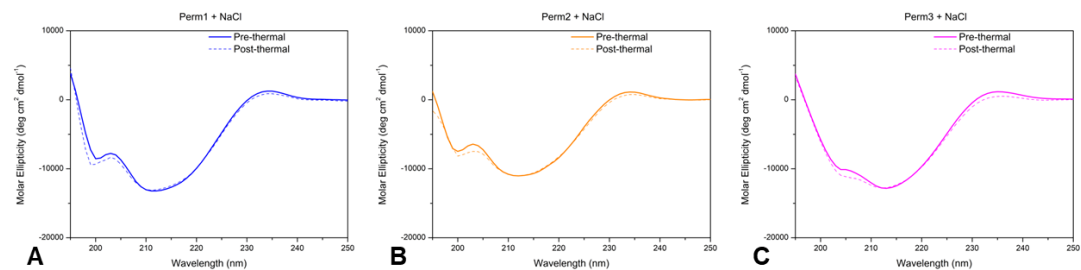


Figure 32. The fraction of unfolded protein (fu), derived from the CD signal at 215 nm, is reported as a function of temperature for the circularly permuted variants Perm1, Perm2 and Perm3 at pH 2.5 (A) and pH 6.8 (B), in the presence of 150 mM NaCl. All constructs display a cooperative thermal transition consistent with a global unfolding process. Differences among the permutants are reflected in the position and slope of the transitions, indicating distinct thermal stabilities associated with the different circular permutation patterns. Comparison between the two pH conditions reveals an overall shift of the transitions towards higher temperatures at pH 6.8, suggesting a stabilising effect of near-neutral pH on the folded state.

It is intriguing to note that, even in the presence of NaCl, none of the mutants show greater thermal stability than MNEI; on the contrary, Perm1 is the closest to MNEI, with a difference of $\sim 10^\circ$ at pH 2.5 and $\sim 3^\circ$ at pH 6.8. In addition, the CD spectra of the three circularly permuted mutants recorded before and after thermal denaturation in the presence of NaCl, as well as in the absence of NaCl, showed that at pH 2.5 denaturation is completely reversible for all three mutants, while at pH 6.8 they are unable to refold completely, suggesting that at neutral pH exposure to high temperatures makes them more susceptible to misfolding. (Figure 33)

pH 2.5



pH 6.8

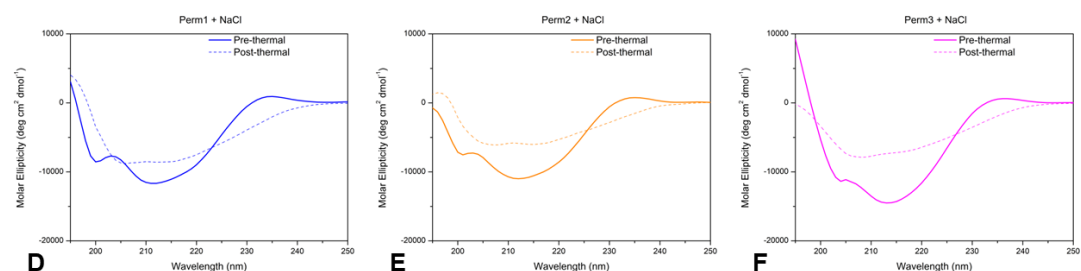


Figure 33. The fraction of unfolded protein (f_u), derived from the CD signal at 215 nm, is reported as a function of temperature for the circularly permuted variants Perm1, Perm2 and Perm3 at pH 2.5 (A) and pH 6.8 (B), in the presence of 150 mM NaCl. All constructs display a cooperative thermal transition consistent with a global unfolding process. Differences among the permutants are reflected in the position and slope of the transitions, indicating distinct thermal stabilities associated with the different circular permutation patterns. Comparison between the two pH conditions reveals an overall shift of the transitions towards higher temperatures at pH 6.8, suggesting a stabilising effect of near-neutral pH on the folded state.

pH 2.5				pH 6.8		
	Perm1	Perm2	Perm3	Perm1	Perm2	Perm3
α-helix (%)	17	13.4	4.4	15.4	12.6	4.3
β-sheet (%)	47.6	46.4	47.1	43.6	40.7	49.4
Turn	4.4	8.3	2.7	7.4	7.4	1.4
Random coil (%)	31	31.9	45.8	33.6	39.3	44.9
T_m (°C)	67 ± 0.5	63 ± 0.5	60 ± 0.5	71 ± 0.5	67 ± 0.5	63 ± 0.5

Table 17. Secondary structure content of Perm1, Perm2 and Perm3 at pH 2.5 and pH 6.8, in the presence of 150 mM NaCl, estimated by deconvolution of far-UV CD spectra with BeStSel [192], together with the corresponding melting temperatures (T_m) derived from thermal denaturation curves. All variants exhibit a predominantly β -sheet-rich architecture, with relatively minor contributions from α -helical elements. The melting temperatures increase systematically at pH 6.8 compared to pH 2.5, consistent with enhanced thermal stability under near-neutral conditions and with a modulation of stability imposed by the circular permutation.

4.2.3.2 ThT analysis and aggregation curves

With the intention of assessing the propensity of circularly permuted mutants to form fibrillar species or amorphous aggregates, the aggregation kinetics were monitored using Thioflavin T (ThT), a fluorophore that increases its emission in the presence of β -repeating structures typical of amyloid fibrils. [196]

The choice of experimental conditions was based on the known aggregation properties of wild-type MNEI. Previous studies have shown that at pH 2.5 and in the presence of NaCl (≥ 150 mM), MNEI is able to form fibrils, thanks to the combination of residual protonation and electrostatic screening that facilitates the alignment of β -strands [195]; at almost neutral pH (≈ 6.8), even at the same protein concentration, the protein does not form fibrils but can give rise to amorphous aggregates under conditions of severe thermal stress, typically at temperatures equal to or higher than its T_m . [79,180] These two conditions therefore represent two structurally distinct scenarios: one that favours ordered (fibrillar) self-association and one that represents a state of non-cooperative destabilisation (amorphous aggregation).

Based on this, the three circularly permuted mutants were evaluated under conditions that maximise sensitivity to the two different aggregation regimes. In particular, each mutant was prepared at pH 2.5 with 150 mM NaCl, at a protein concentration of 2.5 mg/mL, *i.e.*, under conditions that promote fibril formation in the wild type, and was incubated at 55 °C and 50 °C, two temperatures selected on the basis of the thermal stability profiles

obtained by CD, in order to explore a range of conditions close to or below their respective T_{ms}.

At 55 °C, a condition that corresponds to an estimated unfolding percentage of between 10 and 25% for the mutants, the aggregation kinetics show clearly mutant-dependent behaviour. MNEI, which is completely folded at this temperature, exhibits minimal ThT fluorescence with no significant increase over time, suggesting a very low propensity to aggregate. (Figure 34)

In contrast, all permuted mutants show a marked increase in fluorescence, reaching a plateau after approximately 24 hours. The most pronounced response is observed for Perm2, which quickly reaches a higher fluorescence level than both Perm1 and Perm3. Perm1 shows the slowest kinetics and the lowest final fluorescence, while Perm3, despite showing rapid growth in the initial phase, reaches a lower plateau than Perm2. The order of aggregation propensity is therefore Perm2, then Perm3 and Perm1. Analysis of the lag phase further confirms this trend: Perm1 shows an initial limited variation in signal, similar to that of MNEI, while Perm2 and Perm3 already reach over 40-50% of the plateau signal within the first 2 hours. This difference suggests that circular permutation influences not only thermal stability but also the nucleation-dependent mechanisms underlying aggregation under acidic conditions. (Figure 34)

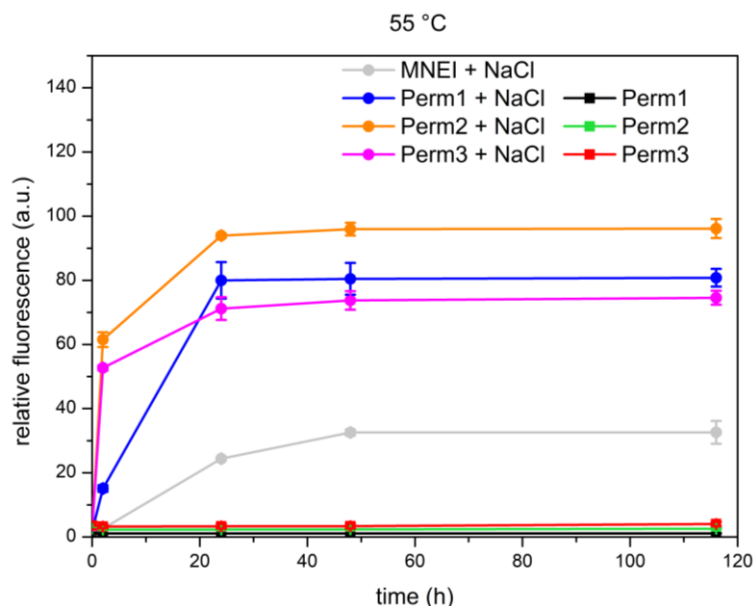


Figure 34. ThT fluorescence values were rescaled by dividing the absolute data by a constant factor (6000) for the sole purpose of facilitating comparison between graphs. Scaling does not alter the kinetic trend or the interpretation of the results.

To determine whether the differences observed were mainly due to the different unfolded fraction present at 55 °C, the mutants were also incubated at 50 °C, a temperature well below ~ 10 °C compared to T_m , at which the percentage of denaturation is reduced, and the aggregation kinetics should be slowed down. As expected, all curves show a reduction in the initial aggregation rate, and a lower plateau compared to the values recorded at 55 °C. (Figure 35)

However, the relative behaviour changes from highly divergent to more homogeneous: Perm2 and Perm3 still reach high levels of ThT fluorescence after approximately 24 hours, while Perm1 shows a more limited increase. Under these conditions, Perm2 and Perm3 maintain very similar kinetics in the early stages, while Perm1 is significantly slowed down, consistent with the greater structural stability identified in the CD analyses. (Figure 35)

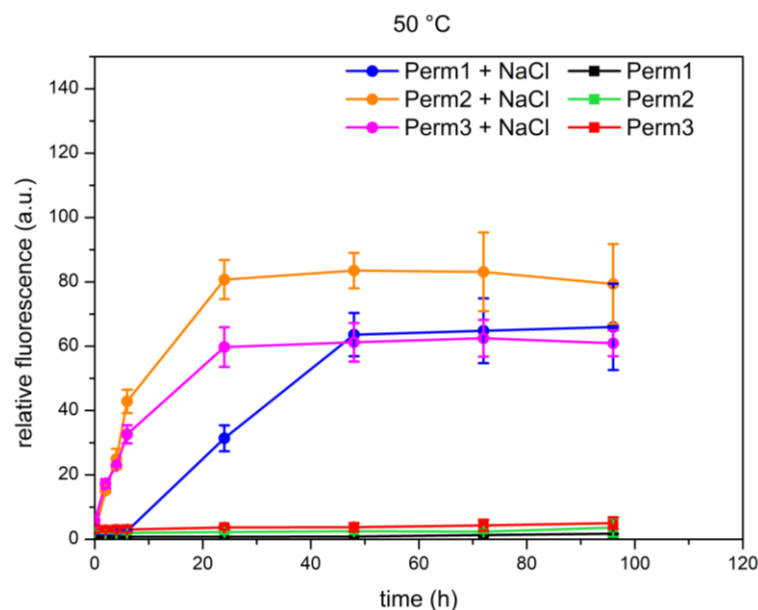


Figure 35. ThT fluorescence values were rescaled by dividing the absolute data by a constant factor (6000) for the sole purpose of facilitating comparison between graphs. Scaling does not alter the kinetic trend or the interpretation of the results.

Lastly, the results showed a clear kinetic difference between the mutants evaluated with NaCl and without NaCl. In fact, at both temperatures considered, in the absence of NaCl, no fluorescence was detected for any of the proteins examined, while, in the presence of NaCl, for each protein, ThT fluorescence peaked after 2 days of incubation (48 hours).

4.2.3.4 SEC and TEM analysis in aggregation conditions

The morphology of the aggregates formed by the three circularly permuted variants was first investigated by transmission electron microscopy (TEM), using samples collected at the plateau of the ThT-monitored aggregation kinetics. Analyses were performed under acidic conditions (pH 2.5, 150 mM NaCl) at 50 °C and 55 °C, temperatures at which clear differences were observed in aggregation behaviour.

At 50 °C, the three variants display markedly distinct morphologies (Figure 36). Perm1 predominantly forms thin, well-defined filamentous structures, which are sparsely distributed and weakly interconnected. These filaments appear relatively homogeneous and structurally preserved, suggesting an

initial propensity toward ordered fibrillar assembly even at this lower temperature. In contrast, both Perm2 and Perm3 are mainly characterised by irregular, amorphous aggregates, consisting of disordered protein material. Only occasional, very short fibrillar elements can be detected, indicating limited fibril nucleation and poor progression toward mature fibrillar structures. (Figure 36)

These differences become more pronounced at 55 °C (Figure 37). Under these conditions, Perm1 gives rise to extended networks of long, thin and frequently paired fibrils, organised into compact and interconnected lattices, indicative of efficient fibril maturation. Perm2 exhibits a mixed morphology, in which short fibrillar elements coexist with amorphous aggregates, suggesting partial and heterogeneous progression toward ordered assembly. In contrast, Perm3 predominantly forms dense, disordered aggregates, with only rare and poorly developed fibrillar structures visible. (Figure 37).

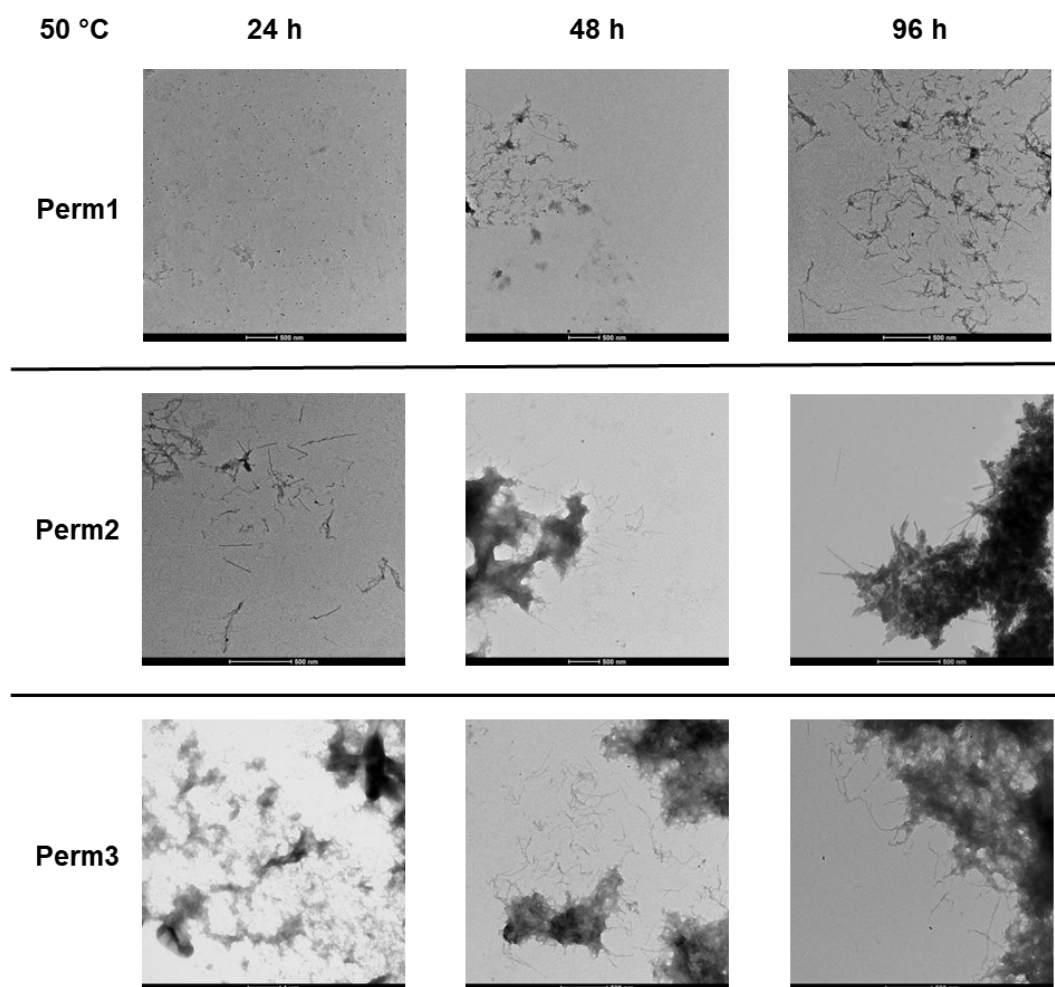


Figure 36. Transmission electron microscopy images of aggregates formed by Perm1, Perm2 and Perm3 at 50 °C under acidic conditions (pH 2.5, 150 mM NaCl), collected after 24, 48, 96 hours of ThT-monitored aggregation kinetics. Perm1 predominantly displays sparse, thin and well-defined filamentous structures, largely isolated and weakly interconnected. In contrast, Perm2 and Perm3 are mainly characterised by irregular, amorphous aggregates with poorly defined morphology, from which only occasional short fibrillar elements emerge. The overall appearance indicates limited fibril maturation at this temperature, with a clear prevalence of disordered assemblies for Perm2 and Perm3. Scale bars as indicated.

Overall, TEM analysis reveals a temperature-dependent divergence in aggregation pathways, with increasing temperature promoting fibril maturation in Perm1, while amplifying amorphous aggregation in Perm2 and Perm3.

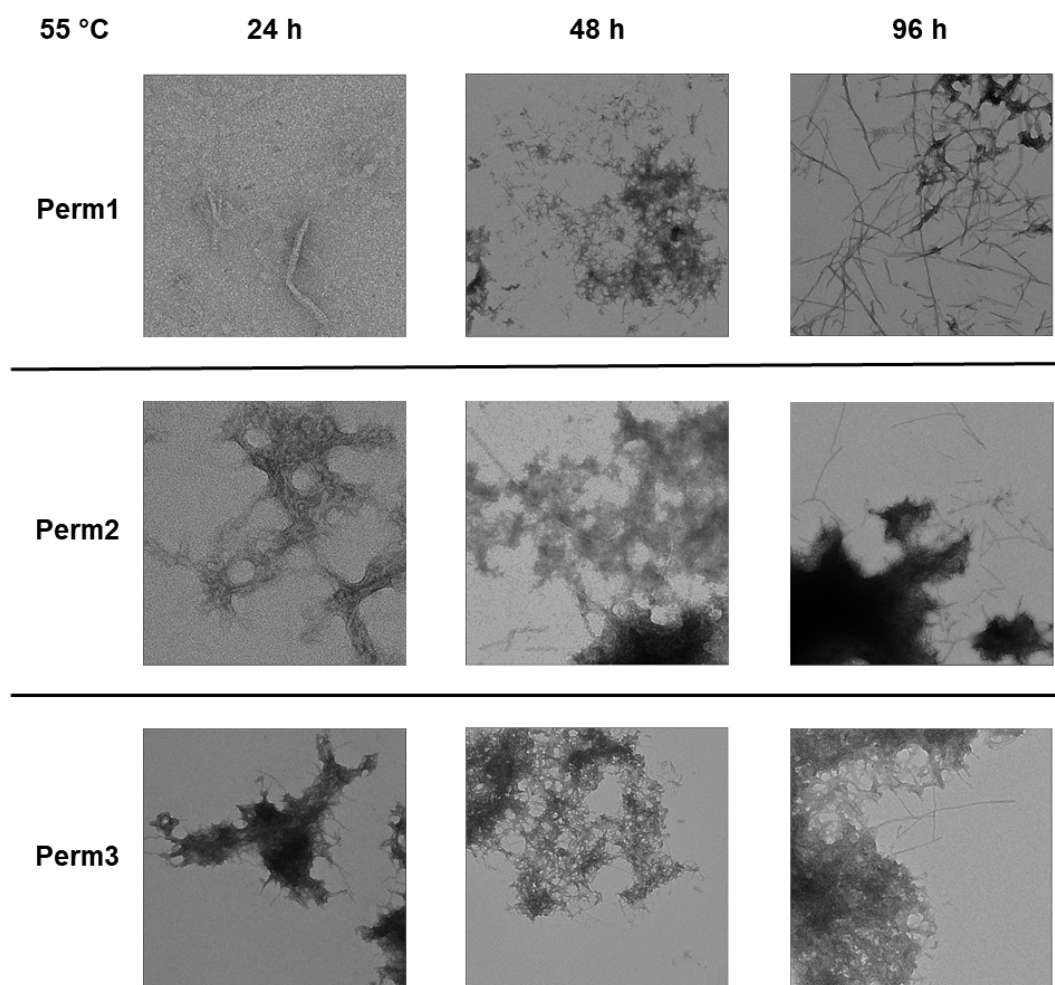


Figure 37. Transmission electron microscopy images of aggregates formed by Perm1, Perm2 and Perm3 at 55 °C under acidic conditions (pH 2.5, 150 mM NaCl), acquired after 24, 48, 96 hours. At this higher temperature, Perm1 forms extended networks of long, thin and frequently paired fibrils, organised into compact and interconnected lattices. Perm2 exhibits a mixed morphology, with short fibrillar elements coexisting with amorphous aggregates, suggesting partial progression toward ordered assembly. In contrast, Perm3 predominantly generates dense, disordered aggregates, with only rare and poorly developed fibrillar structures visible. These images highlight a temperature-dependent divergence in aggregation pathways, with increasing temperature favouring fibril maturation in Perm1 but amplifying amorphous aggregation in Perm2 and Perm3.

Based on these morphological observations, it was then checked if the different aggregation results were associated with distinct oligomeric behaviours in solution. To this end, the three variants were analysed by size exclusion chromatography (SEC) during incubation at 50°C and 55°C (Figure 38). At 50 °C, Perm1 remains essentially monomeric for at least 24

hours, with no detectable accumulation of soluble oligomeric species. A progressive disappearance of the monomeric peak is observed only at 55 °C, a condition under which, consistently with TEM observations, the protein appears to be fully converted into insoluble fibrillar material. Notably, no stable intermediate oligomers accumulate in solution, suggesting that Perm1 follows a relatively direct aggregation pathway, in which monomers convert into fibrils without detectable increase of soluble higher-order species. In contrast, Perm2 and Perm3 exhibit a profoundly different behaviour. (Figure 38)

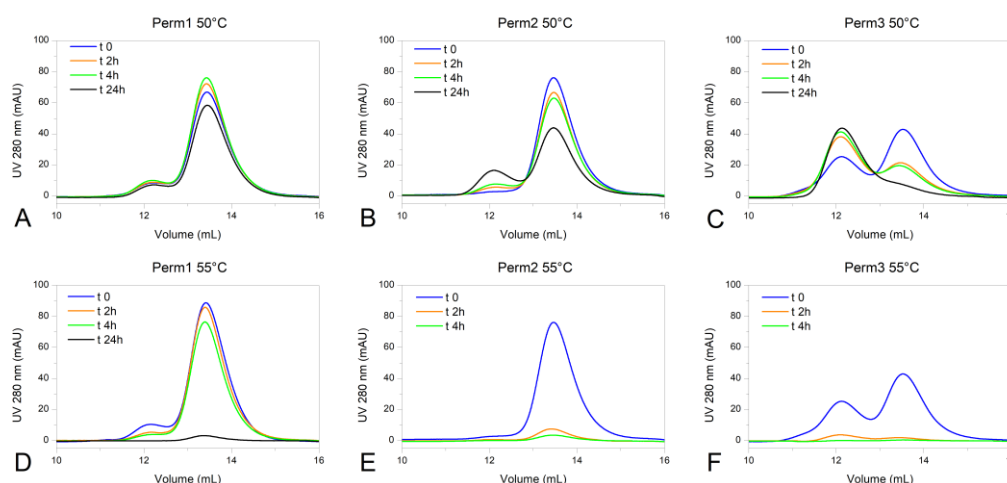


Figure 38. Size exclusion chromatography profiles of the circularly permuted variants Perm1, Perm2 and Perm3 recorded on a Superdex 75 10/300 column in 50 mM sodium acetate buffer, pH 5.5, supplemented with 150 mM NaCl. Samples were incubated at 50 °C (A-C) or 55 °C (D-F) and analysed at different time points (t0, 2 h, 4 h and 24 h) to monitor temperature-dependent changes in elution behaviour. The comparison of chromatograms highlights distinct stability profiles among the permutants, with temperature- and time-dependent alterations in peak shape and intensity, indicative of differential susceptibility to thermal perturbation and aggregation.

At 50 °C, both variants show a progressive reduction of the monomeric peak accompanied by the appearance and growth of a dimeric population, with evidently faster kinetics for Perm3. While Perm2 retains monomer as the predominant soluble species throughout the incubation, Perm3 already

displays a predominance of dimers after 2 hours, indicating a strong intrinsic tendency toward early self-association. (Figure 38)

At 55 °C, both proteins rapidly precipitate, leading to a drastic reduction in soluble species; however, Perm3 transiently retains a detectable dimeric species in the early incubation phase, before complete precipitation. (Figure 38)

To further validate the SEC results and directly visualise the oligomeric states under non-denaturing conditions, the samples incubated at 50 °C were analysed by native PAGE (Figure 39). In line with SEC profiles, Perm1 remains predominantly monomeric over time, showing no significant change after 24 hours of incubation. Perm2 displays a partial reduction of the monomeric band accompanied by the appearance of a dimeric species, while Perm3 exhibits a pronounced shift from monomer to dimer within the same time frame, confirming its strong propensity for early oligomerisation.

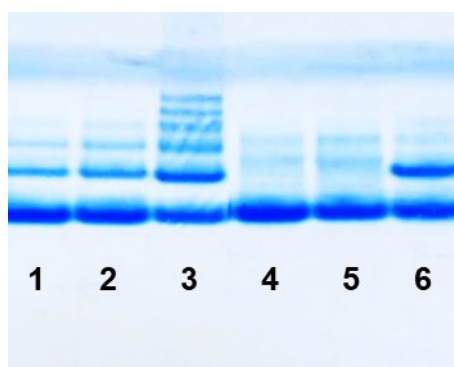


Figure 39. Native PAGE analysis of Perm1, Perm2 and Perm3 incubated at 50 °C. Lanes 1-3 correspond to samples collected at the beginning of the incubation, while lanes 4-6 correspond to samples collected after 24 h. Differences in band migration and relative intensity indicate time-dependent changes in the oligomeric state of the three variants.

Taken together, the integration of TEM, SEC and native PAGE analyses indicates that the circular permutation strategy modulates not only the overall conformational stability of the variants, but also their early self-association behaviour and their ability to mature into ordered fibrillar structures. Perm1, the most stable variant and the least prone to forming soluble oligomers, favours a progressive and orderly fibrillar growth. Perm2

displays an intermediate behaviour, in which partial oligomerisation can coexist with limited fibril formation. Finally, Perm3 shows the highest conformational instability and the strongest tendency toward early dimerization, but these oligomers fail to evolve into ordered fibrils, instead leading predominantly to amorphous and highly disordered aggregates.

4.2.4 Crystal structure of Perm mutants

The structural characterisation of the three circularly permuted mutants was completed using X-ray crystallography, which allowed the three-dimensional structure of Perm1, Perm2 and Perm3 to be determined with resolutions between 1.15 and 1.60 Å (PDB IDs: 9HIJ, 9HKG, 9HIK, respectively). (Table 18) The crystals were obtained from SEC-purified samples, after concentration up to 6 mg/mL, selecting the predominant form in solution for each mutant: the natural dimeric form in the case of Perm3, and the monomeric form in the case of Perm1 and Perm2. (Figure 40) The optimal crystallographic conditions were significantly different between the three mutants, reflecting the different conformational stability and flexibility introduced by circular permutation. Perm1 crystallised at pH 8.5 by vapour diffusion in sitting drop, in the crystallization condition corresponding to well A12 of the Morpheus I screen, characterized by the presence of a complex precipitant consisting of PEG1000, PEG3350 and MPD and high concentrations of divalent ions (Mg^{2+} and Ca^{2+}), conditions typically associated with proteins that require a strongly stabilising environment to crystallise. Crystals appeared after approximately four months of incubation. Perm2 generated crystals after a few months via hanging drop at pH 5.5, in the crystallization condition corresponding to well A7 of the Cryos screen, consisting of PEG4000 and 2-propanol as precipitants and glycerol as an additive; Perm3 crystallised under similar conditions, but at pH 5.6 and a few hours after incubation in the crystallization condition corresponding to well F11 of the JCSG Core III screen, with PEG4000 and ammonium acetate, suggesting greater structural compatibility of the less

stable variants with “softer” precipitants and in a weakly acidic environment. Crystals appeared after approximately one day of incubation. (Figure 40)

Despite differences in crystallisation conditions, the three structures share a common feature of great importance: they all exhibit a quaternary organisation based on domain swapping, in which structured regions of one protomer exchange with their counterparts in the other, generating a highly symmetrical intertwined dimer. In all cases, the single protomer retains the α/β fold typical of MNEI, with C α RMSD values relative to the native structure (PDB 8Q0S) of 0.6 Å for Perm3, 0.4 Å for Perm2 and 0.5 Å for Perm1, confirming that circular permutation does not alter the basic architecture of the protein. (Figures 41-42)

However, the way in which domain swapping occurs differs between mutants. In Perm2 and Perm3, the dimer shows an almost identical geometry, characterised by a compact interface stabilised by a well-defined network of ionic contacts. The most relevant interaction involves residues Asp24 and Arg38, which form a reciprocal salt bridge, and the additional contact between Arg38 and Glu9, which completes a triad of interactions crucial for the stability of the interface. This pattern of bonds is repeated symmetrically in the two protomers, suggesting a mode of association favoured by the repositioning of secondary regions imposed by permutation. (Figure 41)

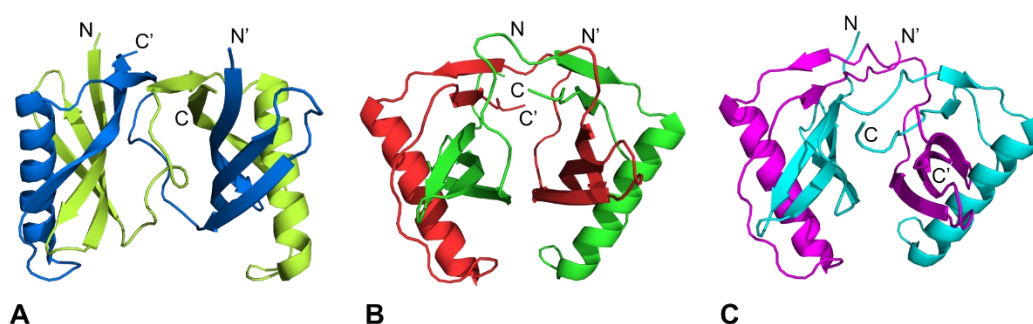


Figure 40. Crystal structures of the three circularly permuted variants Perm1, Perm2, and Perm3. The structure of Perm1 (A) was determined with a resolution of 1.60 Å, that of Perm2 (B) with a resolution of 1.26 Å, and that of Perm3 (C) with a resolution of 1.15 Å. The symmetry of the structure allows us to highlight the differences in the relative

orientation of the protomers within the crystal grid, despite the conservation of the overall fold.

The situation is qualitatively different in Perm1. Although the structure also shows domain swapping, the relative orientation of the protomers is different and the interface is more open, partially involving the new linker introduced to connect the native ends of the protein. The interprotomer contacts are more heterogeneous: the Pro44 residue interacts with Arg22; the first residue of the linker, Gly47, forms a double hydrogen bond with Arg38; finally, a salt bridge between Glu9 and Lys95 completes the interface. (Figure 42)

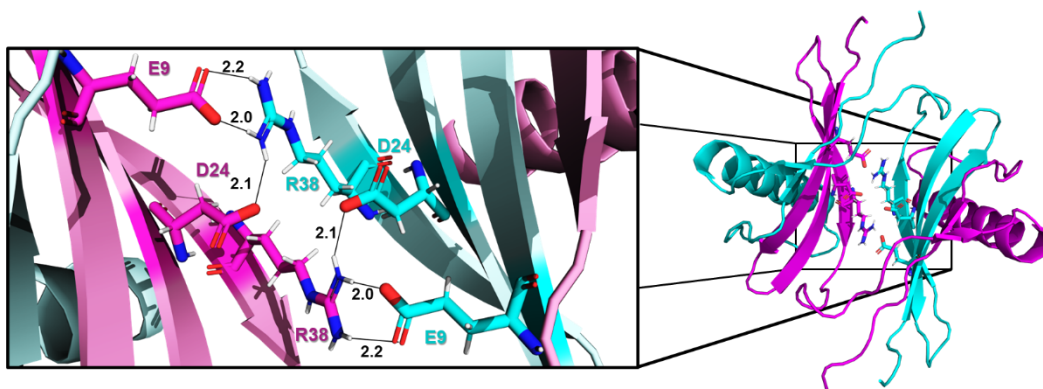


Figure 41. The rectangle shows details of the interprotomer contacts in the interface of the Perm3 (similar to Perm2) crystal structure (right). In bold, the distances and ionic bonds involved in the salt bridges between the side chains of Glu9 and Asp24 with Arg38 of a magenta protomer and vice versa in the cyan protomer.

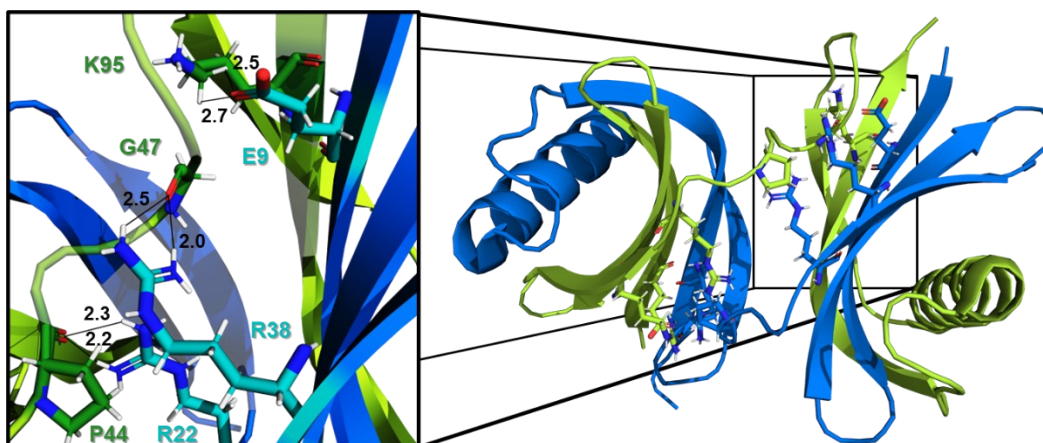


Figure 42. The square shows details of the interprotomeric contacts in the interface of the Perm1 crystal structure (right). The most open interface directly involves the linker introduced to join the native ends. In bold, the distances and ionic bonds involved in the salt bridges between the side chains of Glu9 and Lys95, Arg38 and Gly47, Arg22 and Pro44 of a blue protomer and vice versa in the green protomer.

This larger and less symmetrical configuration is consistent with the increased length of the linker introduced in Perm1, which allows alternative relative orientations of the protomers within the domain-swapped assembly. Importantly, the observation that all three mutants ultimately crystallise as domain-swapped dimers suggests that this quaternary arrangement is thermodynamically favoured under the crystallisation conditions, despite the markedly different oligomeric behaviour observed in solution. In this framework, the prevalence of monomeric species for Perm1 and Perm2 at short incubation times likely reflects kinetic barriers to dimer formation rather than an intrinsic thermodynamic disfavouring of the domain-swapped state. Accordingly, while Perm2 and Perm3 converge toward a highly similar and compact dimeric geometry stabilised by conserved salt bridges, Perm1 adopts a distinct and more open quaternary organisation, reflecting the combined effects of linker length, conformational flexibility and alternative pathways for accessing the domain-swapped minimum.

	Perm1 (PDB: 9HIJ)	Perm2 (PDB: 9HKG)	Perm3 (PDB 9HIK)
Wavelength (Å)	0.999998	0.999990	0.999990
Beamline	SLS, X06SA	SLS, X06SA	SLS, X06SA
Resolution range	40.08 - 1.60 (1.67 - 1.60)	39.1 - 1.264 (1.31 - 1.264)	30.41 - 1.15 (1.24 - 1.15)
Space group	I 1 2 1	P 1 2 1	P 1 2 1
Unit cell a, b, c (Å)			

α, β, γ (°)	45.646, 31.599, 127.006 90, 100.203, 90	39.464, 32.518, 40.824 90, 97.837, 90	39.451 32.446, 41.484 90, 97.583, 90
Total reflections	80831 (7420)	150411 (4001)	210292 (10657)
Unique reflections	23048 (2809)	23742 (1035)	34452 (5551)
Multiplicity	3.4 (3.3)	6.3 (3.8)	6.1 (4.4)
Completeness (%)	96.29 (94.55)	85.59 (37.35)	92.61 (75.33)
Mean I/sigma(I)	12.68 (1.49)	14.46 (1.05)	15.22 (2.17)
Wilson B-factor (Å²)	26.15	21.89	13.74
R-merge	0.0539 (0.7828)	0.06587 (0.8873)	0.0624 (0.5609)
R-meas	0.0642 (0.9433)	0.07209 (1.024)	0.0683 (0.6323)
R-pim	0.0344 (0.5167)	0.02871 (0.4936)	0.0273 (0.2841)
CC1/2	0.998 (0.627)	0.993 (0.68)	0.996 (0.900)
CC*	0.999 (0.878)	0.998 (0.9)	0.999 (0.973)
Reflections used in refinement	23048 (2809)	23710 (1035)	34452 (5551)
Reflections used for R-free	1177 (159)	457 (19)	644 (86)
R-work	0.1704 (0.2772)	0.1594 (0.2663)	0.1589 (0.2368)
R-free	0.1906 (0.3073)	0.1911 (0.3142)	0.1876 (0.2285)
CC(work)	0.966 (0.861)	0.971 (0.468)	0.970 (0.939)

CC(free)	0.971 (0.831)	0.968 (-0.005)	0.963 (0.714)
Average B-factor (Å²)	35.22	35.56	21.26
Number of non-H atoms	1700	901	981
macromolecules	1585	808	822
ligands	32	14	0
solvent	83	85	159
Protein residues	193	98	97
RMS(bonds, Å)	0.005	0.011	0.004
RMS(angles, °)	0.92	0.89	0.67
Ramachandran favored (%)	97.35	96.88	96.84
Ramachandran allowed (%)	2.65	3.12	3.16
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.00	1.18	2.27

Table 18. Data collection and refinement statistics.

4.2.4.1 Topological comparison of PERM domain-swapped dimers with cystatin family members

The domain-swapped architectures observed in the crystal structures of the circularly permuted mutants of the sweet protein MNEI can be further contextualised through comparison with other members of the cystatin family known to undergo domain swapping. Among these, Stefin B (PDB ID: 2OCT) and the engineered construct L1MN (PDB ID: 5YCU) in their domain-swapped conformation were taken as representative examples. [95] At the level of the individual protomer, the PERM mutants retain the

canonical α/β cystatin fold, allowing for significant structural superposition with both systems despite differences in sequence and chain connectivity, and indicating that circular permutation does not alter the fundamental architectural structure of the protein.

The overlap of Perm1 with L1MN reveals good local overlap of the exchanged structural elements, with a low RMSD value (~ 0.4 Å), indicating that the same topological unit is involved in the exchange process. (Figure 43) However, despite this close structural correspondence, the relative orientation of the protomers within the domain-swapped dimer differs substantially between the two systems. In particular, Perm1 adopts a more compact quaternary organization, with tightly grouped protomers, while L1MN exhibits a more extended geometry, characterized by greater spatial separation of the exchanged domains. This observation indicates that circular permutation preserves the structural identity of the exchanged unit while modulating the overall topology of the domain-exchange assembly. Similar behavior is observed for Perm3, which also shows excellent overlap with L1MN at the level of the exchanged regions, but differs in the overall orientation of the interprotomers, in line with the distinct chain connectivity and linker properties of the permuted variants. (Figure 43)

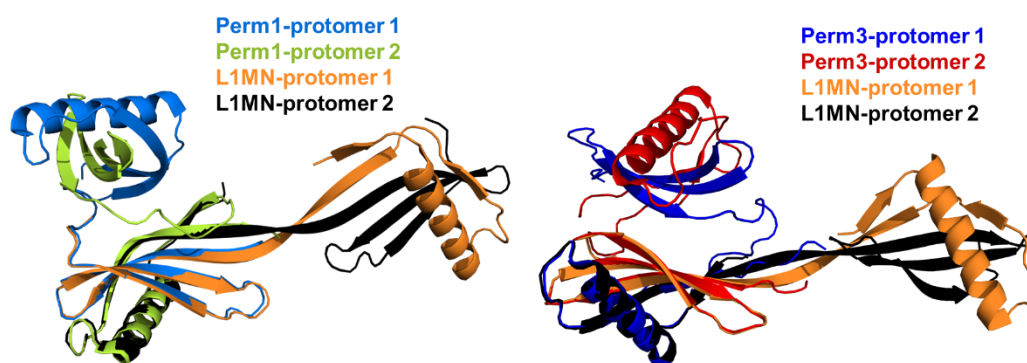


Figure 43. Structural superposition of the Perm1 dimer with domain swapping (PDB ID: 9HIJ) with the engineered cystatin construct L1MN (PDB ID: 5YCU). A protomer of the Perm1 domain-swapped dimer (protomere 1 in blue, protomere 2 in green) is superimposed on a protomer of L1MN (protomere 1 in orange, protomere 2 in black). The exchanged structural elements show excellent local overlap ($\text{RMSD} \approx 0.4$ Å), indicating the involvement of the same topological unit in the swapping process.

However, comparison with stefin B further highlights the distinction between the conservation of the exchange mechanism and the variability of the resulting quaternary architecture. (Figure 44) Due to the limited sequential similarity between the Perm1 and Perm3 mutants and Stefin B, the overall structural overlap is limited. Nevertheless, it is still possible to detect a partial alignment of the exchanged structural elements. This local correspondence indicates that domain swapping in cystatins is based on conserved architectural motifs, even when incorporated into markedly different protomeric orientations and dimeric geometries. This demonstrates that domain swapping represents a general and thermodynamically accessible organisational principle within the cystatin fold, while the specific topology of the resulting dimer is strongly influenced by chain connectivity and linker properties. (Figure 44)

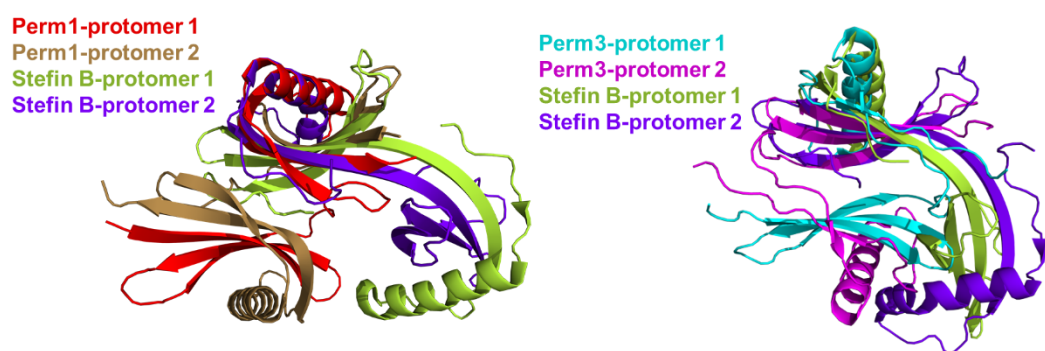


Figure 44. Structural comparison between the Perm1 dimer with domain exchange (left, PDB ID: 9HIJ) and the Perm3 dimer with domain exchange (right, PDB ID: 9HIK) and the stefin B dimer with domain exchange (PDB ID: 2OCT). The two Perm1 protomers (red and sand) and the two Perm3 protomers (cyan and magenta) are shown superimposed on the two stefin B protomers (lemon and purple). Due to the limited sequence similarity, the overall structural alignment in both cases is poor; however, it is still possible to detect a partial overlap of the exchanged structural elements.

4.2.5 Molecular dynamics analysis of circularly permuted mutants

4.2.5.1 RMSD analysis

To study stability under different thermal conditions and identify potential “hot spots” for local unfolding events, molecular dynamics simulations were performed on the three circularly permuted proteins at four different temperatures (300, 350, 400, and 450 K) under acidic conditions. The RMSD trajectories (Figure 45) provide an overview of the conformational drift experienced by each mutant along 200-ns trajectories. At 300 and 350 K, all variants maintain RMSD values between ~2 and 3 Å after the initial equilibration phase, indicating that the native topology of the three permuted constructs is sufficiently stable under mild thermal stress. The trajectories remain compact, without showing evidence of large-scale unfolding events. This suggests that the circular permutation, despite reorganizing the chain connectivity, does not compromise the global fold at physiological or moderately elevated temperatures. (Figure 45)

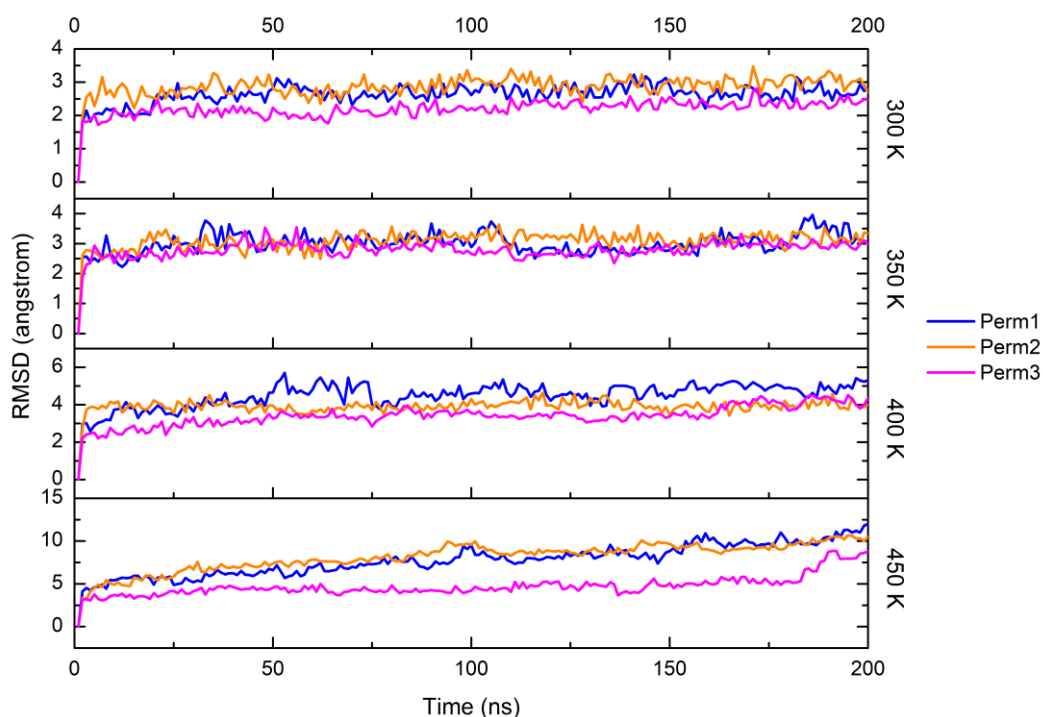


Figure 45. Backbone RMSD profiles of the three circularly permuted variants Perm1, Perm2 and Perm3 calculated at pH 2.5 over 200 ns molecular dynamics simulations performed at increasing temperatures (300, 350, 400 and 450 K, from top to bottom). At 300 K, all variants display low and stable RMSD values, indicating preservation of the native fold. A moderate increase in fluctuations is observed at 350 K, with Perm1 showing slightly larger deviations than Perm2 and Perm3. At 400 K, the divergence between the variants becomes more pronounced, as Perm1 undergoes a marked increase in RMSD, while Perm2 and Perm3 retain comparatively lower values. At 450 K, all systems experience substantial structural destabilisation; however, Perm3 consistently exhibits lower RMSD values than Perm1 and Perm2, suggesting a higher resistance to extreme thermal perturbation.

At 400 K, a divergence emerges among the mutants. Perm1 exhibits a pronounced increase in RMSD fluctuations ($\sim 4\text{-}6\text{ \AA}$), while Perm2 and Perm3 maintain more moderate drifts ($\sim 3\text{-}4\text{ \AA}$). This differential response illustrates a higher thermal sensitivity of Perm1, consistent with its longer and more flexible linker. Nevertheless, at 450 K, all proteins undergo substantial conformational rearrangements, although to varying degrees. Perm1 shows the largest RMSD elevation (up to $\sim 10\text{ \AA}$), followed by Perm2, while Perm3 remains comparatively more compact ($\sim 5\text{-}7\text{ \AA}$). (Figure 45) This result might seem counterintuitive, given the weaker experimental stability of Perm3, but high-temperature MD highlights intrinsic backbone rigidity not necessarily the same stability landscape probed by thermal denaturation experiments. This discrepancy underscores the need to interpret MD-derived instabilities as reflecting local flexibility rather than macroscopic folding transitions. Overall, the RMSD profiles indicate that the three mutants remain structurally robust up to 350 K, begin to differentiate at 400 K, and diverge strongly at 450 K, with Perm1 being the variant most susceptible to thermal destabilization *in silico*.

4.2.5.2 RMSF analysis

The Root Mean Square Fluctuation (RMSF) profiles (Figure 46) reveal the temperature-dependent dynamics of individual residues, enabling the identification of mobility hotspots potentially linked to unfolding nucleation.

Across all temperatures, the regions showing the highest intrinsic flexibility include: the β 4- β 5 loop, the α H- β 2 connecting loop, the extended C-terminal region incorporating the engineered linker (β 5- α H segment). These domains correspond to structurally exposed and functionally relevant regions of MNEI, and their flexibility is expected. However, the degree to which each mutant samples these fluctuations is markedly different. The engineered linker is the most mobile region in all mutants. Perm1, with a five-residue connector, exhibits RMSF values significantly higher than those of Perm2 and Perm3 at all temperatures. The broader conformational freedom provided by the longer linker likely contributes to Perm1's increased thermal sensitivity observed in the RMSD analysis. One could challenge the assumption that increased linker flexibility directly destabilizes the overall fold; however, the consistency between RMSD, RMSF and DSSP data supports this interpretation. All mutants display high fluctuations in the region corresponding to the β 4- β 5 loop, which is known to be a critical hinge region in MNEI and plays a role in early unfolding during aggregation. As temperature increases, Perm1 shows a dramatically larger amplitude of motion compared to the other two proteins. This behaviour suggests a shift in the energy landscape that might facilitate partial unfolding or domain rearrangements. (Figure 46)

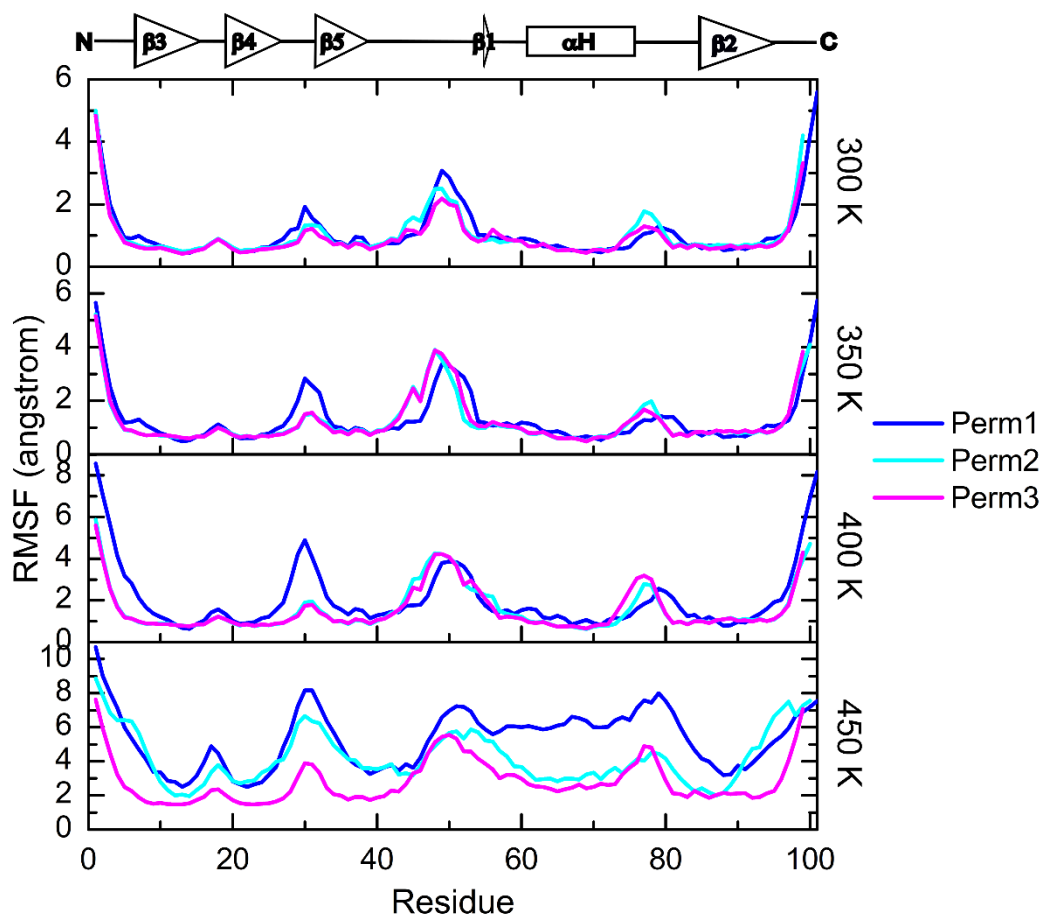


Figure 46. Residue-wise root mean square fluctuation (RMSF) profiles of the three circularly permuted variants Perm1, Perm2 and Perm3 computed from 200 ns molecular dynamics simulations performed at increasing temperatures (300, 350, 400 and 450 K, from top to bottom). RMSF values are reported for Perm1 (blue), Perm2 (cyan) and Perm3 (magenta) as a function of the residue index. The schematic representation of the secondary structure elements is shown above the plots, allowing direct mapping of local flexibility onto β -strands and α -helical regions.

To increase the robustness of MD-derived observations, RMSF values were averaged across triplicate 200-ns simulations at each temperature.

Pairwise t-tests and Cohen's d effect sizes (Tables 19-20-21) confirmed statistically significant increases in mobility with rising temperature for all mutants, with the most dramatic effect observed in Perm1.

Perm1: $1.13 \pm 0.81 \text{ \AA} \rightarrow 5.32 \pm 1.90 \text{ \AA}$ from 300 K to 450 K $\equiv$ Very large effect size ($d = 2.85$)

Perm2: $1.00 \pm 0.64 \text{ \AA} \rightarrow 4.21 \pm 1.83 \text{ \AA}$ $\equiv$ Large effect size ($d = 2.40$)

Perm3: $0.91 \pm 0.51 \text{ \AA} \rightarrow 2.82 \pm 1.41 \text{ \AA} \equiv$ Moderate-to-large effect size ($d = 1.87$)

Inter-protein comparisons at 450 K further emphasize Perm1's enhanced dynamic response. These results confirm a temperature-dependent divergence in intrinsic flexibility, with Perm1 being most sensitive to thermal perturbation, reflecting a behaviour reminiscent of native MNEI under aggregating conditions.

Protein	Metrics	Temperature (K)	Replica	Mean (Å)	StdDev (Å)	Mean Pooled (Å)	StdDev Pooled (Å)
Perm1	RMSF	300	Eq_1	1.131	0.814	1.128	0.809
			Eq_2	1.084	0.806		
			E1_3	1.170	0.807		
		350	Eq_1	1.277	0.847	1.331	0.902
			Eq_2	1.396	0.952		
			E1_3	1.321	0.905		
		400	Eq_1	1.831	1.251	2.064	1.453
			Eq_2	1.886	1.319		
			E1_3	2.476	1.790		
		450	Eq_1	3.495	1.756	5.315	1.905
			Eq_2	5.754	1.979		
			E1_3	6.697	1.980		
Perm2	RMSF	300	Eq_1	0.879	0.524	1.003	0.637
			Eq_2	1.030	0.710		
			E1_3	1.100	0.676		

		350	Eq_1	1.250	0.855	1.188	0.774
			Eq_2	1.218	0.806		
			E1_3	1.096	0.660		
		400	Eq_1	1.390	0.929	1.535	1.048
			Eq_2	1.478	0.982		
			E1_3	1.737	1.232		
		450	Eq_1	2.893	1.714	4.215	1.829
			Eq_2	6.193	1.951		
			E1_3	3.559	1.821		
Perm3	RMSF	300	Eq_1	0.973	0.559	0.914	0.506
			Eq_2	0.919	0.488		
			E1_3	0.850	0.471		
		350	Eq_1	1.228	0.772	1.198	0.786
			Eq_2	1.087	0.702		
			E1_3	1.279	0.883		
		400	Eq_1	1.255	0.768	1.529	1.080
			Eq_2	1.450	1.003		
			E1_3	1.884	1.468		
		450	Eq_1	4.309	1.526	2.817	1.415
			Eq_2	2.246	1.448		
			E1_3	1.897	1.271		

Table 19. Raw and pooled RMSF data for Perm1, Perm2, and Perm3 at different temperatures. The table reports RMSF values (Å) for Perm1, Perm2, and Perm3 at 300, 350, 400, and 450 K, obtained from three independent 200-ns MD simulations (2000 frames each). For each replicate (Eq_1, Eq_2, Eq_3), mean RMSF and standard deviation

are reported. Pooled values include the average of replicate means (Mean Pooled) and the pooled standard deviation (StdDev Pooled), calculated from the individual replicate standard deviations. These RMSF data quantify local residue flexibility and form the basis for the statistical analyses reported in Tables 20 and 21.

Protein	Metrics	Comparison (K)	Mean Temp. 1 (Å)	Mean Temp. 2 (Å)	t-value	Statistical Significance	p-value	Cohen's d	Effect
Perm1	RMSF	300 vs. 350	1.128	1.331	-12.963	**	<0.001	-0.24	Small effect
		300 vs. 400	1.128	2.064	-43.573	***	<0.001	-0.80	Medium effect
		300 vs. 450	1.128	5.315	-156.707	***	<0.001	-2.85	Very large effect
		350 vs. 400	1.331	2.064	-33.198	***	<0.001	-0.61	Medium effect
		350 vs. 450	1.331	5.315	-146.440	***	<0.001	-2.65	Very large effect
		400 vs. 450	2.064	5.315	-105.105	***	<0.001	-1.92	Large effect
Perm2	RMSF	300 vs. 350	1.003	1.188	-14.297	**	<0.001	-0.26	Small effect
		300 vs. 400	1.003	1.535	-33.589	***	<0.001	-0.62	Medium effect
		300 vs. 450	1.003	4.215	-128.095	***	<0.001	-2.40	Very large effect
		350 vs. 400	1.188	1.535	-20.622	***	<0.001	-0.38	Small effect
		350 vs. 450	1.188	4.215	-118.095	***	<0.001	-2.18	Very large effect
		400 vs. 450	1.535	4.215	-98.120	***	<0.001	-1.80	Large effect
Perm3	RMSF	300 vs. 350	0.914	1.198	-23.527	***	<0.001	-0.43	Small effect
		300 vs. 400	0.914	1.529	-39.995	***	<0.001	-0.75	Medium effect
		300 vs. 450	0.914	2.817	-98.120	***	<0.001	-1.87	Large effect
		350 vs. 400	1.198	1.529	-19.247	**	<0.001	-0.35	Small effect

		350 vs. 450	1.198	2.817	-77.511	***	<0.001	-1.47	Large effect
		400 vs. 450	1.529	2.817	-56.041	***	<0.001	-1.02	Large effect

Table 20. Statistical analysis of RMSF differences across temperatures for Perm1, Perm2, and Perm3. The table reports pairwise two-sample t-tests comparing RMSF values (Å) of Perm1, Perm2, and Perm3 across 300, 350, 400, and 450 K, based on pooled data from triplicate 200-ns MD simulations (6000 frames per condition). For each comparison, t-values, degrees of freedom, 95% confidence intervals of the mean difference (Δ = Mean Temp. 2 - Mean Temp. 1), and statistical significance are reported. All comparisons are highly significant ($p < 0.001$, Bonferroni-corrected). Effect sizes were calculated as Cohen's d using pooled standard deviations (Supplementary Table 19); both signed d and |d| are reported, with effect magnitude classified as small (<0.5), medium (0.5-1.0), large (>1.0), or very large (>2.0). The Effect column summarizes the biological relevance of the observed differences in flexibility.

Metrics	Comparison	Temperature (K)	Mean Temp. Prot. 1 (Å)	Mean Temp. Prot. 2 (Å)	t-value	Statistical significance	p-value	Cohen's d	Effect
RMSF	Perm1 vs. Perm2	300	1.1284	1.0030	9.4357	*	<0.001	0.17	ns
	Perm1 vs. Perm3		1.1284	0.9138	17.4234	**	<0.001	0.32	ns
	Perm2 vs. Perm3		1.0030	0.9138	8.4985	*	<0.001	0.16	ns
	Perm1 vs. Perm2	350	1.3311	1.1879	9.3380	*	<0.001	0.17	ns
	Perm1 vs. Perm3		1.3311	1.1976	8.6474	*	<0.001	0.16	ns

	Perm2 vs. Perm3		1.1879	1.1976	-0.6815	-	>0.05	0.01	ns
	Perm1 vs. Perm2	400	2.0641	1.5346	22.8921	***	<0.001	0.42	ns
	Perm1 vs. Perm3		2.0641	1.5294	22.8756	***	<0.001	0.42	ns
	Perm2 vs. Perm3		1.5346	1.5294	0.2677	-	>0.05	0.01	ns
	Perm1 vs. Perm2	450	5.3153	4.2148	32.2838	***	<0.001	0.59	Medium effect
	Perm1 vs. Perm3		5.3153	2.8170	81.5557	***	<0.001	1.49	Large effect
	Perm2 vs. Perm3		4.2148	2.8170	46.8321	***	<0.001	0.85	Medium effect

Table 21. Pairwise statistical comparison of RMSF between Perm1, Perm2, and Perm3 at fixed temperatures. The table reports pairwise two-sample t-tests comparing RMSF values (Å) of Perm1, Perm2, and Perm3 at 300, 350, 400, and 450 K, based on pooled data from triplicate 200-ns MD simulations (6000 frames per condition; pooled means and standard deviations from Table 19). For each comparison, t-values indicate the direction of the difference, statistical significance is reported using asterisk notation ($|t| \geq 3.29$, $p < 0.001$), and effect sizes are expressed as Cohen's d. Effect magnitude is classified as ns (not significant), small (<0.5), medium ($0.5-1.0$), large (≥ 1.0), or very large (>2.0). The Effect column summarizes biologically relevant differences in flexibility, highlighting the stronger thermal response of Perm1 relative to Perm2 and Perm3 at higher temperatures.

4.2.5.3 DSSP analysis

Secondary structure persistence was assessed through DSSP assignment along the trajectories (Figure 47). This analysis provides residue-level

insight into the structural elements most vulnerable to thermal stress and complements RMSF interpretation. At 300 K, all proteins retain a stable β -sheet core, and the α H is fully preserved. However, early signs of helix fraying at the C-terminal end of α H are detectable in Perm2 and Perm3 even at the lowest temperature, indicating a slightly weaker stabilization of this region in the mutants with shorter linkers. At 350 K and 400 K, the divergence becomes more pronounced. Perm1 maintains a stable helical core, but its extended linker begins to populate partially unstructured conformations. In contrast, Perm2 and Perm3 start losing α -helical secondary structure progressively from residue ~72 onward. This observation aligns with the RMSF peaks in the α H- β 2 loop, supporting the view that this region constitutes a thermally sensitive hinge. At 450 K the three proteins undergo substantial helix loss, but with different magnitudes: Perm1 loses approximately 50% of its helical content, consistent with its higher RMSD and RMSF at high temperature. Perm2 and Perm3 retain larger helical segments, despite their experimentally lower stability.

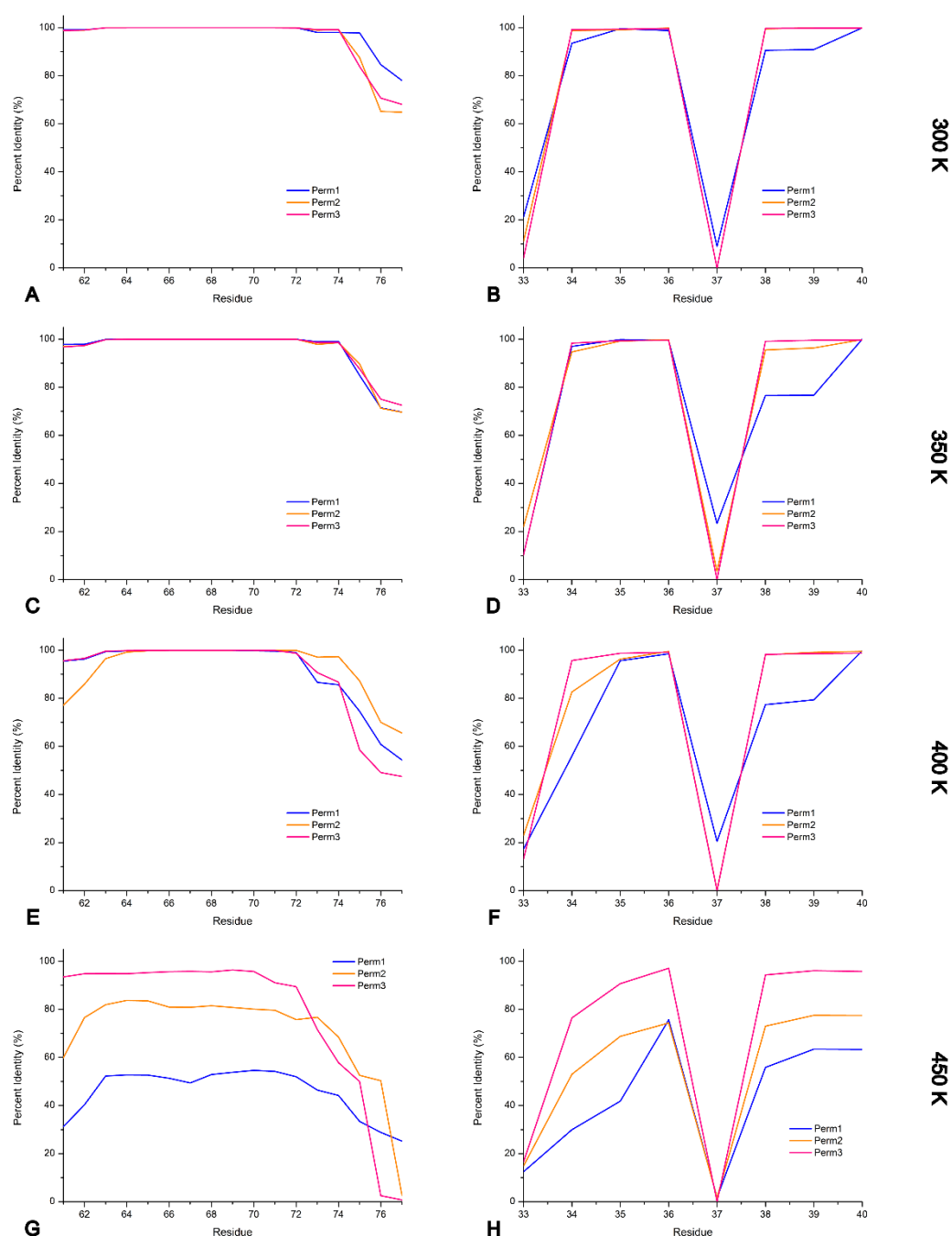


Figure 47. Residue-based secondary structure persistence of the three circularly permuted variants, evaluated by DSSP analysis at pH 2.5 along molecular dynamics simulations performed at increasing temperatures (300, 350, 400 and 450 K). Left panels report, for each temperature, the percentage of simulation time in which residues belonging to the selected α -helical segment retain α -helical assignment. At 300 K, all variants show near-complete conservation of the helix, with only minor losses toward the C-terminal edge. Progressive temperature increase leads to a gradual reduction in helical persistence, which becomes evident at 400 K and pronounced at 450 K, where Perm1 exhibits the most

extensive loss of α -helical character across the analysed region. Right panels show the percentage persistence of secondary structure assignment for a representative β -strand segment. At 300 K, the β -sheet is highly conserved in all three variants, with limited local reductions corresponding to loop-adjacent residues. At 350 and 400 K, transient losses of β -structure become more frequent, particularly in Perm1. At 450 K, β -sheet stability decreases markedly for all variants, although Perm3 retains higher persistence values in both strand and neighbouring loop regions, indicating a comparatively greater resistance of its β -architecture under extreme thermal stress.

4.2.5.4 PCA: calculation setup and trajectory projection

To investigate if molecular dynamics trajectories sample collective motions compatible with domain swapping, conformational fluctuations of permuted MNEI mutants were analysed using principal component analysis, starting in all cases from monomeric protein structures, in order to assess if the intrinsic dynamics of the isolated fold contain hints of hinge opening or large-scale reorganizations. PCA was used as a dimensionality reduction tool to identify the dominant low-frequency movements that explain most of the variance in the trajectories and that are most likely to correspond to cooperative rearrangements of the protein backbone. All analyses were performed starting from MD simulations of monomeric structural models generated using AlphaFold2, ensuring a consistent initial topology and a direct comparison between the different protein variants.

The analysis focuses on backbone dynamics and detects continuous shifts along collective coordinates rather than discrete transitions between predefined states. In this context, individual principal components describe movements involving multiple elements of the secondary structure, allowing local flexibility to be distinguished from large-scale reorganisations potentially associated with hinge opening and domain swapping.

The structural significance of the dominant principal components was assessed by mapping displacement vectors, allowing direct visualisation of the direction and amplitude of collective movements sampled during simulations. In this way, an intuitive link is provided between the statistical descriptors of movement and three-dimensional structural rearrangements,

forming the basis for comparing the dynamic behaviour of different permuted variants and topologically related proteins in relation to their experimentally observed domain swapping architectures.

4.2.5.4.1 PCA in MNEI

Results of principal component analysis on MNEI (PDB ID: 1FA3), show that the first principal component at 300 K is characterised by small-amplitude localised displacements, mainly confined to the region comprising residues Gly50-Phe51. (Figure 48-A) The observed movements remain below ~ 2 Å and do not involve reorganisation of the β -sheets or α -helix relative to the surrounding secondary structure elements. This indicates that the dominant collective movement corresponds to local flexibility of the supporting structure rather than large-scale cooperative movements. However, at 450 K (Figure 48-B), the amplitude of the displacements described by the first principal component increases slightly, reflecting the greater flexibility expected under conditions of severe thermal stress.

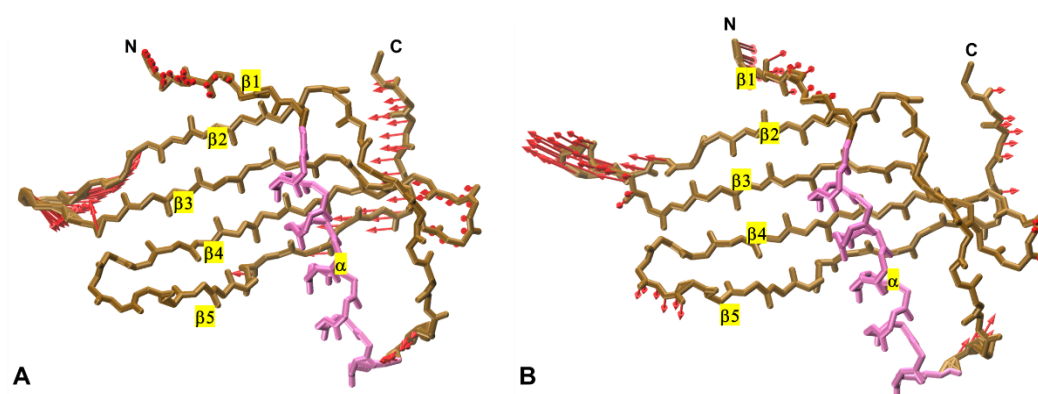


Figure 48. Dominant principal component displacement vectors mapped onto the backbone of wild-type monellin at 300 K (A) and 450 K (B). Arrows indicate the direction and relative amplitude of collective motions. Backbone is shown in ochre and the α -helical region (residues 9-27) in mauve.

Nevertheless, the spatial distribution of the movement remains largely preserved compared to the lower temperature simulation. No additional regions exhibiting coherent long-range displacements emerge, and the

overall integrity of the β -sheet core is preserved. Importantly, the α -helix undergoes no movement or changes.

The two temperatures analysed provide further evidence that the fold remains compact and tightly closed even under destabilising conditions, such as acidic pH and stress temperatures. Furthermore, the absence of extensive cooperative rearrangements compatible with domain exchange supports the role of the MNEI protein as a negative control system and highlights that domain exchange is not an intrinsic property of the cystatin fold, but rather depends on the accessibility of alternative conformational pathways promoted by topological modifications.

4.2.5.4.2 PCA in Perm1

Principal component analysis was applied to Perm1 to assess whether the dynamics of the monomeric state encode collective movements compatible with the domain-switching architecture observed crystallographically. (Figure 49)

At 300 K, the displacement vectors associated with the dominant principal component are largely confined to peripheral regions, with the most prominent contributions located at exposed ends and segments accessible to the solvent. (Figure 50-A) Although smaller vectors can be detected near the hinge region identified in the crystallographic domain-exchange assembly, their amplitude and spatial coherence remain limited, and the overall fold maintains a compact organisation. In this regime, the dominant collective motion therefore mainly reflects localised fluctuations rather than structural rearrangement, consistent with the idea that, at room temperature, Perm1 is not exploring an open conformation, although it is capable of reaching the swapped state under selective conditions (crystallisation, long timescales).

At 450 K, a distinctly different behaviour is observed. Under strong thermal stress, the displacement vectors of the first component concentrate on the hinge loop and extend to the α -helical segment, which shows pronounced collective vectors along its length and near its anchoring region, in line with

reduced mechanical stability of the helix and the onset of large-scale cooperative reorganisation. (Figure 50-B) This redistribution of the dominant mode is consistent with the RMSF analysis, which identifies a sharp increase in flexibility for Perm1 in the $\beta 1$ to α region at 450 K (values sometimes exceeding ~ 5 Å), compared to the more subdued response observed for Perm3. This indicates that Perm1 is mechanically stable at 300 K, while at elevated temperatures the dominant collective motions specifically involve structural determinants relevant to both domain swapping (hinge loop) and the α to β reorganisation correlated with the experimentally observed aggregation, suggesting that circular permutation and linker length predispose Perm1 to access alternative conformational pathways under destabilising conditions.

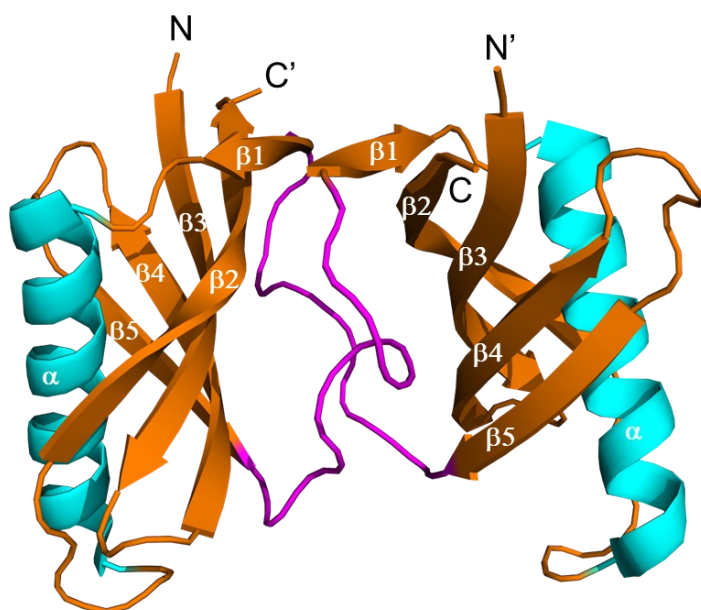


Figure 49. Crystal structure of the Perm1 dimer with domain swapping, where the segment of the main chain associated with the swap is highlighted in magenta to indicate the structural element involved in the quaternary rearrangement, while the β -strands (orange) and α -helix (cyan) are represented schematically.

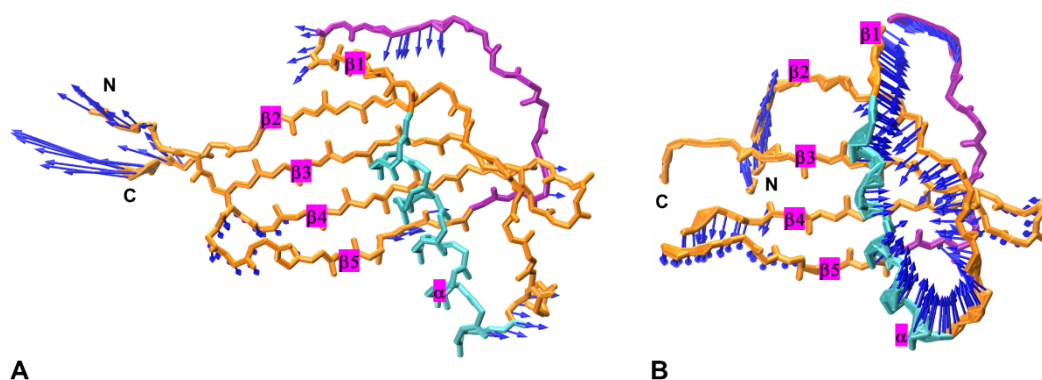


Figure 50. Dominant principal component displacement vectors mapped onto Perm1 at 300 K (A) and 450 K (B). Vectors are shown as blue arrows indicating direction and relative amplitude of collective motions. The backbone is represented in orange, the α -helical segment in cyan, and the hinge/swapping-associated region in purple for comparison with the crystallographic domain-swapped architecture in Figure 49.

4.2.5.4.3 PCA in Perm3

Principal component analysis was applied to Perm3 to assess whether the dominant collective motions sampled by the monomeric state are consistent with the domain swapping architecture observed experimentally. (Figure 51) At 300 K, the displacement vectors associated with the dominant principal component show a non-uniform spatial distribution, with the most significant contributions located along the upper region of the β sheet, particularly near $\beta 1$, and extending towards the hinge loop that constitutes the exchanged element in the crystallographic dimer with domain exchange. (Figure 52-A) In this scenario, the motion is not dominated by diffuse terminal fluctuations as in MNEI, but instead highlights a mechanically coherent region connecting the edge of the β sheet to the hinge segment associated with the exchange. In particular, the α -helical region remains outside the displacement vectors of the first component, indicating that the collective mode mainly reflects the dynamics centred on the β sheet/hinge, while preserving the overall organisation of the fold.

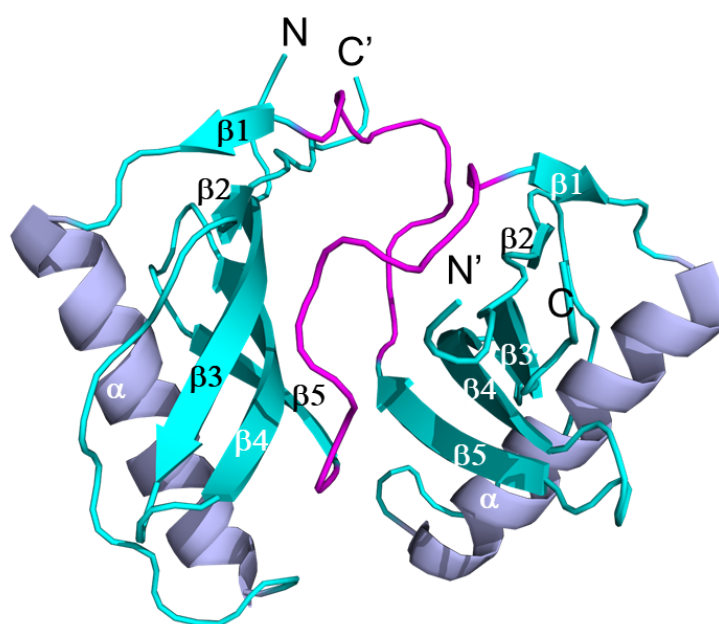


Figure 51. Crystal structure of the Perm3 dimer with domain swapping. The segment of the main chain associated with the swap (swapped) is highlighted in magenta, while the β strands and α helices are represented in stylised form to emphasise the overall cystatin-like folding.

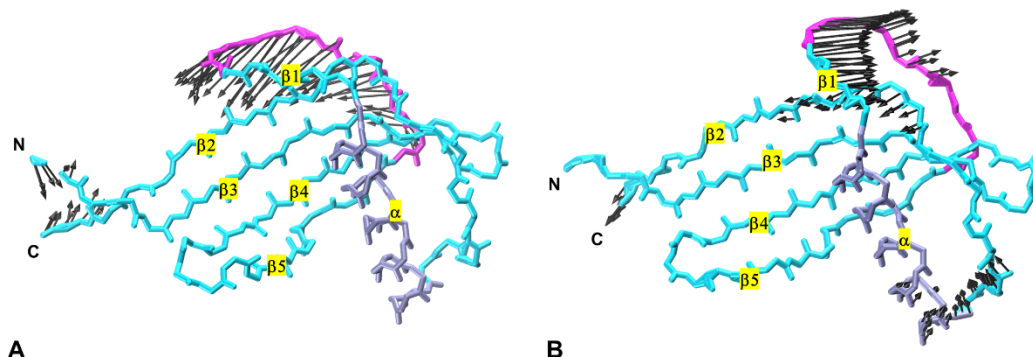


Figure 52. Displacement vectors of the first principal component mapped onto Perm3 at 300 K (A) and 450 K (B). The vectors are represented by black arrows indicating the direction and relative amplitude of collective movements. The protein structure is represented in cyan, the α -helical segment in iceblue, and the hinge region associated with the exchange in magenta.

4.2.5.4.4 PCA in L2MN

Principal component analysis was applied to the engineered L2MN variant, a monellin-derived construct specifically designed to promote domain exchange through the insertion of a five-residue motif (QVVAG) derived

from the cystatin family, as described in detail in the literature by Nandwani & co. [95] In contrast to MNEI, L2MN represents a reference system in which domain swapping has been demonstrated experimentally, making it suitable for correlating collective movements observed in molecular dynamics simulations with a known structural rearrangement. (Figure 53)

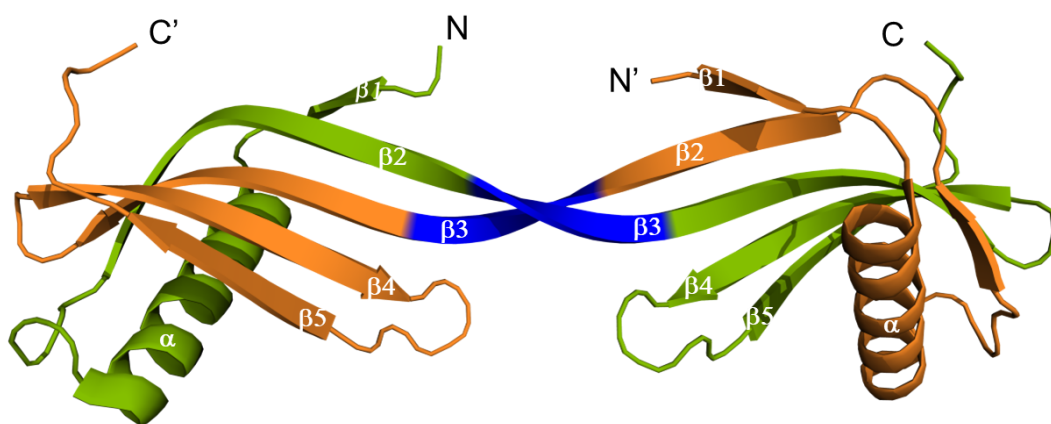


Figure 53. L2MN dimer with experimentally determined domain exchange (PDB ID: 5YCU), highlighting the $\beta 2$ - $\beta 3$ region involved in domain exchange. The exchanged segment corresponds to the engineered motif of five residues (QVVAG, residues 48-52), originally derived from the cystatin family and designed to promote domain swapping.

At 300 K, the displacement vectors associated with the first principal component are mainly located around the $\beta 2$ - $\beta 3$ loop comprising residues 48-52, corresponding to the engineered cystatin motif involved in domain swapping. The movements observed remain relatively limited in amplitude and do not involve a global reorganisation of the fold. (Figure 54) Unlike MNEI, however, the displacements in this region are more consistent and directional, indicating that the loop designed to mediate domain exchange is already dynamically pre-activated in the monomeric state. In fact, although weaker, additional contributions are detected in some β -strands, suggesting a potential reorganisation phenomenon between the loop subject to swap and the β -sheets, while the α -helix remains unchanged. The early and persistent involvement of the engineered $\beta 2$ - $\beta 3$ loop provides a dynamic correlation of the propensity for domain swapping observed

experimentally in L2MN, supporting the hypothesis that the construct is mechanically predisposed to domain swapping even in the monomeric state. (Figure 54)

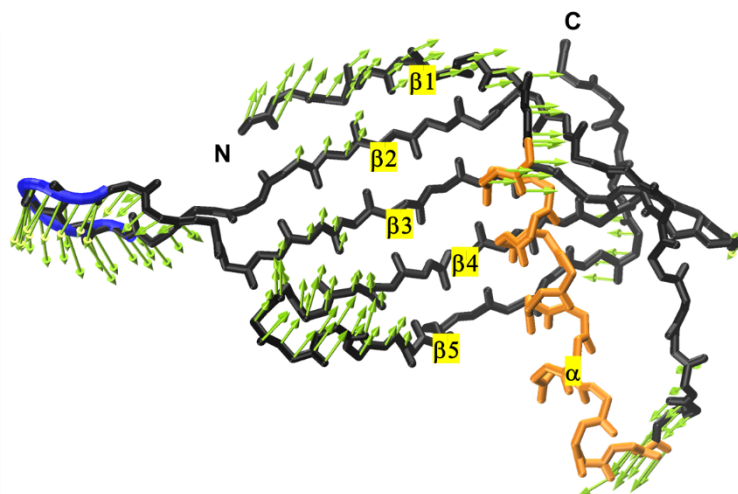


Figure 54. Displacement vectors associated with the first principal component mapped onto the L2MN structure at 300 K. The vectors are indicated by greenish arrows showing the direction and relative amplitude of the collective movements. The protein backbone is shown in black, the α -helical segment (residues 10-28) in orange, and the engineered β 2- β 3 loop involved in domain exchange (residues 48-52) in blue. The colours used allow a direct comparison with the swapped structure shown in Figure 53.

4.2.5.4.5 PCA in StefinB

Principal component analysis was applied to Stefin B as a canonical example of a protein with domain exchange whose quaternary rearrangement has been extensively characterised experimentally. (Figure 55)

At 300 K, the displacement vectors associated with the dominant principal component are scattered and located mainly in the peripheral regions, while the hinge loop involved in domain switching shows no significant collective displacement. In this state, the monomeric state of Stefin B appears dynamically stable, and the dominant fluctuations do not encode a clear pathway towards hinge opening or domain switching. (Figure 56)

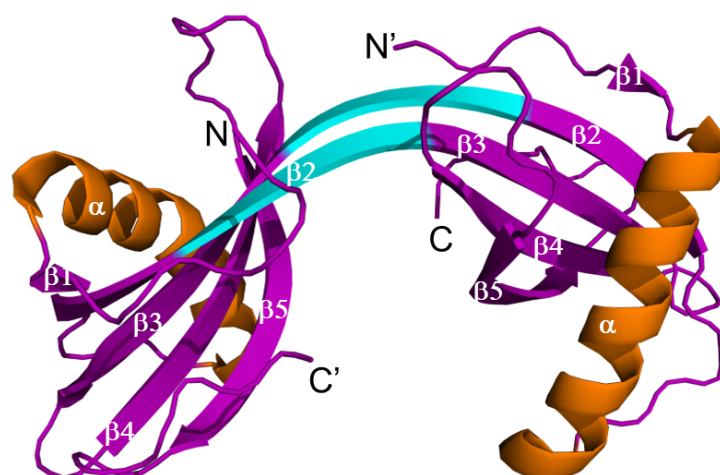


Figure 55. Crystal structure of the StefinB dimer (PDB ID: 2OCT) with domain swapping. The segment of the main chain associated with the swap (swapped) is highlighted in cyan, while the β strands (purple) and α helices (orange) are represented in stylised form to emphasise the overall cystatin-like folding.

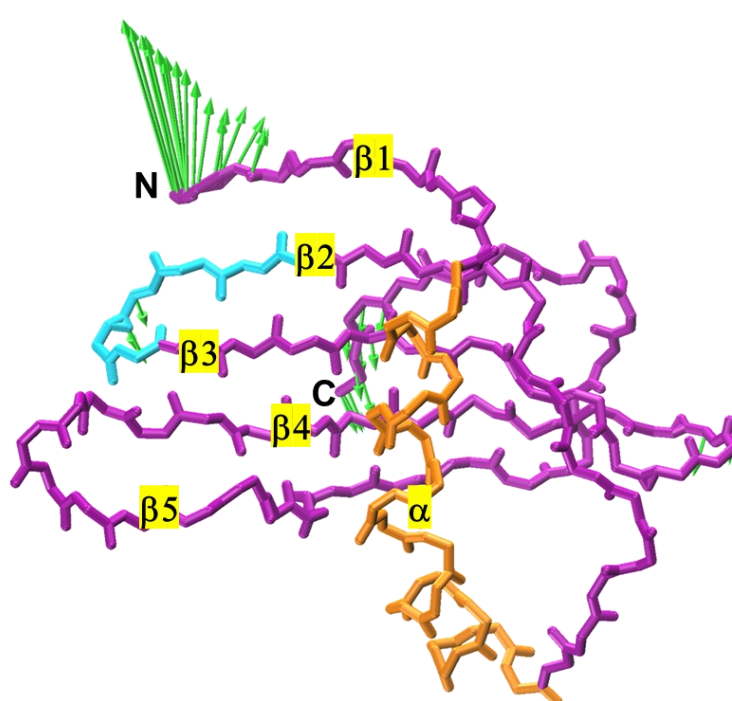


Figure 56. Displacement vectors of the first principal component mapped onto StefinB at 300 K. The vectors are represented by greenish arrows indicating the direction and relative amplitude of collective movements. The protein structure is represented in purple, the α -helical segment in orange, and the hinge region associated with the exchange in cyan.

4.2.6 Discussion

The sweet protein MNEI, a single-chain derivative of monellin, is not only known for its sweetening properties but above all because it is used as a model for studies on protein aggregation under extreme conditions of pH, temperature and salinity. To obtain further information on the molecular basis of aggregation, three new mutants were designed using the circular permutation technique. This technique is not a simple rearrangement of the sequence, but a topological intervention capable of remodelling deeply the structure by connecting the $\beta 2$ and $\beta 3$ loops of MNEI and joining the native ends with linkers of different lengths, thus maintaining the α/β cystatin fold substantially conserved. The application of this technique proved useful from the early stages of expression and purification, as all three mutants are produced with high yields and show good solubility behaviour, indicating that circular permutation does not compromise correct folding. However, crucial initial evidence emerged from the SEC profiles: Perm1, like MNEI, behaves in solution as a mainly monomeric species, Perm2 mostly as a monomeric species, while Perm3 is present in solution in equal amounts of monomer and dimer. These different behaviours in terms of stability suggest that, despite sharing the same opening site and overall fold, the different lengths of the linkers (longer in Perm1, intermediate in Perm2 and shorter in Perm3) can distinctly influence stability, oligomeric behaviour and propensity for aggregation pathways.

Circular dichroism analyses conducted at pH 2.5 and 6.8 show that, like MNEI, each mutant maintains its β -sheet content and that the most significant variations concern the α -helical and random coil content. A clear divergence emerges after evaluating the thermal denaturation profiles, which show that no mutant exhibits thermal stability superior to that of MNEI ($\sim 72^\circ\text{C}$ at pH 2.5 and $\sim 76^\circ\text{C}$ at pH 6.8). Despite this, Perm1 is closer to the native protein in terms of melting temperature ($\sim 67^\circ\text{C}$ at pH 2.5 and $\sim 71^\circ\text{C}$ at pH 6.8), while Perm2 ($\sim 63^\circ\text{C}$ at pH 2.5 and $\sim 67^\circ\text{C}$ at pH 6.8) and Perm3 ($\sim 60^\circ\text{C}$ at pH 2.5 and $\sim 63^\circ\text{C}$ at pH 6.8) deviate more, indicating progressive destabilisation as the linker length decreases. In addition, at

acidic pH, *i.e.* under conditions that promote aggregation for MNEI, the thermal denaturation of the mutants does not stabilise alternative states to the native fold and the protein is able to refold upon cooling. At neutral pH, however, thermal denaturation leads to the formation of more stable alternative states to the native fold that prevent renaturation, regardless of the presence of salt.

In addition, the propensity for aggregation of the engineered mutants was monitored by ThT binding studies and validated by TEM and SEC analyses. These analyses were conducted at 55 and 50 °C, two temperatures corresponding to the initial unfolding states. At pH 2.5 and in the presence of NaCl, all three mutants show an increase in the ThT signal at a temperature close to T_m , but the correlation between fluorescence intensity and final morphology of the aggregates is not linear. Perm1, despite showing slower kinetics and lower ThT levels, progressively evolves into long, thin, organised fibrils at both temperatures, even if with different morphologies. On the contrary, Perm2 and especially Perm3 show a faster and higher ThT response, but TEM images reveal a predominance of amorphous aggregates. This result indicates that circular permutation not only regulates the rate of self-association, but above all the selection of the aggregation pathway, determining whether the system converges towards orderly fibrillar growth or is deviated early towards off-pathway states. This mechanism is further validated by SEC chromatography data obtained on aliquots of mutants incubated under aggregation conditions, which revealed that Perm1 remains in solution mainly as a monomer at both temperatures and that the amount of dimer tends to decrease with increasing incubation time. Perm2, on the other hand, undergoes dimerization at 50 °C, while at 55 °C it rapidly precipitates as the incubation time increases. Finally, Perm3 tends to precipitate like Perm2 as the incubation time increases at 55 °C, while at 50 °C it is progressively converted to a dimer. This data is corroborated by Native-PAGE conducted at 50 °C and under the same aggregating conditions, which shows that after 24 hours of incubation, Perm1 does not accumulate detectable soluble oligomers, suggesting a

monomeric pathway leading to the formation of orderly and stable fibrils. Perm2 and Perm3, on the other hand, show a marked tendency to oligomerisation, favouring the formation of disordered aggregates.

Crystallographic structures complete the picture by showing that all three mutants can adopt a domain-swapped quaternary architecture, despite the marked differences observed in solution. In all cases, structured regions of each protomer exchange with the corresponding regions of the other, giving rise to a highly symmetrical intertwined dimer.

The resulting domain-swapped architecture is partially superimposable, at the level of the individual protomers, on that observed in other engineered constructs belonging to the cystatin family, such as L1MN/L2MN and Stefin B, in terms of α/β fold. [95,197]

However, the structural organization of the two subunits differs significantly. In the L1MN/L2MN and Stefin B constructs, the interprotomeric interface is linked to the exchange that occurs mainly through the $\beta 2$ - $\beta 3$ loop, whereas in MNEI mutants it occurs through the linker introduced with CP between the two protomers. In fact, Perm2 and Perm3 share a more closed interface, characterised by a compact and highly symmetrical geometry, stabilised by a dense network of ionic contacts. Perm1, on the contrary, adopts a more open interface in which the linker contributes directly to the contacts between the protomers.

These differences suggest that the length of the linker plays a decisive role in defining the geometry of the domain-swapped dimer. In particular, the shorter linker of Perm3 imposes a more rigid and reproducible configuration, consistent with the rapid formation of crystals after a few hours of incubation. In contrast, the longer linker of Perm1 reduces the propensity for association in solution, in agreement with the ThT and TEM results, making the formation of the domain-swapped dimer a rare and slow event, which requires significantly longer incubation times (up to four months).

Molecular dynamics simulations and PCA analysis provide an additional level of understanding of the experimental data. In fact, molecular dynamics simulations conducted under experimentally aggregating conditions and at

four temperatures (300, 350, 400, 450 K) on the three circularly permuted proteins highlight possible hot spots involved in the aggregation process. RMSD analyses show that at 300 K, all analysed systems are stable ($< 3 \text{ \AA}$) with RMSF flexibility mostly localised at the terminal ends and in the loop where the linker was inserted; at higher temperatures, all mutants show larger deviations ($> 6 \text{ \AA}$) and progressively higher RMSF flexibility in the α H- β 2 loop, but only Perm1, at a temperature of 450 K, is the most flexible of the three circularly permuted mutants, losing almost completely the α helix structure and showing a higher RMSF than Perm2 and Perm3 in the loop between the β 4 and β 5 strands. These results suggest that the flexibility of this region, together with that of the α H- β 2 loop, could lead to the opening of the monomer from the helix side and thus to the formation of ordered fibres, as already proposed for the aggregation of MNEI and its mutants. [78,195] Statistical analyses confirm this assumption. PCA analysis allows for further reorganisation of the experimental data obtained on the domain-swapped architectures observed in the circularly permuted mutants of MNEI, enabling us to describe how circular permutation redistributes collective motions along the polypeptide chain. In particular, the analysis of the first principal component (PC1) allows us to identify, for each system and thermal condition, the regions most affected by collective movements and potentially involved in the mechanisms of domain opening and exchange.

In this context, MNEI represents a negative reference, since at both temperatures considered it shows small amplitude fluctuations, compatible with the failure of the hinge to open and domain exchange. A very different situation emerges from the analysis of Perm1 and Perm3: in Perm1, even if at 300 K the first principal component reflects fluctuations at the end and localised on the hinge loop, it is precisely at 450 K that is observed a marked reorganisation of collective motions involving both the hinge loop and the α -helix, which loses its structural behaviour; in Perm3, on the other hand, already at 300 K the collective motions are distributed on the hinge loop involved in the domain exchange and on the edge of the β -sheet, but without

the involvement of the α -helix. This difference could indicate that Perm3 accesses the swapped state through a more local opening mode (hinge/ β -sheet), while Perm1 accessed it through a global opening mechanism, compatible with the domain swapping observed crystallographically and with its marked ability to form ordered fibrils.

Overall, these results are consistent with crystallographic findings and suggest that circular permutation affects both stability and collective dynamics related to domain swapping. However, it is important to note that PCA analysis provides a description of the most evident collective motions along the simulations but does not directly identify the structural regions responsible for domain swapping. This limitation is clearly evident from the analysis of the reference systems. In L2MN, in fact, although the β 2- β 3 loop is designed to mediate domain swapping, the main motions are distributed over several elements of the β -sheet, without isolating the loop as the only dynamically relevant element. Similarly, in Stefin B, the absence of significant collective motions on the hinge loop does not contradict its well-known propensity for swapping, but reflects the fact that, in single-chain simulations, the movements associated with swapping do not necessarily emerge as dominant collective motions. It should be noted that all MD simulations were performed on monomeric structures, and therefore do not explicitly take into account the intermolecular interactions that might occur in dimeric or domain-swapped states. While this approach allows isolation of intrinsic conformational dynamics, it does not capture how two protomers might interact before domain swapping. Future studies including simulations of dimeric systems could provide further insight into the early stages of intermolecular association and the role of domain interfaces in guiding aggregation pathways. Together, these results outline a coherent framework in which the circular permutation technique acts as a powerful topological tool capable of modulating, at best, the structural and functional properties of circularly permuted mutants of MNEI. Comparisons made on MNEI and other engineered variants have shown that, despite maintaining the α/β cystatin fold, small differences in the connectivity of the polypeptide chain

of circularly permuted mutants, such as in the length of the linker, are sufficient to profoundly alter the balance between monomeric stability, domain exchange, and aggregation pathways.

Previous studies conducted on MNEI and other protein systems, indicate that the stability of the monomeric state in solution is a determining factor for the formation of ordered fibrillar aggregates, while a greater propensity to form dimers with domain exchange is correlated with the formation of amorphous oligomeric aggregates off-pathway. [195] This supports the findings, as Perm1 appears to be predominantly monomeric in solution, favouring the progressive and ordered formation of fibrils, while Perm2 and Perm3, found in solution in dimeric form, not only go through earlier domain swapping, but also deviate from the aggregation pathway towards amorphous aggregates.

These results support the idea that domain swapping and aggregation are not coupled processes, even within structurally related protein families, but instead depend on a complex interaction between topology, dynamic flexibility, and environmental conditions. Indeed, rather than acting as a generic precursor to fibrillation, domain swapping appears to represent an alternative pathway of self-assembly that competes with and often suppresses the ordered growth of fibrils, in agreement with observations reported for other globular proteins and ribonuclease-based systems. [198] Beyond their fundamental implications for understanding protein aggregation mechanisms, these observations open up relevant perspectives for protein engineering and biotechnology. The possibility of understanding the interaction between the different factors that influence the stability of MNEI in solution provides useful insights for preventing aggregate formation, considering its potential as a powerful sweetener.

At the same time, controlling the balance between monomeric stability and fibrillar aggregation highlights circular permutation as an effective strategy for designing nanostructured biomaterials. [15,136,137]

4.3 PIRs

4.3.1 Concept and design of PIR constructs

Based on the modular reduction rationale outlined in the Introduction (see Section 1.3), the experimental activity was set up with the aim of assessing whether a single PIR repeat was enough to preserve the main biochemical and structural properties attributed to multi-repeat PIR proteins. To this end, constructs containing a single internal repeat (r1) derived from different PIR proteins of *S. cerevisiae* were designed and analysed, regardless of the number of repeats present in the respective native proteins.

In the initial stages of the work, attention was focused on constructs already available in the host laboratory, particularly the GFP-PIR1r1-8 fusion, containing the complete PIR1 protein with eight internal repetitions. This system was used as an experimental reference to optimise heterologous expression and purification conditions in *E. coli*, exploiting GFP fluorescence as a direct tool to monitor protein expression, solubility and recovery during chromatography steps. In parallel, the study of PIR1r1, containing a single repetition, was introduced as the first example of a reduced construct, useful for evaluating the stability and biochemical behaviour of an isolated PIR module.

Subsequently, once the plasmids for the other assigned proteins became available, the same modular reduction strategy was extended to the PIR2, PIR4 and PIR5 variants, all designed in a single-repeat form (PIR2r1, PIR4r1 and PIR5r1). This approach allowed us to build a set of homologous constructs, differentiated only by the sequence of the PIR region, to be used for a systematic comparison of expression properties, solubility, stability and structural behaviour.

All the single-repeat constructs analysed in this work have a common N-terminal architecture, comprising a His₆ purification tag followed by a cleavage site for the TEV protease, designed to ensure maximum experimental comparability. The C-terminal portion of each construct includes the region derived from the respective native PIR protein, selected to include the repeated module of interest. This design choice minimises the

influence of the vector, tag or cloning differences on the observed properties, allowing any variations to be attributed primarily to the PIR module sequence.

The definition of r1 constructs is based on the identification of the consensus motif characteristic of PIR proteins, associated with the sequence SQGDGQQA. [126] In the PIR1r1, PIR4r1 and PIR5r1 constructs, there is a single, clearly identifiable canonical repetition. In the case of PIR2r1, even though the construct was designed as a single repetition, sequence analysis reveals the presence, in addition to the canonical repetition, of an adjacent segment corresponding with a partial or degenerate repetitive motif. This region has been retained in the construct as part of the natural context immediately surrounding the complete repetition, configuring PIR2r1 as an intermediate case useful for evaluating the contribution of incomplete repetitive portions to the biochemical and structural properties of the PIR family.

To directly compare the sequences of the r1 modules and verify their degree of conservation, a multiple sequence alignment (MSA) was performed using ClustalΩ (Figure 57). The alignment shows marked conservation in the central region of the constructs, corresponding to the core of the PIR motif, compared to greater variability in regions of low complexity, characterised by enrichment in polar residues and frequent insertions and deletions. These results support the idea that PIR proteins share a conserved functional core embedded in a flexible and variable context, consistent with their structural role in the cell wall.

PIR5_r1	MGSSHHHHHHHSSGLVPRGSHMSSGENLYFQSWATLTPSSKMDGGTTEYRTSFGLAIVPFT	60
PIR4_r1	MGSSHHHHHHHSSGLVPRGSHMSSGENLYFQSWSTLTPTGSISSGAEEYTTFGIAVQAIT	60
PIR1_r1	MGSSHHHHHHHSSGLVPRGSHMSSGENLYFQSWSTLTPTSATYKGGITDYSSSTFGIAVEPIA	60
PIR2_r1	MGSSHHHHHHHSSGLVPRGSHMSSGENLYFQSWSTLTPTATYSGGVTDYASTFGIAVQPIS *****::**::*:***:**::	60
PIR5_r1	VTESVKV-----RVVSHIVSQIGDGQLQITTAKNVVTKSTIAVPskVTATATSSGGSG	114
PIR4_r1	SSKA-----KRDVISQIGDGOVQATSAAATAQATDSQAATTAT--TPTSSSGSGG	108
PIR1_r1	TTASSK-----AKR-----AAAAAISQT- G DGQIQATTK	86
PIR2_r1	TTSSASSAATTASSAKRAASQIGDGOVQAATTTASVSTKSTAAAVSQT- G DGQISGGSG * ..:	119
PIR5_r1	GSGGSGGNEKVIKVQACKSSGTLAITLQGGLVIDSSGRIGSIVANRQFQFDGPPPPQAGAI	174
PIR4_r1	GS-GSGGG-PSLKDSSCKNSGTLETLTKDGLVTDAGKRIGSIVANRQFQFDGPPPPQAGAI	166
PIR1_r1	TSGGSGGG-DPVTAETCKSSGTLEMNLKGGILTdgkgRIGSIVANRQFQFDGPPPPQAGAI	145
PIR2_r1	GSG-GSGG-DPVDVASCkTSGTLEMNLKGGILTdgkgRIGSIVANRQFQFDGPPPPQAGAI * .* : :*:.**** .:*.:* *..*****	177
PIR5_r1	YAGGWSITKHGTLAIGDNDVFYQCLSGTFYNLYDQSIGGCNCPVHLQTVGLVDC	228
PIR4_r1	YAAGWSITEDGYLAGDSDFYQCLSGNFNLYDQNVAEQCSAHLEAVSLVDC	220
PIR1_r1	YAAGWSITPEGNLAIGDQDTFYQCLSGNFNLYDEHIGTCQNAVHLQAIDLLNC	199
PIR2_r1	YAAGWSITPDGNLAIGDNDVFYQCLSGTFYNLYDEHIGSQCTPVHLEAIDLIC *.***** .* **:*.*.*****.*****: .: *:***:~::~*	231

4.3.2 Exploratory expression and solubility tests

4.3.2.1 GFP-PIR1r1-8

GFP-PIR1r1-8 was expressed in *E. coli* Shuffle T7 cells in a total volume of 4 L (see Section 3.2.1.3), inducing expression at 30 °C with 0.1 mM IPTG for 4 hours. To follow the temporal trend of expression, culture aliquots were collected at time 0 (immediately before induction) and at subsequent intervals up to 4 hours post-induction. SDS-PAGE analysis of these samples (Figure 58) shows a progressive increase in the band corresponding to

GFP-PIR1r1-8, indicating efficient and stable expression over time, with no obvious signs of degradation.

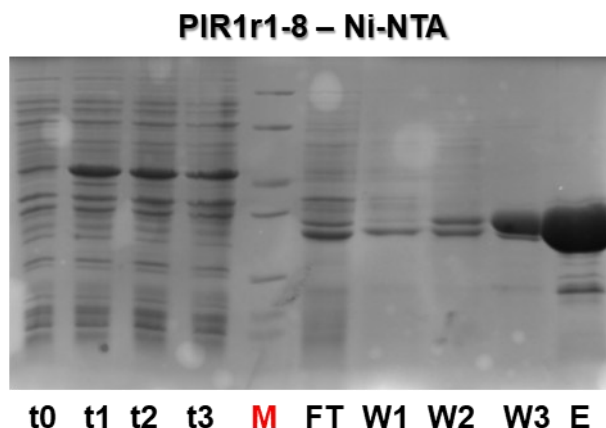


Figure 58. SDS-PAGE analysis of GFP-PIR1r1-8 expression and purification. Whole-cell lysates collected before induction (T0) and at increasing times after IPTG induction at 30 °C (T1-T3) show a progressive accumulation of GFP-PIR1r1-8. Lane M, molecular weight marker; FT, flow-through; W1-W3, wash fractions; E, elution fraction, containing the target protein at the expected molecular weight.

At the end of expression, the cells were harvested and lysed by high-pressure homogenisation using an Avestin® Emulsiflex-C5 system (see Section 3.2.2.4). (Figure 59) The crude lysate was then clarified by centrifugation and filtration of the supernatant (see Section 3.3.1). (Figures 59-B-C) Purification of the recombinant protein was performed by affinity chromatography on Ni-NTA resin (see Section 3.3.2.2), using PIR buffer solution (50 mM Tris, 200 mM NaCl, pH 7.0; see Section 2.3.1), a buffer commonly used in the laboratory for GFP-PIR1r1-8 and PIR1r1. However, it is important to note that GFP-PIR1r1-8 is a large protein (422 amino acids; theoretical molecular weight of 45058 Da) and has a calculated isoelectric point of approximately 7.3. Consequently, the purification conditions adopted are close to the isoelectric point of the protein, which is potentially critical in terms of solubility and conformational stability.

During affinity chromatography, a structured washing and elution protocol was therefore adopted, slightly different from the standard protocol

described in Section 3.3.2.2, which still included the subsequent washing steps, but at the end, elution was performed with an imidazole cleaning-elution solution containing 500 mM imidazole (see Section 2.3.1), all maintained at pH 7.0. Representative aliquots of each fraction (flow-through, washes and eluate) were analysed by SDS-PAGE (Figure 58), showing effective retention of the protein on the resin and concentrated elution in the final fractions.

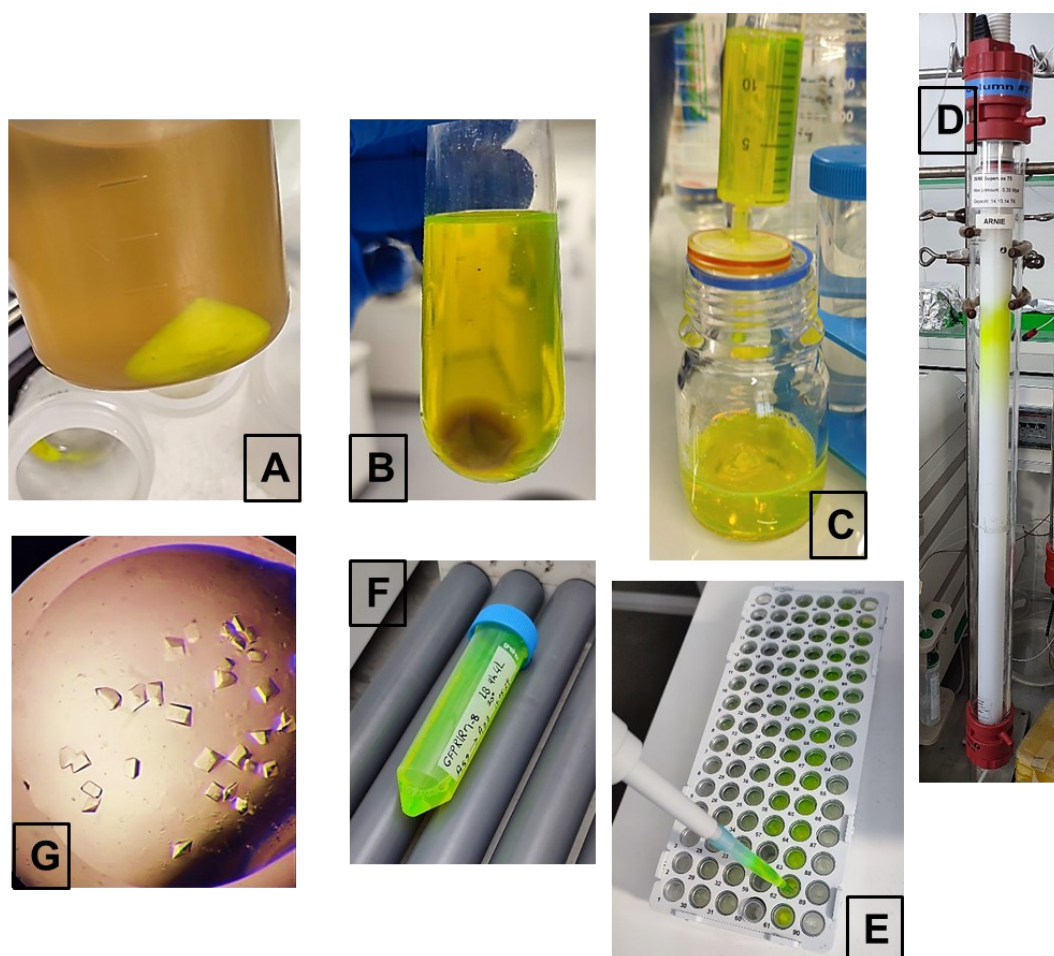


Figure 59. Representative steps of expression, purification, and downstream processing of GFP-PIR1r1-8. (A-B) Cell lysis and clarification by centrifugation, showing the characteristic green colouration of GFP-containing samples. (C-D) Loading and elution by affinity chromatography. (E) Collection of fractions eluted from SEC. (F) Protein sample concentrated with TEV protease on a roller for cleavage of the HIS tag. (G) Representative crystals obtained from purified GFP-PIR1r1-8.

The eluate containing GFP-PIR1r1-8 was then concentrated using centrifugal filters with a cut-off of 30 kDa (see Section 3.3.4) and subsequently subjected to size exclusion chromatography (SEC) on a HiLoad™ Superdex 75 pg (26/60) column (Cytiva) column, with a bed volume of approximately 320 mL (see Section 3.3.4). In this step, the presence of GFP allowed direct visual monitoring of the purification progress, making it possible to follow the behaviour of the protein along the column and during fraction collection in real time. (Figures 59-D-E)

The chromatographic profile obtained (Figure 60, on the right) shows the presence of a small initial peak, followed by a well-defined main peak. SDS-PAGE analysis of the corresponding fractions (Figure 60, on the left) indicates that the main peak contains most of the GFP-PIR1r1-8 in an apparently monodisperse form, while the initial minor peak is compatible with an aggregated or oligomeric fraction. The consistency between the SEC profile and the electrophoretic analysis confirms the suitability of the GFP-PIR1r1-8 system as a reference model for evaluating solubility, chromatographic behaviour and protein preparation quality.

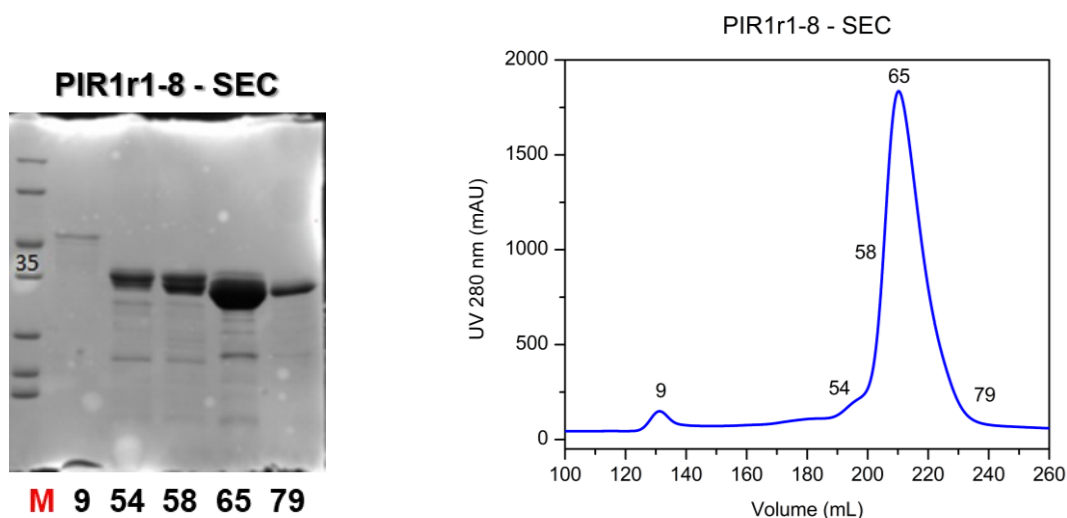


Figure 60. Size-exclusion chromatography (SEC) of GFP-PIR1r1-8 and analysis of collected fractions. SEC profile shows a minor early-eluting peak and a dominant main peak. SDS-PAGE of selected fractions indicates that fraction 9, corresponding to the initial peak, doesn't contain the target protein, whereas fractions 54-79 from the main peak contain GFP-PIR1r1-8 in a homogeneous form.

Overall, these results provided a solid experimental basis for optimising the expression and purification conditions of PIR proteins and represented a fundamental starting point for the analysis of GFP-free single-repeat constructs, in particular PIR1r1, discussed in the next section.

The fractions corresponding to the main peak of size exclusion chromatography were pooled and used for a subsequent stage of protein processing, aimed at removing GFP by proteolytic cleavage. (Figure 59-E). For this purpose, TEV protease was added to the SEC eluate collected in a Falcon container at an enzyme:substrate ratio of 1:100 (w/w), using a 200 μ L aliquot of TEV protease at a concentration of 6.5 mg/mL. The cleavage reaction was carried out by incubating the mixture at 4 °C for 24 hours under slow agitation on a roller, in order to promote accessibility of the recognition site while maintaining conditions compatible with the stability and specificity of the enzyme. (Figure 59-F).

At the end of incubation, the solution was centrifuged and filtered to remove any precipitates or aggregates, then concentrated for subsequent crystallisation. Under these conditions, the cleaved protein showed favourable crystallisation behaviour, leading to the formation of well-defined crystals after approximately 21 days of incubation. The crystals were obtained at a final protein concentration of 25.2 mg/mL, using a crystallisation buffer containing 25% (w/v) PEG 4000, 0.2 M NaOAc and 0.1 M Tris at pH 8.5 (Figure 59-G).

These results indicate that, despite the use of initial purification conditions at the limit for a large protein such as GFP-PIR1r1-8, the cleavage of the GFP tag and subsequent concentration of the protein made it possible to obtain material suitable for structural studies, further confirming the usefulness of this construct as a reference system for setting up experimental strategies applied to single-repeat PIR constructs.

4.3.2.2 PIR1r1

Unlike the GFP-PIR1r1-8 construct, PIR1r1 does not have a fused fluorescent protein and consists exclusively of the single-repeat PIR module. The PIR1r1 construct has a length of 199 amino acids, a theoretical molecular weight of 20779.93 Da and a calculated isoelectric point of 6.10. These characteristics make it significantly different, from a physicochemical point of view, from GFP-PIR1r1-8, both in terms of size and charge distribution, with potential implications for solubility and behaviour in heterologous expression. Initial attempts to express PIR1r1 under the same conditions used for GFP-PIR1r1-8 (*E. coli* Shuffle T7 cells, induction at 30 °C for 4 hours with 0.1 mM IPTG) did not produce detectable amounts of soluble protein. Following these results, expression was repeated using the same host strain but modifying the induction conditions, reducing the temperature to 21 °C and prolonging expression overnight. Under these conditions, PIR1r1 showed more consistent expression, although overall lower than that observed for the GFP-fused construct.

Like GFP-PIR1r1-8, PIR1r1 was purified using PIR buffer solution (50 mM Tris, 200 mM NaCl, pH 7.0; Section 2.3.1) and Ni-NTA affinity chromatography. SDS-PAGE analysis of the different fractions collected during purification (Figure 61) shows the presence of the band corresponding to PIR1r1 mainly in the elution fractions, but in significantly reduced quantities compared to the protein background, indicating a limited overall yield. Despite this, the eluted protein is sufficiently defined to proceed to the next stages of concentration and characterisation.

The eluted sample was concentrated and subsequently subjected to size exclusion chromatography, whose chromatographic profile (Figure 62) shows a well-defined main peak, compatible with a predominantly monomeric population, accompanied by a minor signal at lower elution volumes, indicative of the presence of aggregated or oligomeric species in limited quantities. The fractions corresponding to the main peak were collected and used for crystallisation assays.

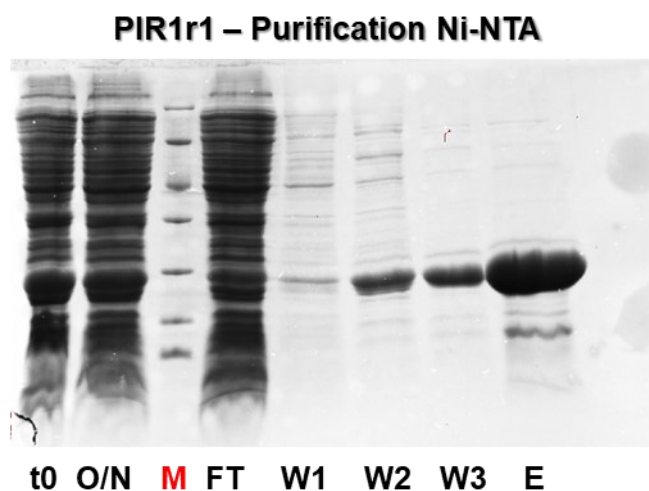


Figure 61. SDS-PAGE analysis of PIR1r1 expression and purification. Whole-cell lysates collected before induction (T0) and after IPTG induction at 21 °C overnight show a progressive accumulation of PIR1r1. Lane M, molecular weight marker; FT, flow-through; W1-W3, wash fractions; E, elution fraction, containing the target protein at the expected molecular weight.

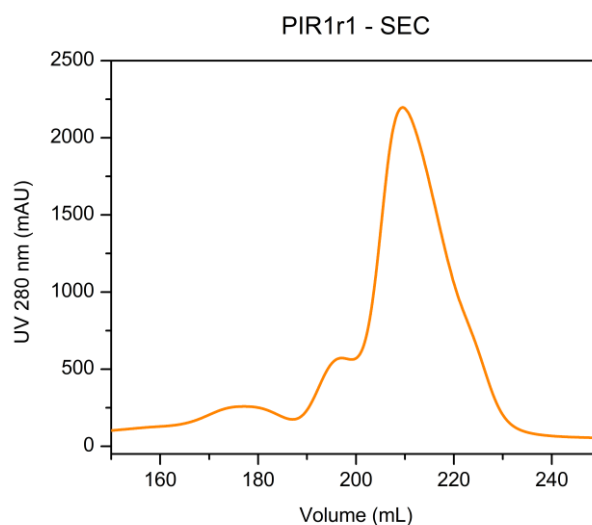


Figure 62. Size-exclusion chromatography (SEC) profile of PIR1r1. A dominant main peak is observed, preceded by a minor early-eluting signal.

Despite the low overall yield of the purification, PIR1r1 proved to be suitable for crystallisation. The crystals were obtained by the hanging/pendant drop vapour diffusion method at 8 °C, using protein solutions at two different concentrations: 25.2 mg/mL (Job1) and 14.8 mg/mL (Job2), both prepared

in 20 mM Tris, 200 mM NaCl, pH 7. The crystallisation drops were prepared by mixing equal volumes of protein solution and precipitating solution (1:1 ratio) and equilibrated against a reservoir in 96-well plates. The precipitating solution contained 30% (w/v) PEG 4000, 0.2 M NaOAc and 0.1 M Tris at pH 8.5. The first crystals appeared after about 8 days and continued to grow for up to 21 days (Figure 63).

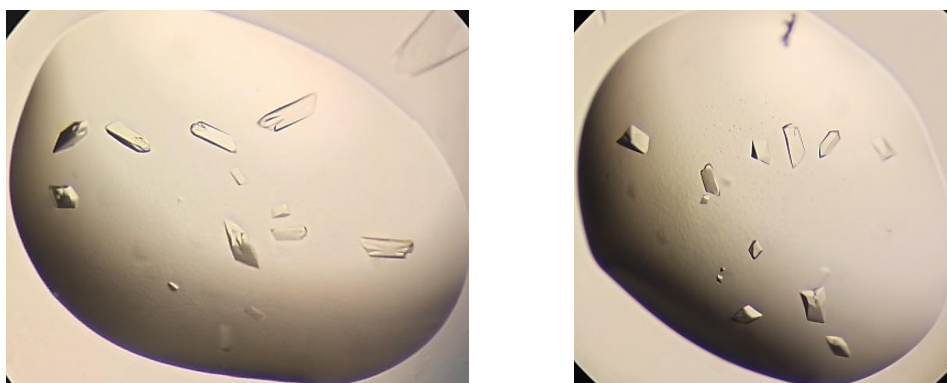


Figure 63. Representative crystals of PIR1r1 obtained by hanging-drop vapour diffusion at 8 °C under PEG 4000-based conditions.

Overall, these results indicate that, despite presenting significant difficulties in terms of expression and yield, PIR1r1 is capable of forming crystals under conditions relatively similar to those used for GFP-PIR1r1-8 after tag removal, suggesting that the single-repeat PIR module has an intrinsic propensity for structural order, despite the absence of the multiple repeats present in the native protein.

4.3.2.3 PIR2r1

PIR2r1 is a single-repeat protein composed of 231 amino acids, with a theoretical molecular weight of 23406.57 Da and a calculated isoelectric point of 5.59. Compared to PIR1r1, PIR2r1 is therefore slightly larger but has a more acidic charge distribution, which is potentially relevant for its behaviour during lysis and affinity purification. Considering the difficulties encountered with PIR1r1 in terms of yield and recovery, extensive screening

of expression conditions was initially carried out for PIR2r1 before proceeding with larger-scale purification attempts.

The plasmid DNA encoding PIR2r1, obtained by miniprep from *E. coli* DH5 α cells, was tested in two *E. coli* strains commonly used for recombinant protein expression: BL21(DE3)Gold and Shuffle-T7. For each strain, two growth media (LB and TB) and two induction regimes were evaluated: a higher temperature condition (30 °C for 4 hours) and a milder condition (21 °C overnight). (Figure 64-A) Screening was initially conducted in 3 mL overnight cultures, followed by inoculation of 500 μ L into 50 mL of fresh culture. For each condition, a 1 mL aliquot of culture was taken, centrifuged, and the pellet resuspended in 1 $\times$ SDS buffer to directly compare uninduced and induced samples by SDS-PAGE. (Figure 65)

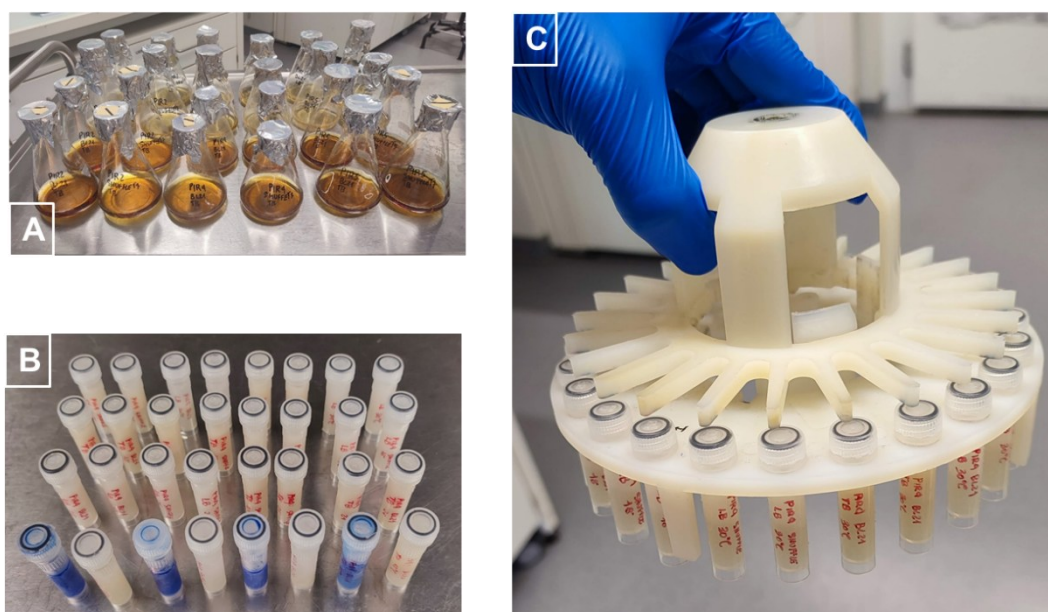
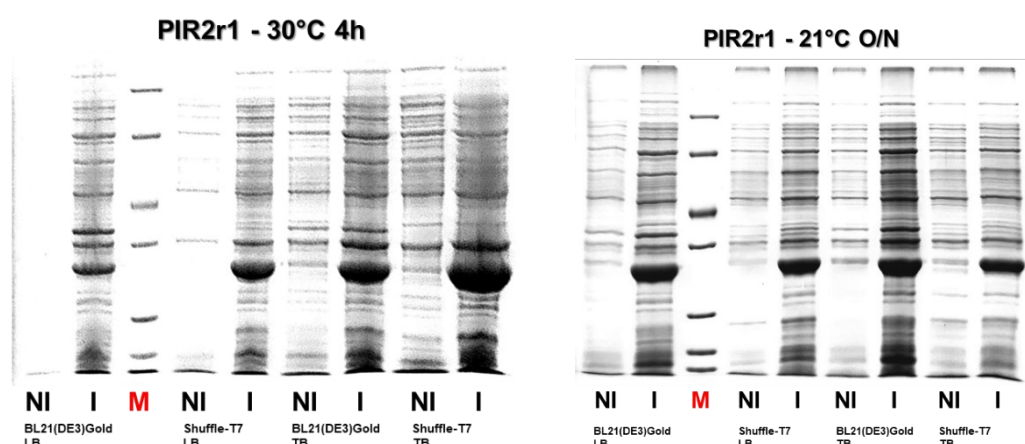


Figure 64. (A) Bacterial cultures pre-harvest for PIR protein expression assays. (B) Clarified cell lysates distributed into individual screw-cap tubes with glass beads. (C) Samples loaded into a FastPrep centrifuge rotor prior to downstream purification steps.

Gel analysis shows that PIR2r1 is expressed under all conditions tested, both in BL21(DE3) Gold and in Shuffle T7. In screening gels, the band attributable to PIR2r1 is reproducibly located between the 25 and 35 kDa markers, thus at an apparent weight (~30 kDa) higher than the theoretical

molecular weight of the construct (~23.4 kDa). This discrepancy is consistent with abnormal migration in SDS-PAGE, frequently observed for proteins enriched in regions of low complexity and/or with an unbalanced amino acid composition, which can bind SDS in a non-canonical manner and exhibit reduced mobility. Minor contributions may arise from the cloning sequence (His₆-tag and TEV site) and, potentially, from the presence of partially SDS-resistant oligomeric species; however, the presence of a consistent and comparable main band across the various conditions suggests that expression is robust and that the mobility anomaly does not affect the qualitative interpretation of the expression screening.



Figures 65. SDS-PAGE analysis of PIR2r1 expression testing under different induction conditions, showing non-induced (NI) and IPTG-induced (I) samples expressed in different *E. coli* strains and culture media, comparing induction at 30 °C for 4 hours and at 21 °C overnight. M, molecular weight marker.

In particular, expression is more pronounced in cultures induced at 30 °C for 4 hours, suggesting that, unlike PIR1r1, PIR2r1 tolerates more aggressive expression conditions without an apparent loss of production. Based on these results, the condition Shuffle T7, TB, 30 °C for 4 hours was selected as a reference for subsequent purification attempts. In the first phase, it was decided to keep the PIR buffer solution already used for GFP-PIR1r1-8 and PIR1r1 (50 mM Tris, 200 mM NaCl, pH 7.0) unchanged, with

the idea of limiting the number of variables and verifying whether PIR2r1 could be purified under similar conditions.

The cultures under the various conditions were collected by centrifugation (see Section 3.3.1) and the cell pellet was resuspended in 2 mL of PIR buffer solution. Lysis was performed using the FastPrep homogenisation system in the presence of glass beads (see Section 3.2.2.1), using screw-cap tubes (Figure 64-B-C). After centrifugation, the supernatant was carefully separated from the beads and purified using Ni-NTA in spin column format (see Section 3.3.2.1). The different fractions were collected for subsequent analysis on SDS-PAGE.

Despite the clear expression of the protein observed in the screening gels, the first purification attempt did not result in the recovery of PIR2r1 in the elution fractions. To analyse this seemingly contradictory result in more detail, the Shuffle T7, 30 °C, 4 hours condition was further investigated by way of example using anti-His Western blotting (Figure 66). Staining of the membrane with Ponceau S confirms the presence of protein material and effective transfer, while chromogenic development of DAB (see Section 3.4.2.1) shows a specific signal for PIR2r1 exclusively in the pellet, load and post-induction sample, but not in the wash or elution fractions. This result indicates that PIR2r1, although expressed, is not effectively recovered by the Ni-NTA resin under the buffer conditions initially adopted.

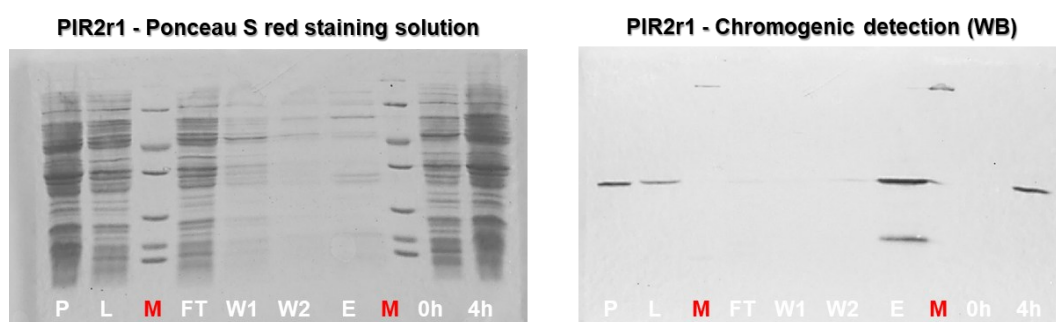


Figure 66. Ponceau S staining confirms uniform protein transfer across all lanes, while anti-His chromogenic Western blot detection shows a specific PIR2r1 signal in the pellet, load and post-induction samples, but not in the wash or elution fractions, indicating lack of recovery by Ni-NTA under the tested conditions.

At this point, the failure of purification was interpreted not as an expression problem, but as an indication of a possible incompatibility between the lysis/purification conditions and the physicochemical properties of the protein. In particular, the pH 7.0 of the buffer used is relatively close to the isoelectric point of PIR2r1, a condition that can promote non-specific interactions, reduce effective solubility, or compromise the accessibility of the His tag to the resin. Furthermore, a moderate ionic strength (200 mM NaCl) may not be sufficient to shield unwanted electrostatic interactions in a protein context rich in low-complexity regions.

Based on these considerations, it was decided to modify the lysis and purification buffer by increasing both the ionic strength and pH to 50 mM Tris, 300 mM NaCl, pH 8.0. PIR2r1 was then re-expressed under the selected conditions (Shuffle T7, TB, 30 °C for 4 hours) and the entire lysis and purification workflow using spin Ni-NTA was repeated using the optimised buffer (see Section 2.3.1). Under these new conditions, SDS-PAGE analysis of the fractions shows the appearance of a band corresponding to PIR2r1 in the elution fractions, indicating effective protein recovery and confirming that buffer choice is a critical factor for the purification of this construct. (Figure 67)

To confirm the presence of PIR2r1 in the elution fractions, the band corresponding to the expected molecular weight for PIR2r1 was manually isolated from the SDS-PAGE gel by cutting with a scalpel (Figure 67, on the right, red box on the elution band) and subjected to qualitative analysis by mass spectrometry. Mass spectrometry analysis allowed the identification of PIR2r1 with 91% sequence coverage, robustly and unambiguously confirming the identity of the protein recovered in elution, validating the effectiveness of the purification strategy adopted after optimisation of the buffer conditions.

The experience gained with PIR2r1 showed that, given good expression, the success of purification critically depends on adapting the buffer conditions to the physicochemical properties of the construct. In particular, the optimisation of pH and ionic strength proved to be decisive for the

recovery of the protein in elution. Based on this evidence, the analysis of PIR4r1 was set up by directly applying the optimised purification conditions, while still maintaining a preliminary screening of the expression conditions.

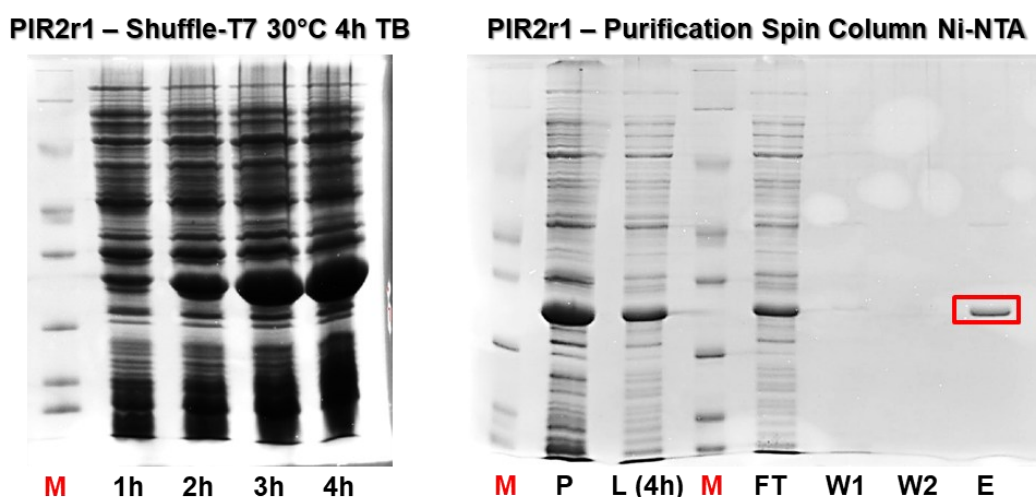


Figure 67. SDS-PAGE analysis of PIR2r1 purification performed with the PIR optimized buffer solution (see Section 2.3.1) shows the appearance of a band in the elution fractions, indicating successful recovery of the target protein. The corresponding band was excised from the gel (right panel, red box) and identified by mass spectrometry, confirming the presence of PIR2r1 in the eluted sample.

4.3.2.4 PIR4r1

PIR4r1 is a single-repeat protein consisting of 220 amino acids, with a theoretical molecular weight of 22572.57 Da and a calculated isoelectric point of 5.30. Compared to PIR2r1, PIR4r1 is of comparable size but has a slightly more acidic pI, suggesting potentially similar behaviour during lysis and purification, particularly in relation to pH and ionic strength conditions of the buffer. Similar to PIR2r1, the PIR4r1 plasmid (cloned in *E. coli* DH5α cells) was initially screened for expression conditions (Figure 68) using *E. coli* Shuffle T7 and BL21 (DE3) Gold strains, LB and TB growth media, and two induction regimes: 30 °C for 4 hours and 21 °C overnight. SDS-PAGE analysis of the induced samples shows that PIR4r1 is clearly expressed under conditions of 30 °C for 4 hours in Shuffle T7, regardless of the medium used, while expression is lower overall in BL21(DE3) Gold. Considering the

lower temperature conditions, expression in Shuffle T7 grown in TB at 21 °C overnight showed a more intense band than the other conditions, suggesting higher production or better protein stability under these conditions. (Figure 68)

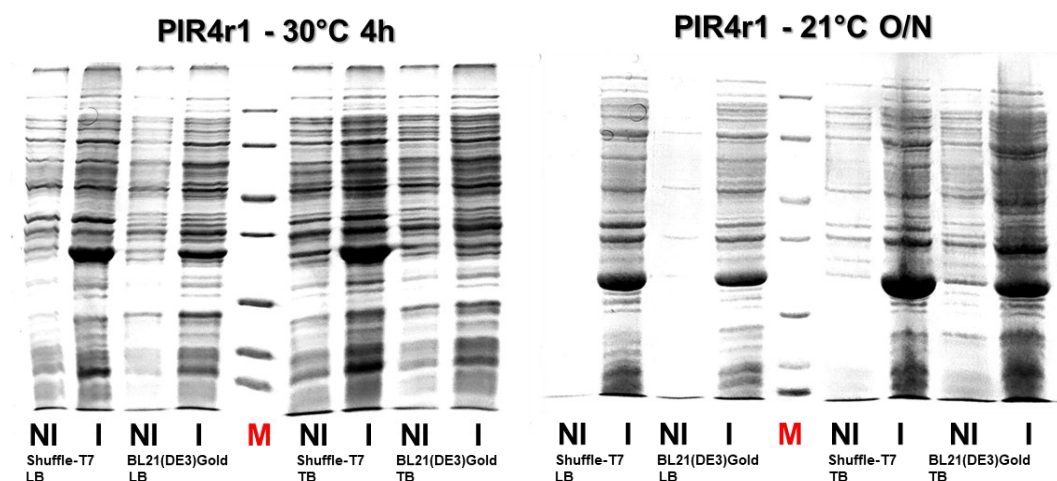


Figure 68. SDS-PAGE analysis of PIR4r1 expression testing under different induction conditions, showing non-induced (NI) and IPTG-induced (I) samples expressed in different *E. coli* strains and culture media, comparing induction at 30 °C for 4 hours and at 21 °C overnight. M, molecular weight marker.

Because of these observations, and in line with what was learned for PIR2r1, PIR4r1 was directly re-expressed in *E. coli* Shuffle T7 in TB medium, inducing expression at 21 °C overnight, and purified using the previously optimised buffer (50 mM Tris, 300 mM NaCl, pH 8.0). The cells were collected by centrifugation, resuspended in the same buffer and subjected to mechanical homogenisation using the FastPrep® homogenisation system in the presence of glass beads, followed by centrifugation to separate the supernatant. The clarified supernatant was then purified using Ni-NTA in spin column format. Monitoring of the different fractions by SDS-PAGE shows the presence of a band at the expected molecular weight for PIR4r1 in the elution fractions; however, this band is of reduced intensity and close to the detection limit in the acquired images. This observation was interpreted as a limitation of the sensitivity of the acquisition system rather than an actual absence of the protein, also

considering the consistency of the signal between the different replicates and subsequent evidence of identification.

Similarly to what was done for PIR2r1, for PIR4r1 the region of the SDS-PAGE gel corresponding to the expected molecular weight was isolated manually by cutting with a scalpel (Figure 69, red box) and subjected to qualitative analysis by mass spectrometry. In this case, qualitative analysis by mass spectrometry of the peptides extracted from the gel region corresponding to the expected molecular weight confirmed the identification of PIR4r1, with 76% sequence coverage. Although lower than that obtained for PIR2r1, this value is consistent with the lower intensity of the band observed in SDS-PAGE and with the presence of low-complexity regions in the PIR sequence, which can reduce peptide sampling efficiency. Overall, the coverage obtained is fully compatible with a reliable identification of the eluted protein as PIR4r1.

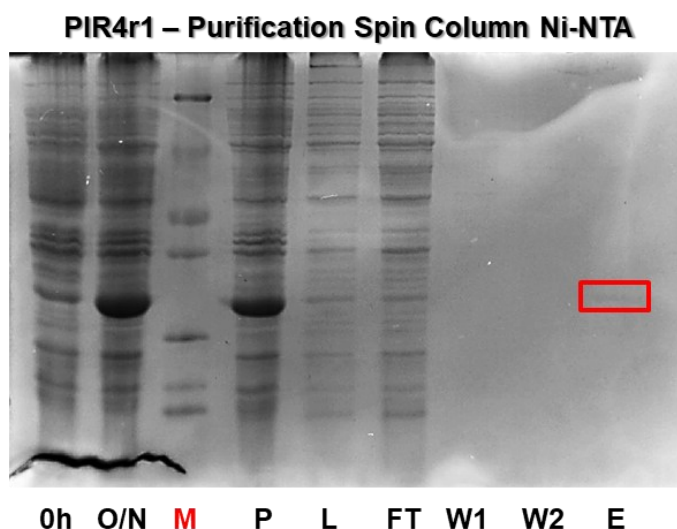


Figure 69. SDS-PAGE analysis of PIR4r1 purification performed using the PIR optimised buffer (50 mM Tris, 300 mM NaCl, pH 8.0) shows a faint band in the elution fractions, close to the detection limit. The corresponding gel region was excised (red box) and analysed by mass spectrometry, confirming the presence of PIR4r1 in the eluted sample.

Overall, the results obtained for PIR4r1 indicate that, despite limited yield and weak signals in SDS-PAGE, the application of optimised buffer

conditions allows the recovery of the protein in purified form, confirming the effectiveness of the rational approach derived from the analysis of PIR2r1.

4.3.2.5 PIR5r1

PIR5r1 is a single-repeat protein composed of 228 amino acids, with a theoretical molecular weight of 22572.57 Da and a calculated isoelectric point of 8.68. Compared to PIR2r1 and PIR4r1, PIR5r1 is therefore more basic, suggesting potentially different behaviour during the lysis and purification phases, particularly in relation to the pH and ionic strength of the PIR optimised buffer (see Section 2.3.1).

Since the results obtained for PIR2r1 and PIR4r1, the expression conditions for the PIR5r1 plasmid (duplicated by miniprep from *E. coli* DH5α cells) were screened using the same experimental approach. Expression was tested in *E. coli* Shuffle T7 and BL21(DE3) Gold, using LB and TB growth media and comparing two induction regimes: 30 °C for 4 hours and 21 °C overnight. (Figure 70)

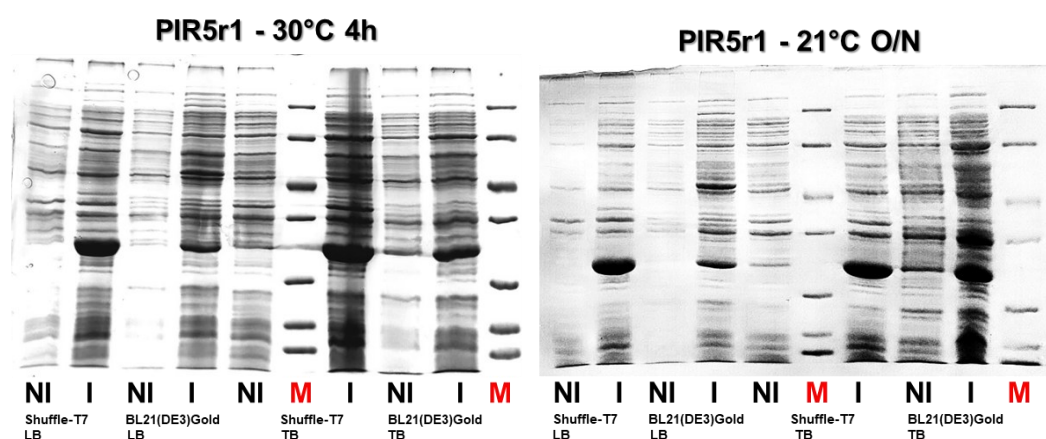


Figure 70. SDS-PAGE analysis of PIR5r1 expression testing under different induction conditions, showing non-induced (NI) and IPTG-induced (I) samples expressed in different *E. coli* strains and culture media, comparing induction at 30 °C for 4 hours and at 21 °C overnight. M, molecular weight marker.

Based on these observations, and considering the experience gained with PIR2r1 and PIR4r1, it was decided to proceed directly with purification using

the optimised high ionic strength buffer (50 mM Tris, 300 mM NaCl, pH 8.0) in order to assess whether the increase in salt concentration could favour protein recovery, particularly in view of the relatively high pI of the construct. The cells were collected by centrifugation, resuspended in the same buffer and subjected to mechanical homogenisation using the FastPrep® homogenisation system in the presence of glass beads, followed by centrifugation to separate the supernatant.

The clarified supernatant was then purified using Ni-NTA in spin column format. SDS-PAGE analysis of the collected fractions shows the presence of a band at the expected molecular weight for PIR5r1 in the elution fractions. (Figure 71)

The signal intensity is lower than that observed for PIR2r1, but clearly higher than that obtained for PIR4r1, indicating a more efficient protein recovery than the latter, although not optimal.

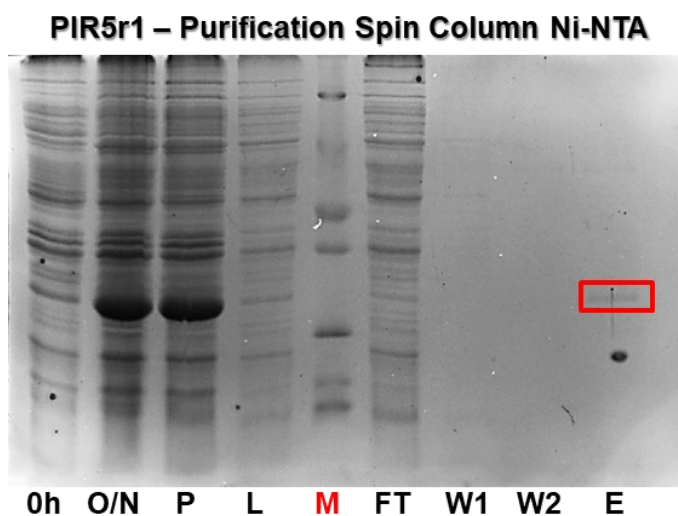


Figure 71. SDS-PAGE analysis of PIR5r1 purification performed using the optimised buffer (50 mM Tris, 300 mM NaCl, pH 8.0) shows a distinct band in the elution fractions, close to the detection limit. The corresponding gel region was excised (red box) and analysed by mass spectrometry, confirming the presence of PIR5r1 in the eluted sample.

To confirm the identity of the recovered protein, the band corresponding to the expected molecular weight was manually isolated from the SDS-PAGE gel and subjected to qualitative analysis by mass spectrometry. Peptide

identification confirmed the presence of PIR5r1 in the eluted sample, with 63% sequence coverage. Although lower than the values obtained for PIR2r1 and PIR4r1, this coverage is consistent with the lower intensity of the band and with the characteristics of the PIR sequence, which includes regions of low complexity that are potentially less favourable for peptide sampling. Overall, the results obtained confirm the recovery of PIR5r1 in purified form and support the effectiveness of the approach based on optimising the ionic strength of the PIR optimized buffer solution.

4.3.2.6 PIR sequencing

To further prove the identity of the experimentally analysed PIR constructs, the plasmids encoding PIR2r1, PIR4r1 and PIR5r1 were sequenced. This analysis was designed as a quality control following the cloning, expression and purification phases, with the aim of confirming full correspondence between the sequences actually expressed and those predicted during the design phase of the constructs.

Sequencing was performed on plasmid DNA obtained by miniprep from *E. coli* DH5 α cells, prepared as described in Section 3.1.3, and entrusted to a specialised external service. Analysis of the sequences obtained confirmed complete correspondence with the expected constructs, with no evidence of point mutations, insertions, deletions or reading frame alterations that could affect the length, composition or properties of the expressed proteins. This check is particularly relevant considering the differences observed in the apparent molecular weight of the PIR2r1, PIR4r1 and PIR5r1 proteins in SDS-PAGE gels compared to the calculated theoretical values. Sequencing allows us to unequivocally rule out that these discrepancies are due to cloning errors or unexpected variations in the primary sequence, supporting the interpretation that the electrophoretic behaviours observed reflect intrinsic characteristics of the proteins, such as amino acid composition, the presence of low-complexity regions, or abnormal behaviour under denaturing conditions.

It is also important to note that this verification was carried out at a stage of the project when specific antibodies against PIR2r1, PIR4r1 and PIR5r1 were not yet available, making direct immunological checks on the identity of the expressed proteins inaccessible. In this context, the sequencing of plasmid constructs was the most reliable tool for ruling out upstream errors in plasmid construction and providing preliminary validation of the results obtained in the expression and purification phases. More specific verifications based on dedicated antibodies were subsequently carried out once these reagents became available, as described in the following sections.

Taken together, the sequencing results provide independent and complementary confirmation of the identity of the PIR constructs analysed, reinforcing the evidence obtained by qualitative mass spectrometry and ensuring that the experimental data reported in the previous sections refer to the correct protein sequences. This downstream quality control is an essential step for the overall reliability of the biochemical analysis of PIR proteins presented in this work.

4.3.3 Large-scale expression and purification

4.3.3.1 PIR2r1

Once the small-scale expression and purification conditions had been optimised, large-scale production of PIR proteins was carried out, starting from the construct that had shown the most robust and reproducible behaviour, namely PIR2r1. The aim of this phase was to obtain enough purified protein for subsequent biochemical, structural and immunological analyses.

PIR2r1 expression was carried out in *E. coli* Shuffle T7 using TB medium, inducing expression at 30 °C for 4 hours, as described in Section 3.2.1.3. The cultures were progressively scaled up from 2 L to a total volume of 12 L, maintaining the growth and induction conditions identified during the screening phase. At the end of expression, the cells were harvested by centrifugation as described in Section 3.3.1.

The cell pellet obtained was resuspended in PIR optimised buffer solution (see Section 3.2.2) and the various stages of collection, pellet resuspension and cell suspension preparation were documented photographically (Figure 72) to support the reproducibility of the process on a large scale. Cell lysis was performed using high-pressure systems, specifically French Press (see Section 3.2.2.3) or Avestin® Emulsiflex-C5 System (see Section 3.2.2.4), depending on the availability of the equipment. Lysis was followed by a centrifugation step to remove cell debris and clarification of the supernatant by filtration through 0.45 µm membranes (see Section 3.3.1), obtaining a clear protein extract suitable for chromatographic purification. (Figure 72)

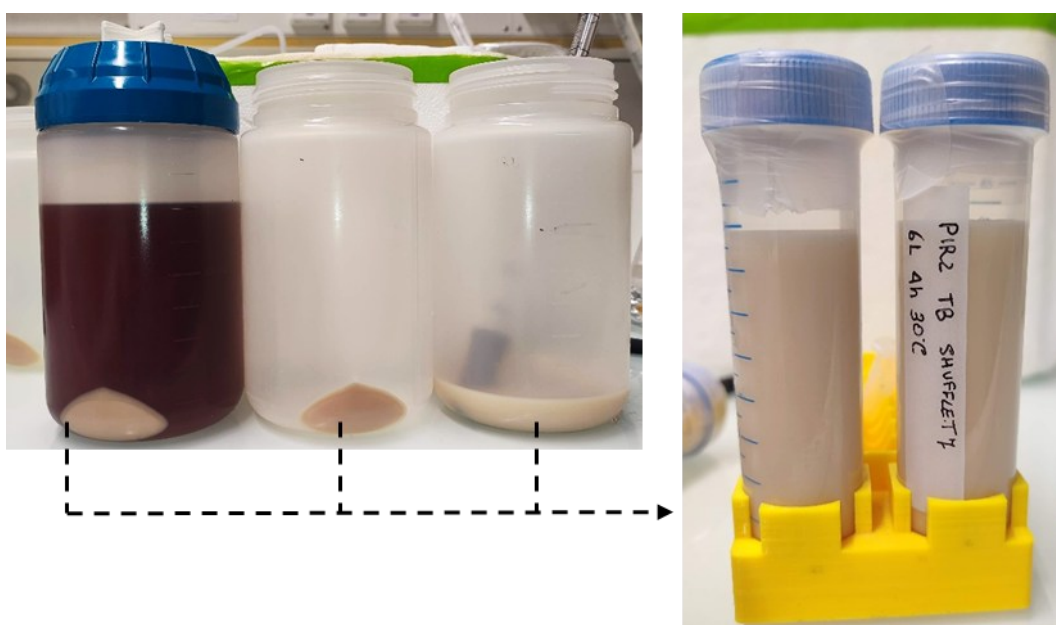


Figure 72. Photographic documentation of large-scale PIR2r1 production, showing harvested cell pellets, resuspension in PIR optimised buffer solution and preparation of cell suspensions prior to high-pressure lysis and downstream purification.

Large-scale purification of PIR2r1 was performed by affinity chromatography on a 5 mL Ni-NTA Protino column, following the procedure described in Section 3.3.2.2. During purification, representative aliquots were collected from each phase (loading, washing and elution) and subsequently analysed by SDS-PAGE. As shown in Figure 73, the eluate

shows an intense band at the expected molecular weight for PIR2r1, indicating efficient protein recovery under large-scale conditions.

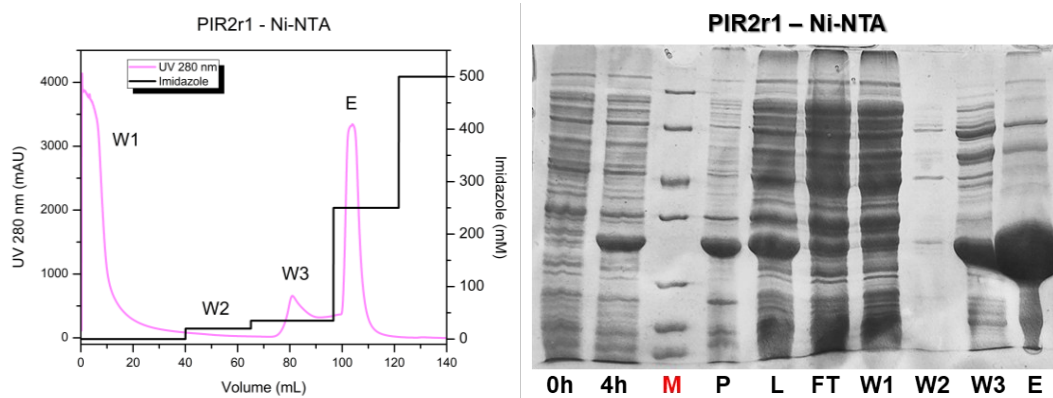


Figure 73. Affinity purification of PIR2r1 on a 5 mL Ni-NTA column under large-scale conditions. The chromatographic profile (left) shows the washing and elution steps, while the corresponding SDS-PAGE analysis (right) reports representative samples collected throughout expression and purification. Lanes 0 h and 4 h correspond to cell aliquots collected before induction and after 4 h of IPTG induction, respectively; P, insoluble pellet; L, clarified supernatant; FT, flow-through; W1, wash with PIR optimized buffer solution only; W2, wash with PIR optimized buffer solution supplemented with 20 mM imidazole; W3, wash with PIR optimized buffer solution supplemented with 35 mM imidazole; E, elution with Ni-NTA elution buffer (250 mM imidazole at pH 8.0, see Section 2.3.1). The elution fraction shows a very intense band for PIR2r1, indicating high protein recovery, despite loading only 3 μ L of eluted sample mixed with 1 $\times$ SDS buffer.

The eluate was then concentrated using Centricon filters with a cutoff of 10 kDa and subjected to size exclusion chromatography (SEC), as described in Section 3.3.4. The chromatographic profile obtained shows the presence of two main elution peaks (Figure 74), suggesting the coexistence of protein populations with different oligomeric states or degrees of purity. The first peak, eluting at lower volumes, has a broader and more asymmetrical shape, suggesting the presence of high molecular weight species or heterogeneous populations, potentially attributable to aggregates or oligomers. The second peak, centred around ~60 mL of elution volume, is narrower and more symmetrical, indicative of a more homogeneous protein population. The fractions corresponding to the two peaks were analysed by

SDS-PAGE, and in particular, the fractions associated with the first peak show the presence of additional bands with higher and lower molecular weights than expected for PIR2r1, suggesting lower purity or co-elution of contaminating species. In contrast, the fractions corresponding to the second peak, in particular fractions 29-31, show a single prevailing band at the molecular weight expected for PIR2r1, with minimal background and no obvious contaminants.

Based on these observations, fractions 29-31 were selected as the most suitable for subsequent characterisation analyses, including Western blotting experiments and crystallisation assays. This choice reflects a selection criterion based not exclusively on UV signal intensity, but on the integration between chromatographic profile and electrophoretic purity, favouring the most homogeneous and structurally defined protein population.

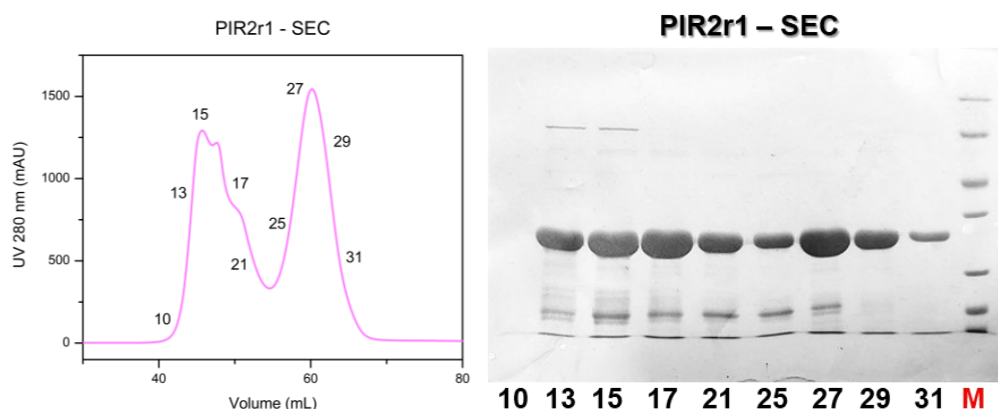


Figure 74. Size-exclusion chromatography profile of concentrated PIR2r1 eluate (left) and SDS-PAGE analysis of selected SEC fractions (right). The chromatogram shows two main elution peaks, with an early, broader peak indicative of heterogeneous or higher-molecular weight species and a later, narrower peak corresponding to a more homogeneous PIR2r1 population. SDS-PAGE analysis confirms the presence of additional bands in fractions associated with the first peak, whereas fractions 29-31 from the second peak display a predominant band at the expected molecular weight for PIR2r1 with minimal background, supporting their selection for downstream analyses.

The final purified fractions were used for subsequent crystallisation experiments and Western blotting analyses, described in the following sections. Overall, this phase demonstrates that the small-scale optimised expression and purification workflow for PIR2r1 is fully transferable to culture volumes in the order of litres, allowing the production of significant quantities of purified protein with a high degree of reproducibility.

4.3.3.2 PIR4r1

For PIR4r1, once the optimal expression condition in *E. coli* Shuffle T7 in TB medium at 21 °C overnight had been identified, large-scale production was carried out, followed by purification using Ni-NTA affinity chromatography and subsequent size exclusion chromatography (SEC), according to the same workflow adopted for PIR2r1. SDS-PAGE analysis of the elution fractions from Ni-NTA (Figure 75) shows a substantial amount of recovered protein, although accompanied by a significant protein background, making an additional purification step necessary to increase the homogeneity of the sample.

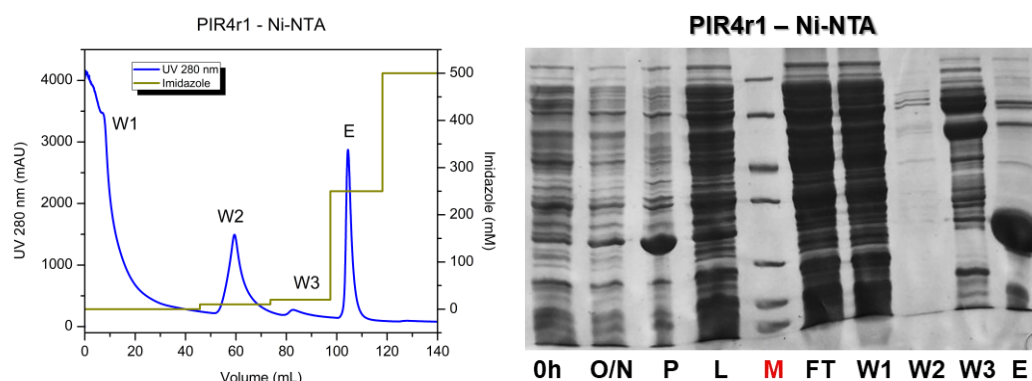


Figure 75. Affinity purification of PIR4r1 on a 5 mL Ni-NTA column under large-scale conditions. The chromatographic profile (left) shows the washing and elution steps, while the corresponding SDS-PAGE analysis (right) reports aliquotes collected throughout expression and purification. Lanes 0 h and O/N correspond to cell aliquots collected before induction and after overnight of IPTG induction, respectively; P, insoluble pellet; L, clarified supernatant; FT, flow-through; W1, wash with PIR optimized buffer solution only; W2, wash with PIR optimized buffer solution supplemented with 20 mM imidazole; W3, wash with PIR optimized buffer solution supplemented with 35 mM imidazole; E, elution with 250 mM

imidazole at pH 8.0. The elution fraction shows an intense band for PIR4r1, indicating high protein recovery, despite loading only 3 μ L of eluted sample mixed with 1 $\times$ SDS buffer.

Similarly, the eluate was concentrated using Centricon filters with a 10 kDa cutoff and subjected to size exclusion chromatography (SEC), as described in Section 3.3.4. The SEC profile of PIR4r1 (Figure 76) shows the presence of a main peak centred around ~60 mL of elution volume, accompanied by secondary signals at lower volumes. Similar to what was observed for PIR2r1, the peaks eluting at lower volumes are compatible with the presence of high molecular weight species or heterogeneous populations, potentially attributable to aggregates or oligomeric forms. The main peak, which is narrower and more symmetrical, suggests the presence of a predominant and more homogeneous protein population. SDS-PAGE analysis of the fractions collected along the SEC profile confirms this interpretation, showing that the fractions corresponding to the main peak contain a predominant band at the expected molecular weight for PIR4r1, with a significant reduction in the non-specific bands observed after affinity purification. (Figure 76)

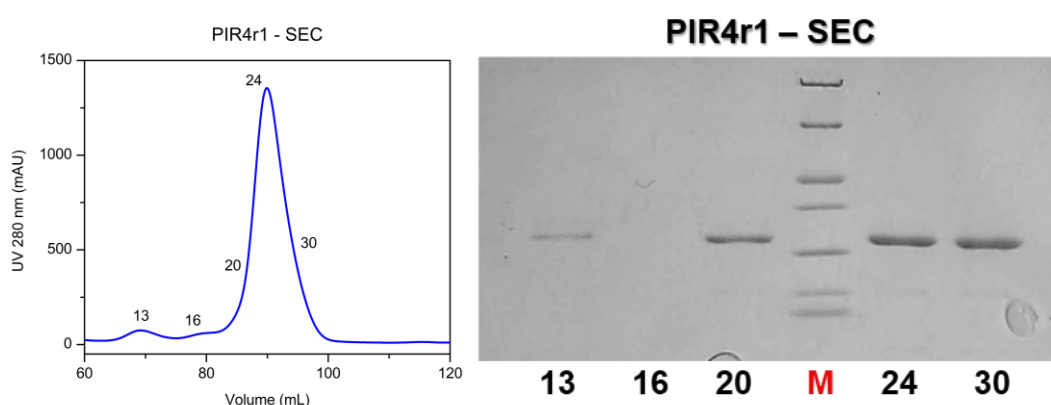


Figure 76. Size-exclusion chromatography profile of concentrated PIR4r1 eluate (left) and SDS-PAGE analysis of selected SEC fractions (right). The chromatogram shows a dominant, symmetrical peak corresponding to a homogeneous PIR4r1 population, together with minor early-eluting signals compatible with higher-molecular weight species. SDS-PAGE analysis confirms that fractions associated with the main peak contain a predominant band at the expected molecular weight for PIR4r1, with reduced background compared to the affinity-purified sample.

4.3.3.3 PIR5r1

Based on the results obtained during the preliminary screening phase, PIR5r1 was produced on a large scale using the Shuffle T7 strain in TB medium, with induction at 21 °C overnight, a condition selected because it was associated with a cleaner overall expression profile than the alternatives tested, while showing comparable expression levels even at 30 °C for 4 h.

In view of its chemical-physical characteristics, and considering the results obtained with PIR2r1 and PIR4r1, it was initially decided to maintain the optimised high ionic strength purification buffer (50 mM Tris, 300 mM NaCl, pH 8) to reduce non-specific electrostatic interactions and potential aggregation or precipitation phenomena during the purification stages.

Cell collection, pellet resuspension based on OD600, cell lysis (French press or Avestin® Emulsiflex-C5 System), and lysate clarification were performed according to the standard protocols described in Sections 3.2.2 and 3.3.1. The clarified supernatant was then subjected to Ni-NTA affinity chromatography, collecting representative aliquots from each step for analysis by SDS-PAGE.

The Ni-NTA chromatographic profile of PIR5r1 shows a defined elution peak at the increase in imidazole concentration, indicative of effective binding to the matrix (Figure 77). SDS-PAGE analysis of the elution fractions reveals the presence of a band at the expected molecular weight, with an intensity intermediate between that observed for PIR2r1 (more intense) and PIR4r1 (weaker). This result suggests an overall good yield, although lower than that obtained for PIR2r1.

During large-scale production, however, episodes of variability in protein solubility were observed. In some preparations, PIR5r1 was predominantly soluble and recoverable in Ni-NTA elution, while in others the protein tended to remain largely in the pellet already in the early stages of purification. This variability is likely attributable to a combination of factors, including the proximity of the purification pH to the isoelectric point of the protein and a

possible sensitivity of PIR5r1 to misfolding or aggregation phenomena during large-scale expression. Nevertheless, when the protein was soluble, the final yield was still satisfactory and compatible with subsequent structural applications.

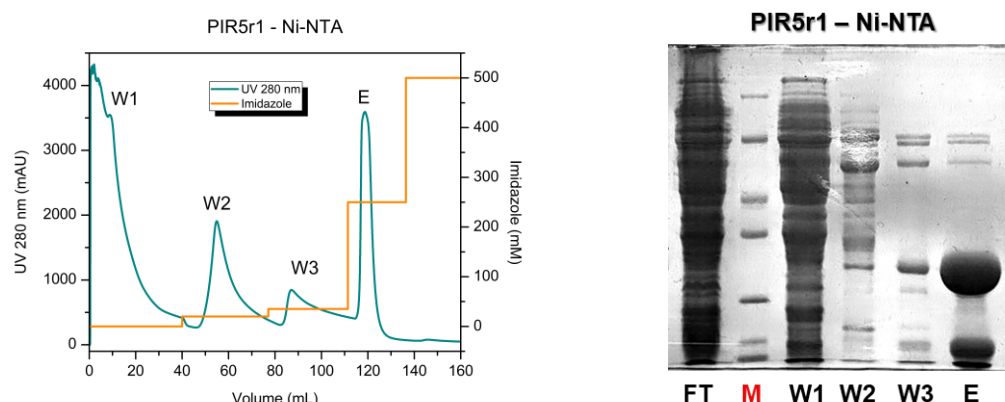


Figure 77. Affinity purification of PIR5r1 on a 5 mL Ni-NTA column under large-scale conditions. The chromatographic profile (left) shows the washing and elution steps, while the corresponding SDS-PAGE analysis (right) reports representative samples collected throughout expression and purification. Lanes FT corresponds to flow-through; W1, wash with PIR optimized buffer solution only; W2, wash with PIR optimized buffer solution supplemented with 20 mM imidazole; W3, wash with PIR optimized buffer solution supplemented with 35 mM imidazole; E, elution with 250 mM imidazole at pH 8.0. The elution fraction shows an intense band for PIR5r1, indicating high protein recovery, despite loading only 3 μ L of eluted sample mixed with 1 $\times$ SDS buffer.

The Ni-NTA eluate was then concentrated and subjected to size exclusion chromatography (SEC). The SEC chromatogram of PIR5r1 shows a single, well-defined main peak, consistent with a predominantly monomeric protein population (Figure 78). Although it was not possible to recover the SDS-PAGE gel image for the SEC fractions, the chromatographic profile and subsequent analyses indicate a good level of sample homogeneity.

Overall, the results obtained indicate that PIR5r1 can be produced and purified in adequate quantities, even if with greater sensitivity to expression and purification conditions than the other PIR proteins analysed. This observation highlights how, within the same family of single-repeat modules, even moderate differences in sequence properties, such as the isoelectric

point, can translate into distinct biochemical behaviours during large-scale production.

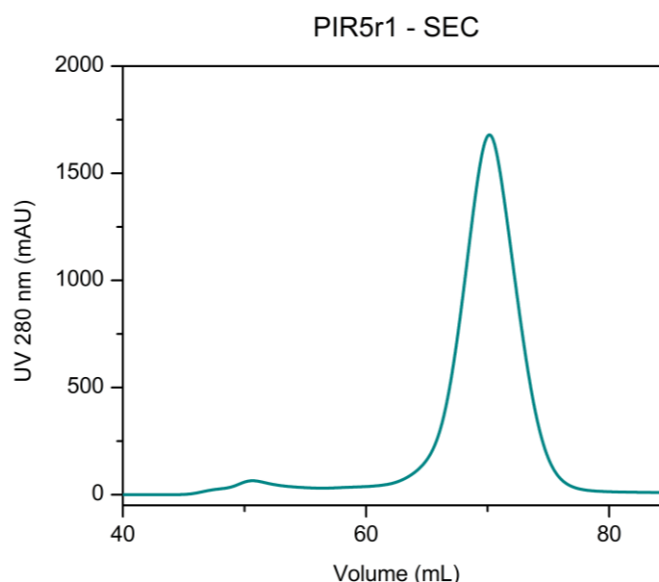


Figure 78. Size-exclusion chromatography profile of concentrated PIR5r1 Ni-NTA eluate, showing a single, well-defined main peak consistent with a predominantly monomeric and homogeneous protein population, despite variability observed during large-scale expression and solubility.

4.3.4 Western blot: immunological validation of single-repeat constructs

The availability of anti-PIR antibodies allowed for immunological validation of the purified single-repeat constructs in this study. In a preliminary phase, the specificity of anti-PIR2 antibodies (see Section 2.3.5) was evaluated by colleague Amon Victor Constantin Nemeth in the Genetics Laboratory of Prof. H.-U. Mösch (Dept. of Biology, Philipps-Univ. Marburg, Germany) using Western blot on cell extracts from *S. cerevisiae* wild-type and strains with combinatorial deletions of PIR genes (Figure 79). This analysis showed that the antibodies reliably recognise PIR2r1 and show reactivity towards PIR1r1 and PIR4r1, while they are not suitable for the detection of PIR5r1. Based on this evidence, Western blot analysis of recombinant samples was

limited to PIR1r1, PIR2r1 and PIR4r1, excluding PIR5r1 as it was not informative with the available antibody. (Figure 79)

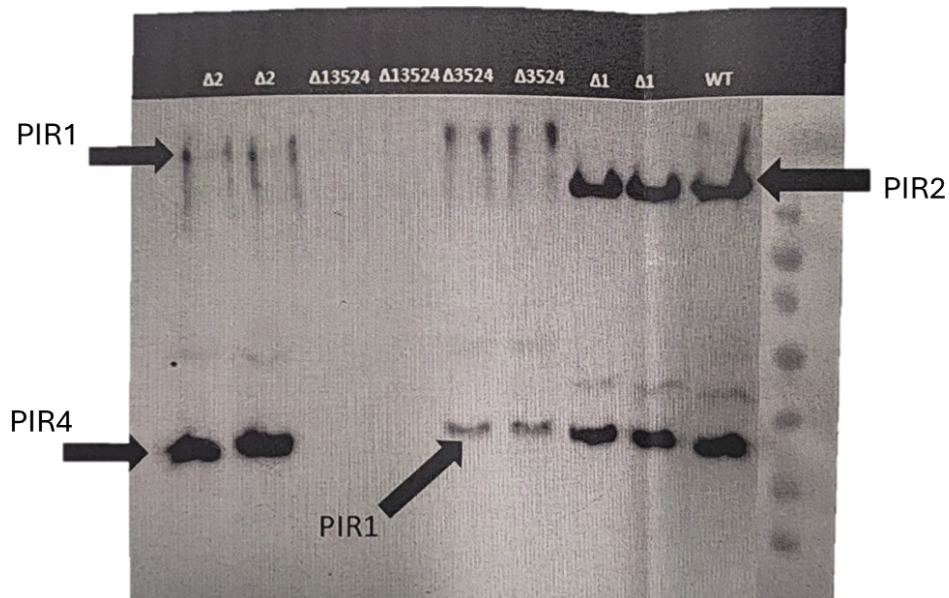


Figure 79. Western blot analysis performed by Amon Victor Constantin Nemeth (Genetics Laboratory of Prof. H.-U. Mösch, Dept. of Biology, Philipps-Universität Marburg) on *S. cerevisiae* cell extracts using anti-PIR2 antibodies. Samples include wild-type (WT) cells and strains carrying defined deletions of PIR genes: Δ2 (PIR2 deletion), Δ1 (PIR1 deletion), and combinatorial deletions Δ13524 and Δ3524 affecting PIR1, PIR3, PIR5 and PIR2/PIR4, as indicated above the lanes. A strong PIR2-specific signal is detected in WT and Δ1 strains and is absent in Δ2 backgrounds, confirming antibody specificity for PIR2. Cross-reactive signals corresponding to PIR1 and PIR4 are observed in strains retaining these genes. No detectable signal attributable to PIR5 is observed in any deletion background, indicating that the anti-PIR2 antibody does not recognise PIR5.

For each protein, only fractions obtained by size exclusion chromatography (SEC) showing the highest degree of purity were used. Protein concentrations were determined using Nanodrop (see Section 3.4.4) and, referring to the sensitivity levels observed in the collaborating laboratory, dilutions were prepared so as to load extremely low amounts of protein per well. In all cases, 10 μL per well was loaded for each sample, corresponding to protein masses of 5 ng, 2.5 ng, 0.5 ng and 0.25 ng per lane.

To obtain these quantities, progressive dilutions were made from the concentrated stocks: for the 5 ng and 2.5 ng per well loads, dilutions between 1:50 and 1:250 were used, while for the 0.5 ng and 0.25 ng loads, an additional dilution factor (up to 1:1000) was applied, bringing the overall dilution factor to values in the order of 10^4 - 10^5 . This approach allowed to rigorously check the sensitivity of the antibody and the reliability of immunological recognition even at extremely low loads.

Overall, Western blots show that all three recombinant proteins analysed are effectively recognised by anti-PIR2 antibodies, despite PIR2r1 and PIR4r1 showing an apparent molecular weight in SDS-PAGE higher than the theoretical one.

This result is particularly relevant as it indicates that immunological recognition is not compromised by electrophoretic migration abnormalities, confirming the identity of the purified constructs as correctly obtained monorepeat PIR proteins.

In the case of PIR1r1, immunological recognition is evident even at the lowest concentrations (0.5-0.25 ng per well). At higher concentrations, however, additional bands with lower molecular weights are observed, consistent with the presence of fragmented species or degradation products. This behaviour suggests a possible greater susceptibility of PIR1r1 to fragmentation or proteolysis processes, or the recognition by the polyclonal antiserum of epitopes shared with stable fragments of the protein. In any case, the presence of a clear and reproducible main signal confirms the immunological identity of the construct. (Figure 80)

In the case of PIR2r1, the signal is intense and well defined under all loading conditions tested, from 5 ng to 0.25 ng per well. This behaviour indicates very high antibody reactivity and robust exposure of the recognised epitope, in full agreement with the data obtained on cell extracts in the collaborating laboratory, where PIR2 represents the most evident and specific band.

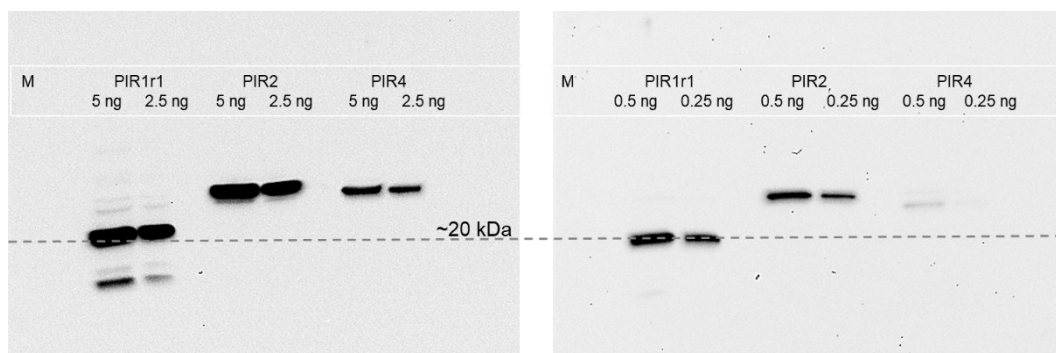


Figure 80. Western blot analysis of purified recombinant PIR1r1, PIR2r1 and PIR4r1 using anti-PIR2 antibodies, performed with decreasing protein amounts (5-2.5 ng per lane on the left panel; 0.5-0.25 ng per lane on the right panel). All three constructs are specifically recognised despite the anomalous electrophoretic migration of PIR2r1 and PIR4r1. PIR1r1 shows clear immunoreactivity down to the lowest amounts tested, with additional lower-molecular weight bands appearing at higher loads, consistent with fragmentation or partial proteolysis. PIR2r1 displays a strong and well-defined signal across all concentrations, indicating high antibody reactivity. PIR4r1 is readily detected at higher loads, while the signal decreases markedly at lower amounts, approaching the detection limit at 0.25 ng.

For PIR4r1, the signal is clearly visible at higher concentrations (5 ng and 2.5 ng per well), while it decreases significantly at 0.5 ng and is almost at the detection limit at 0.25 ng. This trend is consistent with what was observed in the specificity control on yeast extracts, where PIR4r1 is recognised by antibodies but with lower intensity than PIR2r1. The reduction in signal at lower concentrations likely reflects a lower abundance or accessibility of the epitopes recognised by the polyclonal antiserum. (Figure 80)

In accordance with the antibody specificity data (Figure 79), PIR5r1 was not analysed by Western blot, as the available anti-PIR2 antibodies are not suitable for its detection. For this protein, identity validation was therefore entrusted to a combination of qualitative mass spectrometry, chromatographic behaviour and consistency with the expected molecular weight.

Taken together, the results of expression, purification, immunological validation and chromatographic analysis provide a consistent picture of the biochemical properties of the single-repeat PIR constructs analysed.

Selective recognition of the PIR1r1, PIR2r1 and PIR4r1 constructs by Western blot provided independent confirmation of the identity of the purified proteins, while the absence of reactivity towards PIR5r1 highlighted intrinsic limitations of the available immunological tools, which were appropriately compensated for by alternative validation approaches. The selection of the purest SEC fractions, supported by these controls, made it possible to obtain sufficiently homogeneous and characterised samples for the start of crystallisation experiments. In parallel, the availability of a well-defined and validated single-repeat construct provided the rational basis for computational analysis using molecular dynamics simulations, aimed at exploring the intrinsic structural and dynamic properties of PIR modules in the absence of constraints imposed by the native multi-repeat organisation.

4.3.5 Crystallisation

4.3.5.1 PIR1r1 and rationale for computational analysis

The PIR1r1 protein was not crystallised during the thesis work, and determining its structure was not initially part of the experimental objectives of the assigned research project. However, during the research period carried out at the Structural Biochemistry Laboratory of the host institution, a crystallographic structure of PIR1r1 was solved, which subsequently proved to be of great relevance for the interpretation and development of the computational analyses presented in this work. The crystallisation of the protein was carried out by my colleague, PhD student Maisa Alkheder Alahmad, while the structural refinement steps were handled by Dr Po-Hsun Wang (a post-doc with experience in crystallography and structural refinement), both members of the same lab group led by Prof. Lars-Oliver Essen.

The availability of this crystallographic structure made it possible to directly analyse the 3D organisation of a single PIR repeat module in a well-defined structural context. In particular, the structure provided a fundamental experimental reference for evaluating the role of the protein context in the structural and functional support of the PIR repeat, offering a solid basis for

the structural and functional considerations discussed in the following paragraphs (Figure 81-A).

A structural element of particular interest is the portion of the protein that does not correspond to the repetitive PIR sequence. At first glance, the 3D observation of this region (Figure 81-B) could recall an organisation associated with the phenomenon of domain exchange. However, the crystallographic structure contains a single polypeptide chain in the asymmetric unit and shows neither intermolecular exchange of structural elements nor cis/trans alternation between monomeric and oligomeric states. Consequently, the observed architecture cannot be interpreted as a case of domain exchange in the classical sense.

What makes this structural organisation particularly interesting is the fact that it exhibits an approximately symmetrical organisation, giving rise to a compact architecture whose final structure cannot be immediately deduced from the linear sequence alone.

The topology emerging from Figure 81-B therefore reveals a dense intramolecular organisation characterized by ternary pseudosymmetry. The secondary structure elements are organized into three subdomains, or structural lobes, which converge toward a central core stabilized by disulfide bridges. Together, these lobes define an overall architecture that does not exhibit perfect C3 symmetry but consists of three structurally equivalent regions arranged around a central core, giving rise to an overall trefoil-like topology consistent with models of pseudosymmetry described for other protein architectures. [199,200]

In this context, the PIR motif should not be understood as a structurally isolated element, but rather as functionally integrated into a more complex structure, which provides support with its non-repetitive regions to the linker containing repetitions. Therefore, this organisation suggests that the functional properties associated with the PIR motif depend not only on the repetitive sequence itself, but also on the structural context in which it is integrated. This is important in light of the design hypothesis underlying the single-repeat constructs analysed in this work, according to which a single

PIR unit may be sufficient to preserve key functional properties, provided that it is incorporated into an appropriate structural scaffold.

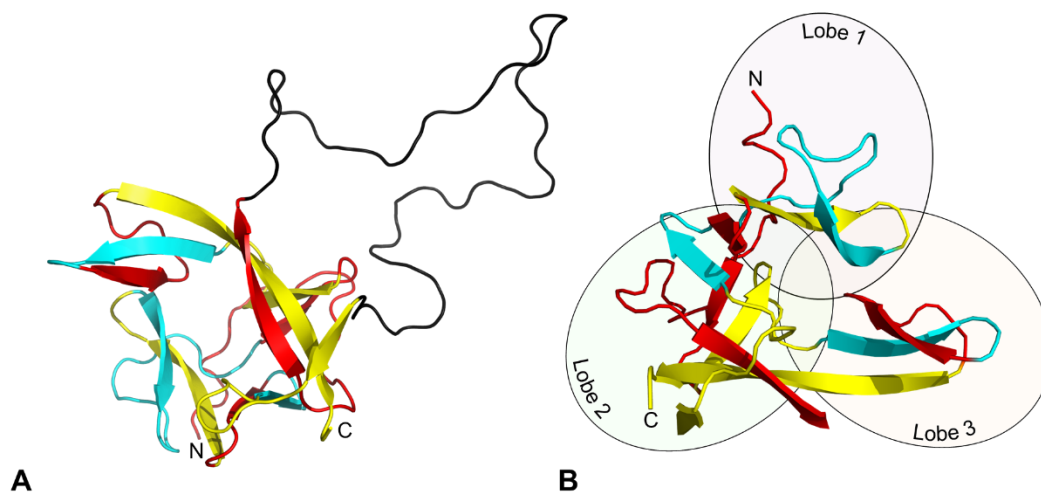


Figure 81. 3D representation of the crystallographic structure of PIR1r1 highlighted according to its internal topological organisation. On the left (A) is the complete protein, with the connecting region containing the PIR repeat (black), while on the right (B) is a detail of the “trefoil” topology domain, obtained by emphasising the three main structural lobes (Lobe 1-3), indicated by semi-transparent ellipses. The secondary structure elements are manually coloured according to their order along the primary sequence (red, yellow and cyan), highlighting how consecutive segments in the sequence contribute to different lobes and how each lobe is made up of non-contiguous regions of the polypeptide chain. Statistical information and the PDB code for the crystallographic structure of PIR1r1 are not reported since the structure has not yet been deposited.

A further element of interest, in line with the functional interpretation of PIR proteins, derives from experiments conducted in the host laboratory in which PIR1r1 crystals were soaked in the presence of glucose.

Under these conditions, a crystallographic structure was obtained in which a glucose molecule is clearly located near the PIR motif, in a surface region adjacent to the repeat. The binding region does not correspond to a rigid, preformed pocket on the structure, but rather the binding is more superficial. This type of interaction is consistent with ligand incorporation by soaking, which typically highlights accessible regions of the protein compatible with the crystal network. From a functional perspective, however, this superficial

binding mode is particularly relevant, as it supports the interpretation of PIR proteins as cell wall components involved in interactions with polysaccharides. The observed localisation suggests that carbohydrate recognition may be mediated by structurally organised yet conformationally adaptable surfaces, rather than by rigid and highly specific binding pockets. This observation is consistent with the low complexity and intrinsic flexibility of the PIR motif, indicating that these structural features may be functional for carbohydrate recognition rather than representing elements of structural disorder. Although the glucose structure (Figure 82) is not directly included among the systems used for the molecular dynamics simulations presented in this thesis, it provides independent experimental evidence supporting the functional interpretation of the PIR motif.

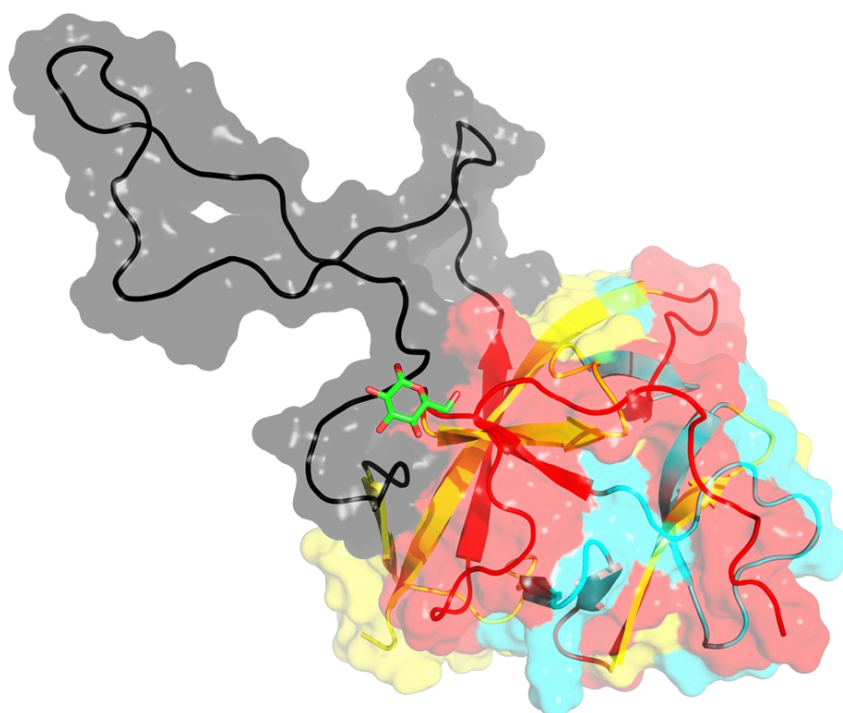


Figure 82. Structure of PIR1r1 in complex with glucose. The molecular surface of the structured domain is coloured according to the three lobes that define the trefoil-like architecture (red, yellow and cyan), highlighting the pseudo-symmetrical organisation of the fold. The glucose molecule is shown as a stick model (green) and is located near a surface region at the interface between the structural lobes, adjacent to the segment containing the PIR repeat. The absence of a deeply defined binding pocket and the position of the ligand on an exposed surface suggest an interaction mediated by structurally

adaptable regions, consistent with the intrinsic flexibility of the PIR motif and its functional role in the recognition of cell wall carbohydrates.

Furthermore, the localisation of glucose in proximity to the single PIR repeat suggests that even an isolated repetitive module, when inserted into an appropriate structural context, can contribute significantly to carbohydrate recognition.

Therefore, the crystallographic data reinforce the validity of the approach adopted in this thesis, and the convergence between experimental structural observations and computational analyses supports the idea that the study of individual PIR repetitive constructs is sufficient to capture the key aspects of the functional behaviour of this domain.

To support the quality of the structural model in the apo form, a Ramachandran analysis was performed considering the general case (excluding Gly, Pro, and pre-Pro residues). (Figure 83)

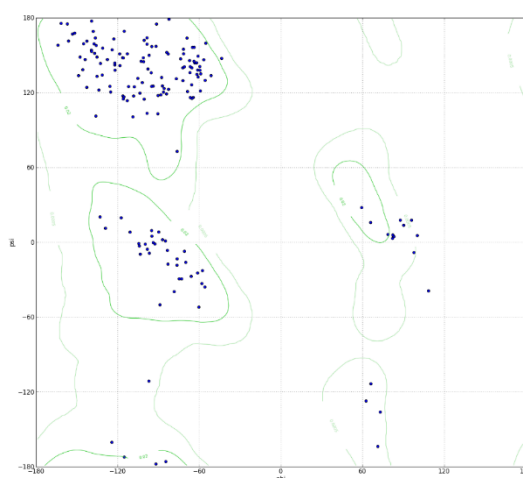


Figure 83. Ramachandran plot of PIR1r1 (general case, excluding Gly, Pro, and pre-Pro). Residues are represented as blue dots; favoured and allowed regions are indicated by density curves.

The distribution of ϕ/ψ angles shows that the vast majority of residues fall within the favoured or allowed regions, with a limited number of outliers. These deviations are mainly located in peripheral and flexible regions of the

structure, compatible with loops exposed to the solvent and not involved in the structural core or the PIR motif. (Figure 83)

Since this analysis confirms the good stereochemical quality of the model, the crystallographic structure of PIR1r1 has been included in this thesis as a structural reference in the apo form to explore, through molecular dynamics simulations, the stability of the PIR motif and the contribution of non-repeating regions (trefoil-like domain) to the overall conformational equilibrium of the protein, integrating the experimental results and providing an overview of the structural behaviour of single-repeat PIR modules.

4.3.5.2 Crystallisation of other PIRs

In parallel with the structural studies conducted on PIR1r1, extensive crystallisation attempts were also carried out for the PIR2r1, PIR4r1 and PIR5r1 proteins, with the aim of obtaining crystalline samples suitable for high-resolution structural analysis. These experiments were systematically conducted downstream of the purification processes described in the previous sections, using only the fractions eluted from size exclusion chromatography that showed the highest degree of purity in SDS-PAGE.

For each protein, the selected fractions were progressively concentrated until a wide range of protein concentrations was achieved, typically between 6, 12 and 50 mg/mL, in order to generate different crystallisation jobs and increase the probability of nucleation. The trials were set up using standard vapour diffusion (hanging drop) conditions, exploring a variety of precipitants and pH conditions, in line with what had already been successfully adopted for PIR1r1.

Despite extensive screening, none of the three proteins produced authentic and reproducible protein crystals. Crystalline-looking structures were observed under numerous conditions (Figure 84), initially interpreted as possible micro- or macro-crystals. However, manual analysis by loop picking revealed that these structures dissolved immediately upon contact, unequivocally indicating their nature as salt crystals or non-protein aggregates rather than ordered protein crystals. (Figure 84)

The absence of protein crystallisation for PIR2r1, PIR4r1 and PIR5r1 can be attributed to a combination of factors. On the one hand, the intrinsic characteristics of these proteins, and on the other hand, it must be recognised that the fine optimisation of crystallisation parameters requires a high level of specialist expertise and an advanced iterative approach, which is partly beyond the primary scope of this thesis.

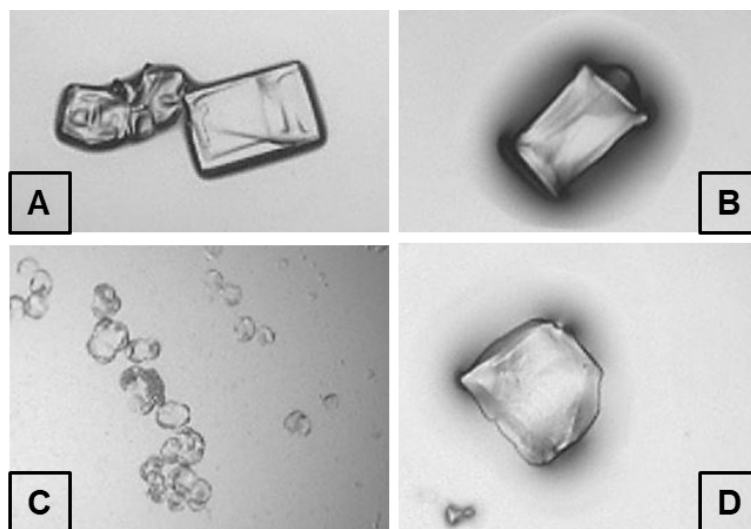


Figure 84. Representative examples of crystalline-looking objects observed during crystallisation screening of PIR2r1, PIR4r1 and PIR5r1 under vapour diffusion conditions. Panels A-D show morphologies initially compatible with protein crystals, including well-defined edges and faceted shapes; however, loop-picking tests revealed immediate dissolution upon handling, indicating that these structures correspond to salt crystals or non-protein aggregates rather than ordered protein crystals.

In this context, the negative results obtained should not be interpreted as mere experimental failures, but rather as relevant information that contributes to delineating the structural profile of the PIR proteins analysed. In particular, the systematic failure of crystallisation attempts for PIR2r1, PIR4r1 and PIR5r1 indirectly reinforces the hypothesis of greater conformational plasticity compared to PIR1r1, consistent with the biochemical, biophysical and computational data presented in the following sections. This evidence further justifies the use of alternative approaches,

such as molecular dynamics, to investigate the structural and dynamic behaviour of these proteins in solution.

4.3.6 Computational characterisation of PIR1r1: APO molecular dynamics simulations

Molecular dynamics simulations were performed on the crystallographic structure of PIR1r1 in apo form to characterise its dynamic behaviour in solution and to distinguish the relative contribution of the different structural elements to the overall stability of the protein. The system, consisting of a total of 174 residues, was simulated for 200 ns in three independent replicates, the results of which were subsequently averaged as described in Section 3.7.2.

Based on the structural architecture emerging from the crystallographic analysis, the protein was divided into two main regions:

- the repeated PIR motif (residues 28-74), characterised by a low-complexity sequence and a relatively compact structural organisation but poor in extended secondary elements;
- the trefoil-like domain (residues 1-27 and 75-174), which constitutes the tri-lobed portion of the protein and represents the most organised structural core.

This division allowed for the decoupled analysis of the overall stability of the fold and the intrinsic flexibility of the PIR module.

4.3.6.1 RMSD analysis

RMSD analysis was performed on the PIR1r1 structure at 300 K in order to evaluate the relative structural stability and dynamic behaviour of different regions of the protein under conditions close to physiological temperature. The RMSD profile calculated for the trefoil-like domain shows a rapid initial increase during the early stages of the simulation, followed by a stable plateau throughout the remainder of the trajectory. After equilibrium, the RMSD values fluctuate within a narrow range, centred around ~1.5-2.0 Å (Figure 85), indicating a well-preserved global fold with no evidence of large-

scale conformational reorganisation. This behaviour is consistent with the presence of a compact, cooperatively stabilised structural core and supports the interpretation of the trilobed architecture as an intrinsically stable scaffold under physiological conditions.

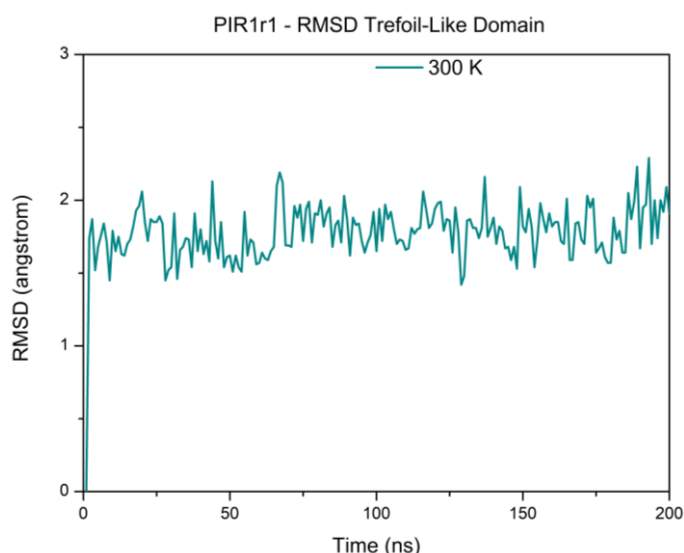


Figure 85. Time evolution of the RMSD calculated for the trefoil-like domain of PIR1r1 at 300 K which remains low and stable throughout the simulation, indicating preservation of the native fold.

In contrast, the RMSD calculated for the region containing the PIR shows a markedly different behaviour. After an initial increase, the RMSD quickly reaches significantly higher values, fluctuating mainly in the range ~ 7 -12 Å during the simulation (Figure 86). These large amplitude fluctuations reflect a high degree of conformational variability and indicate that the PIR motif does not adopt a single well-defined conformation, even at 300 K. Interestingly, no clear plateau corresponding to a rigid structural state is observed, suggesting that this region explores a large conformational ensemble rather than undergoing discrete folding-unfolding transitions. The framework revealed by this analysis reflects the presence of a structurally rigid trefoil-like domain that remains stable over time and a highly flexible region containing the internal PIR repeat characterised by high

conformational freedom, consistent with the architectural organisation inferred from the crystallographic structure. (Figure 86)

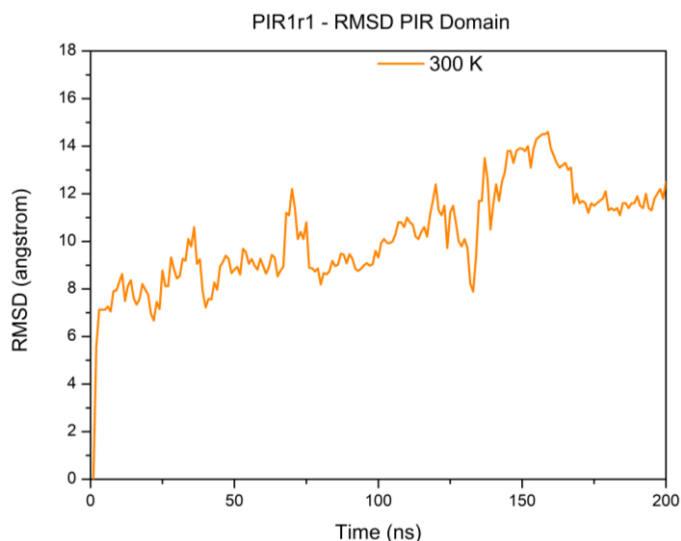


Figure 86. Time evolution of the RMSD calculated for the PIR motif of PIR1r1 at 300 K, at which temperature the RMSD assumes higher values than in the trefoil-like domain, reflecting intrinsic conformational flexibility.

4.3.6.2 RMSF analysis

The RMSF analysis was performed on the apo form of PIR1r1 at 300 K in order to characterise the distribution of local flexibility along the protein sequence under near-physiological conditions. Residue-by-residue fluctuations were analysed both for the full-length protein and for the trefoil-like domain considered independently, allowing the intrinsic mobility of the PIR-containing region to be directly compared with that of the structured core.

The RMSF profile calculated over the entire PIR1r1 sequence reveals a pronounced heterogeneity in residue flexibility (Figure 87). While most regions display moderate fluctuations, a marked peak is observed in the segment corresponding to the PIR motif (residues 28-74), where RMSF values reach significantly higher levels compared to the rest of the protein. This behaviour indicates that the PIR-containing region is characterised by enhanced conformational mobility, consistent with a combination of local flexibility and collective motions. Outside this region, fluctuations remain

comparatively limited, suggesting a more constrained structural organisation. (Figure 87)

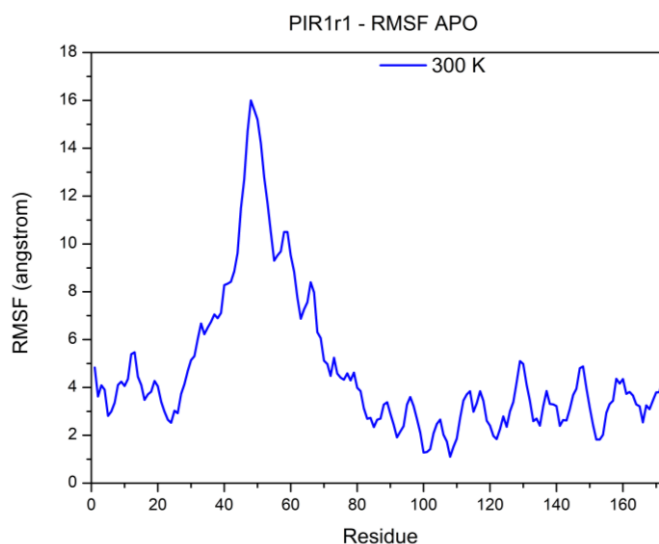


Figure 87. Residue-by-residue RMSF profile of apo PIR1r1 calculated at 300 K over the entire protein sequence. A pronounced increase in RMSF is observed in the region corresponding to the PIR motif (residues 28-74), indicating enhanced local flexibility and conformational mobility. Outside this region, residue fluctuations remain comparatively limited, consistent with a more constrained structural organisation of the protein.

To further isolate the behaviour of the structured core, RMSF analysis was restricted to the trefoil-like domain, excluding residues corresponding to the PIR motif (Figure 88). In this case, residue fluctuations remain relatively low and smoothly distributed along the sequence, with no dominant peaks indicative of local destabilisation. The observed RMSF values, typically in the range of $\sim 1\text{-}4$ Å, are consistent with a compact and well-defined fold that tolerates only limited local motions while preserving its overall architecture. It should be noted that the molecular dynamics simulations presented in this work were conducted exclusively on the apo form of PIR1r1. (Figure 88)

Although a crystallographic structure of PIR1r1 in complex with glucose was obtained in the host laboratory, this system was not included in the simulations for methodological reasons. In particular, an accurate description of the interaction between the PIR motif and a carbohydrate ligand would require the use of advanced sampling approaches (e.g.,

accelerated molecular dynamics techniques or enhanced sampling methods) capable of reliably capturing recognition and stabilisation events mediated by flexible surfaces.

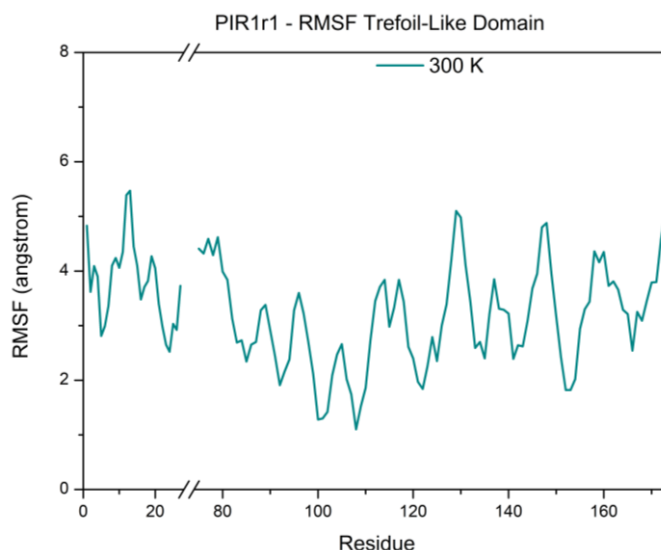


Figure 88. RMSF profile calculated for the trefoil-like domain of PIR1r1 at 300 K. Residues corresponding to the PIR motif (28-74) are excluded, resulting in a discontinuous x-axis. Fluctuations within the trefoil-like domain remain relatively low and smoothly distributed, consistent with a compact and structurally stable scaffold that undergoes only limited local motions under physiological conditions.

In the context of classical simulations at the atomic scale and on accessible timescales, the analysis of the bound complex would have provided limited and potentially misleading information, especially in light of the high intrinsic mobility of the PIR motif that emerged from the apo simulations.

For this reason, the computational analysis deliberately focused on characterising the stability of the tri-lobed fold and the intrinsic dynamics of the PIR motif in the absence of ligands, laying the foundations for future computational studies aimed at carbohydrate recognition using more appropriate methodological approaches.

4.3.7 Discussion

The data obtained in this section are the outcome of expression, purification and crystallisation tests on single-repeat PIR constructs in order to correlate

the design rationale with the observed biochemical behaviour and, in the case of PIR1r1, with the structural and dynamic properties observed experimentally and computationally.

The design of the single-repeat constructs was based on native PIR proteins, ensuring that they shared the same basic N-terminal architecture (His-tag and TEV cleavage site) and a C-terminal region with at least one canonical repeat with the consensus motif.

Initially, the analysis was conducted on the GFP-PIR1r1-8 system, a protein system of approximately 45 kDa and eight repeats, already extensively characterised in the host laboratory. The peculiarity of this construct lies in the presence of GFP, which facilitates the monitoring of expression at 30 °C for 4 hours and the purification stages, making the correlation between the fluorimetric signal, the SEC profile and the SDS-PAGE bands immediate. The subsequent removal of the His tag and the formation of crystals meant that it could be used as a reference system to set up the experimental phases of the single-repeat constructs. This new construct was first screened for expression under multiple conditions in order to identify the best expression conditions for large-scale production. The results obtained in the initial expression phase showed that, despite its theoretical weight of ~23 kDa, the protein appeared in all tested conditions as a stable and clearly visible band in SDS-PAGE at an apparent weight of ~30 kDa. However, initial purification did not result in any recovery in Ni-NTA elution, and anti-His Western blotting performed on a representative aliquot of PIR2r1 revealed a compatibility issue between the state of the protein and the purification conditions adopted.

Subsequent optimisation of the buffer, by varying the pH and NaCl concentration, significantly changed the outcome of the purification. In fact, with the same strain, medium and induction conditions, PIR2r1 was effectively recovered in the Ni-NTA elution. Mass spectrometry on the eluted band, again with a higher apparent molecular weight than the theoretical one, confirmed the identity of the protein, ruling out any doubts about the nature of the purification product.

The optimisation of the buffer solution thus transformed an apparently negative result into a clear reading of the system, showing how the sequence properties of the PIR region and the surrounding context directly influence the choice of chromatographic conditions. Similarly, the same screening procedure was applied to the PIR4r1 and PIR5r1 constructs.

In addition, in light of the abnormal electrophoretic mobility observed for the PIR2r1, PIR4r1 and PIR5r1 constructs in SDS-PAGE, there was legitimate doubt about the integrity of the constructs, potentially attributable to cloning errors or unexpected mutations. For this reason, plasmid sequencing was included as an additional quality control. The complete concordance between the obtained and expected sequences excludes the presence of frameshift mutations, insertions or other reading frame alterations, and the abnormal migration could be attributed to intrinsic characteristics of the proteins such as their amino acid composition or interaction with SDS. Once the best expression and purification conditions had been established, each construct was systematically produced on a large scale, with protein yields adequate for structural characterisation. The SEC profiles revealed distinct oligomeric behaviours: PIR2r1 occurs in solution predominantly as a dimeric species, PIR4r1 as a predominantly monomeric species with a minor fraction of dimers, while PIR5r1 is essentially monomeric. These differences suggest that, despite sharing a similar overall architecture, the linker containing the repetitive elements is able to modulate stability, oligomeric behaviour in solution and propensity for crystallisation pathways in distinct ways.

Despite good expression and purification yields of the constructs considered, crystallisation trials from monodisperse fractions using SEC did not lead to the formation of protein crystals, indicating that for these constructs, the level of homogeneity and stability is not sufficient to draw definitive conclusions about their intrinsic structural properties, but rather reflects experimental limitations related to the combination of local flexibility, oligomeric behaviour, and sensitivity to operating conditions.

Although not the central focus of this thesis, PIR1r1 is the only example of a single-repeat construct for which it has been possible to obtain a high-resolution crystallographic structure in the host laboratory, as well as dynamic characterisation through molecular dynamics simulations. The crystallographic structure of PIR1r1 reveals a compact trilobed domain, similar to a trefoil, which acts as a stable structural scaffold. The region containing the PIR motif is located outside this trilobed domain, in a peripheral portion of the protein, topologically distinct from the scaffold. Molecular dynamics simulations in apo form show a clear dynamic separation between the trilobed domain, characterised by low RMSD values and a uniform RMSF profile, and the outer region associated with the PIR motif, which has significantly higher residue-by-residue mobility. This modular organisation emerges as an intrinsic feature of the PIR module and provides a useful interpretative key to understanding the biochemical behaviour observed in the PIR2r1, PIR4r1 and PIR5r1 constructs. Furthermore, in the host laboratory, the PIR1r1 structure was obtained by soaking pre-formed crystals in the presence of glucose, and refinement processes revealed that the ligand is located near the repeat, providing structural evidence supporting the involvement of the PIR motif in carbohydrate interaction, in line with the proposed role of PIR proteins in the cell wall.

Among other biochemical analyses performed, Western blotting with anti-PIR2 antibodies provided final immunological validation of the identity of the single-repeat constructs analysed. Despite the anomalous electrophoretic mobility observed in SDS-PAGE, the PIR1r1, PIR2r1 and PIR4r1 constructs were specifically recognised by the antibodies, demonstrating that migration anomalies do not compromise antigenic recognition and do not reflect substantial alterations in the sequence or overall fold of the purified proteins. In particular, robust immunoreactivity was observed even at extremely low loads (0.25 ng), confirming the exposure and accessibility of the epitope, in full agreement with the data obtained on yeast cell extracts.

Together, these data provide convergent validation of the identity and integrity of single-repetition PIR constructs, allowing reliable interpretation of the observed differences in solubility, oligomeric behaviour, and crystallisation propensity as consequences of intrinsic structural and dynamic properties of PIR modules.

5 Conclusions and future perspectives

Aim of this thesis was to describe how different protein design approaches could affect the stability and the functionality of two different protein systems interesting for biotechnological applications.

The sweet protein MNEI is known not only for its sweetening properties but also because it is used as a model protein in studies on aggregation under stress conditions. Two distinct approaches were adopted to improve its physicochemical properties: point mutation by substitution and circular permutation.

The point mutations by substitution approach applied to MNEI represents a completely strategy in which the chain is not cut and sewn, but rather repaired by replacing Glu23 with Gln, allowing the elimination of a negative charge buried in a hydrophobic cavity and simultaneously giving it greater stability. The mutant created, called Mut4, not only has increased thermal stability which increases with increasing pH and salt, but also has a marked attenuation of structural sensitivity to pH variations. This improved stability is accompanied by a drastic reduction in access to the aggregation pathways typical of MNEI. Unlike what has been observed in the literature for Y65R or Mut9 [78,78,90], where the introduction or removal of charged residues changes the balance of the chain, favouring helix opening and aggregation in a wide range of pH and salt levels, in Mut4 neither fibrils nor amorphous aggregates are observed, even in conditions that are strongly aggregating for MNEI. MD simulations confirm that the mutated region does not become more flexible, but looks more rigid, as if it were less prone to triggering local openings, thus preventing access to partially unfolded states that, in MNEI and other mutants, are associated with the beginning of aggregation processes. In perspective, Mut4 can be interpreted as an example of targeted stabilisation, in which the removal of a negative charge from an unfavourable environment does not introduce new interactions but rather removes a source of instability, resulting in a more stable protein that maintains its fold and does not easily explore partially unfolded states that lead to aggregation. A similar approach could be applied to other proteins in

which single internal charges compromise stability. One question that remains open concerns the sweetness of Mut4, and since E23Q does not change the fold or the distribution of surface charges, it is reasonable to assume that the interaction with the T1R2/T1R3 receptor is not substantially affected, although small changes in dynamics could influence the intensity of the response. Checking this would require specific functional assays via sensory tests and in vitro studies on the sweet taste receptor. Indeed, if sweetness remained unchanged, Mut4 would represent a case of rational optimisation in which enhanced stability does not compromise function, providing a solid basis for the development of more stable protein variants suitable for food applications.

The circular permutation technique, instead, represents a profound topological modification, as the MNEI polypeptide chain was cut at a specific point (between $\beta 2$ and $\beta 3$) and re-sewn at the native ends using linkers of different lengths, generating three new mutants, Perm1, Perm2 and Perm3. In fact, although the connectivity of the chain changed, the cystatin α/β fold remained essentially conserved. This detail was crucial in light of the chemical-physical and computational characterisations that converge in showing that the length of the linker is not a minor detail, but rather a parameter capable of influencing the equilibrium in solution between the three mutants: Perm1, with the longest linker, behaves as a predominantly monomeric species and, under aggregating conditions, evolves more slowly and with sigmoidal kinetics towards long, thin and ordered fibrils, with a morphology coherent at both selected temperatures. In contrast, Perm2, which is mostly monomeric but with a tendency to dimerise, and Perm3, which is completely dimeric and has the shortest linker of the three, evolve rapidly towards off-pathway amorphous aggregates under the same aggregation conditions. In addition, X-ray crystallography shows that all circularly permuted mutants can form domain-swapped dimers, but with different interfaces, more compact in mutants with short linkers and more open in Perm1. Molecular dynamics simulations and principal component analysis support this interpretation, since in Perm1 the opening involves

larger regions, including the α -helix, favouring a global rearrangement compatible with fibril formation, while in Perm3 localised fluctuations are consistent with early access to the swapped state and consequently disordered aggregation. What clearly stands out is that domain swapping is not related to ordered fibrillation. In fact, in the systems studied, a greater predisposition to domain swapping is associated with conversion into amorphous aggregates, while monomer stability appears to be a key factor for allowing a slower and more organised pathway that leads to the formation of ordered fibrils. These results help to understand how the interplay between different factors affecting the stability of MNEI in solution may be a determining element in modulating aggregation, given its potential as an intense sweetener. At the same time, if it is true that fibrils can be a problem, especially in biological and pathological contexts, in the field of bio-nanomaterials and nanotechnology they are considered versatile elements for the design of nanostructured biomaterials. There are several potential biotechnological applications, from the use of β -rich fibrils as scaffolds for filter membranes, explored for the removal of contaminants in wastewater (metals, dyes, organic micropollutants), to their use as reinforcement in hydrogels, as matrices for trapping molecules and releasing them slowly (drug delivery), and even as scaffolds for obtaining nanowires or conductive composites when coupled with nanoparticles or metal salts. [15,201–203]

The value of this phenomenon is that it can be controlled and developed, making it a useful tool for biotech design and applications.

The latest system investigated is represented by yeast PIR (Proteins with Internal Repeats) proteins, which explore another application of protein sequence tailoring. PIR proteins are a family of structural components of the yeast cell wall, characterised by the presence of well-conserved repeated modules, located mainly in the N-terminal region and containing the consensus motif SQGDGQQA. Each repeated module acts as a functional unit capable of covalently anchoring the protein to the β -1,3-glucans of the wall. The presence of multiple repeats along the same chain allows for

multiple anchoring points, distributed spatially along the matrix, contributing cooperatively to the mechanical stability of the cell wall. In this way, the cooperative effect of the repeated modules contributes to the mechanical strength of the cell wall, increasing its ability to absorb and distribute stress. In this case, the aim is not to change the protein's topology, but to reduce its modularity in a controlled way, moving from a structure with multiple internal repetitions to constructs with a single module, to see if a single repetition is enough to keep it attached to the wall and to check out the main structural properties linked to PIRs. The constructs analysed were first subjected to systematic screening. Once the best conditions for expression and purification had been identified, they were analysed through biochemical, structural and computational analyses. Despite sharing a similar basic architecture, they show distinct behaviours in SEC: PIR2r1 occurs in solution mainly as a dimeric species, PIR4r1 as a predominantly monomeric species with a small dimeric fraction, while PIR5r1 is essentially monomeric. Even though the constructs under investigation showed excellent expression and high purification yields, no crystals were obtained under the conditions explored, suggesting possible structural flexibility or conformational heterogeneity incompatible with the ordered packing required for crystallisation. However, this result does not exclude the possibility of obtaining structural information through alternative approaches or through further optimisation of the crystallisation conditions, thus representing a potential future development of the work.

Although not the central topic of this thesis, PIR1r1 is the only example of a single-repeat construct whose crystallographic structure has been solved in the host laboratory and has a compact trilobed domain that acts as a stable scaffold, with the region containing the PIR motif being more mobile and peripheral. Molecular dynamics simulations performed on this protein show a clear dynamic separation between the stable module and the more flexible repetitive portion. Moreover, crystallisation trials performed by soaking pre-formed PIR1r1 crystals in the presence of glucose revealed the localisation of a sugar molecule near the repetitive module, providing a structural hint

supporting the role of PIR proteins in the interaction with carbohydrates in the cell wall.

From a biological point of view, understanding how a single isolated repetition behaves could help us to better interpret the function of complete PIR proteins, which in the yeast wall must combine structural stability with the ability to bind polysaccharides. To truly understand their role in the cellular context, however, further biological, genetic and proteomic analyses are necessary. Knockout or overexpression studies are an effective strategy for assessing if and how the loss of a specific PIR affects cell wall integrity or the response to stresses that disrupt its structure, such as Calcofluor White, Congo Red, and other wall-disrupting agents. [56,204,205] In this way, structural and biochemical data obtained in vitro may be directly linked to the physiological function of the protein in yeast. Not only that, but a deep understanding of structure, dynamics and function opens the way to biotechnological applications, allowing PIRs to be exploited as modular platforms for controlling the surface properties of yeast, for the stable anchoring of enzymes or antigens, and for the design of protein-polysaccharide systems in industrial and biomedical fields. [206,207]

6 References

- [1] S.N. Cohen, A.C. Chang, H.W. Boyer, R.B. Helling, Construction of biologically functional bacterial plasmids in vitro, *Proc Natl Acad Sci U S A* 70 (1973) 3240–3244. <https://doi.org/10.1073/pnas.70.11.3240>.
- [2] P.V. Pham, Medical Biotechnology, in: *Omics Technologies and Bio-Engineering*, Elsevier, 2018: pp. 449–469. <https://doi.org/10.1016/B978-0-12-804659-3.00019-1>.
- [3] S.B. Ebrahimi, D. Samanta, Engineering protein-based therapeutics through structural and chemical design, *Nat Commun* 14 (2023) 2411. <https://doi.org/10.1038/s41467-023-38039-x>.
- [4] R.M. Blizzard, History of Growth Hormone Therapy, *Indian J Pediatr* 79 (2012) 87–91. <https://doi.org/10.1007/s12098-011-0609-4>.
- [5] Z. Song, Q. Zhang, W. Wu, Z. Pu, H. Yu, Rational design of enzyme activity and enantioselectivity, *Front Bioeng Biotechnol* 11 (2023) 1129149. <https://doi.org/10.3389/fbioe.2023.1129149>.
- [6] F.H. Arnold, Design by Directed Evolution, *Acc. Chem. Res.* 31 (1998) 125–131. <https://doi.org/10.1021/ar960017f>.
- [7] F.H. Arnold, A.A. Volkov, Directed evolution of biocatalysts, *Current Opinion in Chemical Biology* 3 (1999) 54–59. [https://doi.org/10.1016/S1367-5931\(99\)80010-6](https://doi.org/10.1016/S1367-5931(99)80010-6).
- [8] E.T. Farinas, T. Bulter, F.H. Arnold, Directed enzyme evolution, *Current Opinion in Biotechnology* 12 (2001) 545–551. [https://doi.org/10.1016/S0958-1669\(01\)00261-0](https://doi.org/10.1016/S0958-1669(01)00261-0).
- [9] J.L. Watson, D. Juergens, N.R. Bennett, B.L. Trippe, J. Yim, H.E. Eisenach, W. Ahern, A.J. Borst, R.J. Ragotte, L.F. Milles, B.I.M. Wicky, N. Hanikel, S.J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S.V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T.S. Jaakkola, F. DiMaio, M. Baek, D. Baker, De novo design of protein structure and function with RFdiffusion, *Nature* 620 (2023) 1089–1100. <https://doi.org/10.1038/s41586-023-06415-8>.

- [10] J. Yao, X. Wang, Artificial intelligence in de novo protein design, *Medicine in Novel Technology and Devices* 26 (2025) 100366. <https://doi.org/10.1016/j.medntd.2025.100366>.
- [11] G. Zhang, C. Liu, J. Lu, S. Zhang, L. Zhu, The Role of AI-Driven De Novo Protein Design in the Exploration of the Protein Functional Universe, *Biology* 14 (2025) 1268. <https://doi.org/10.3390/biology14091268>.
- [12] A.B. Guo, D. Akpinaroglu, C.A. Stephens, M. Grabe, C.A. Smith, M.J.S. Kelly, T. Kortemme, Deep learning–guided design of dynamic proteins, *Science* 388 (2025) eadr7094. <https://doi.org/10.1126/science.adr7094>.
- [13] J.A. Morris, R.H. Cagan, Purification of monellin, the sweet principle of *Dioscoreophyllum cumminsii*, *Biochimica et Biophysica Acta (BBA) - General Subjects* 261 (1972) 114–122. [https://doi.org/10.1016/0304-4165\(72\)90320-0](https://doi.org/10.1016/0304-4165(72)90320-0).
- [14] N. Louros, J. Schymkowitz, F. Rousseau, Mechanisms and pathology of protein misfolding and aggregation, *Nat Rev Mol Cell Biol* 24 (2023) 912–933. <https://doi.org/10.1038/s41580-023-00647-2>.
- [15] G. Wei, Z. Su, N.P. Reynolds, P. Arosio, I.W. Hamley, E. Gazit, R. Mezzenga, Self-assembling peptide and protein amyloids: from structure to tailored function in nanotechnology, *Chem. Soc. Rev.* 46 (2017) 4661–4708. <https://doi.org/10.1039/C6CS00542J>.
- [16] M. Peydayesh, S. Bolisetty, T. Mohammadi, R. Mezzenga, Assessing the Binding Performance of Amyloid–Carbon Membranes toward Heavy Metal Ions, *Langmuir* 35 (2019) 4161–4170. <https://doi.org/10.1021/acs.langmuir.8b04234>.
- [17] T.M. Cezar, M. Lozančić, A. Novačić, A. Matičević, D. Matijević, B. Vallee, V. Mrša, R. Teparić, B. Žunar, Boosting N-terminally anchored yeast surface display via structural insights into *S. cerevisiae* Pir proteins, (2023). <https://doi.org/10.1101/2023.04.25.538238>.
- [18] J.D. Bloom, J.J. Silberg, C.O. Wilke, D.A. Drummond, C. Adami, F.H. Arnold, Thermodynamic prediction of protein neutrality, *Proc. Natl.*

- Acad. Sci. U.S.A. 102 (2005) 606–611.
<https://doi.org/10.1073/pnas.0406744102>.
- [19] P.-S. Huang, Y.-E.A. Ban, F. Richter, I. Andre, R. Vernon, W.R. Schief, D. Baker, RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design, *PLoS ONE* 6 (2011) e24109.
<https://doi.org/10.1371/journal.pone.0024109>.
- [20] P.-S. Huang, S.E. Boyken, D. Baker, The coming of age of de novo protein design, *Nature* 537 (2016) 320–327.
<https://doi.org/10.1038/nature19946>.
- [21] K. Wüthrich, NMR studies of structure and function of biological macromolecules., *Bioscience Reports* 23 (2003) 119–68.
- [22] A. McPherson, J.A. Gavira, Introduction to protein crystallization, *Acta Crystallogr F Struct Biol Commun* 70 (2014) 2–20.
<https://doi.org/10.1107/S2053230X13033141>.
- [23] W. Kühlbrandt, Cryo-EM enters a new era, *eLife* 3 (2014) e03678.
<https://doi.org/10.7554/eLife.03678>.
- [24] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature* 596 (2021) 583–589.
<https://doi.org/10.1038/s41586-021-03819-2>.
- [25] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, R.D. Schaeffer, C. Millán, H. Park, C. Adams, C.R. Glassman, A. DeGiovanni, J.H. Pereira, A.V. Rodrigues, A.A. Van Dijk, A.C. Ebrecht, D.J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M.K. Rathinaswamy, U. Dalwadi, C.K. Yip, J.E. Burke, K.C. Garcia, N.V. Grishin, P.D. Adams, R.J. Read, D. Baker, Accurate prediction of protein structures and

- interactions using a three-track neural network, *Science* 373 (2021) 871–876. <https://doi.org/10.1126/science.abj8754>.
- [26] N. Balasco, L. Esposito, L. Vitagliano, Structural Biology in the AlphaFold Era: How Far Is Artificial Intelligence from Deciphering the Protein Folding Code?, *Biomolecules* 15 (2025) 674. <https://doi.org/10.3390/biom15050674>.
- [27] N. Tokuriki, D.S. Tawfik, Protein Dynamism and Evolvability, *Science* 324 (2009) 203–207. <https://doi.org/10.1126/science.1169375>.
- [28] H.N. Motlagh, J.O. Wrabl, J. Li, V.J. Hilser, The ensemble nature of allostery, *Nature* 508 (2014) 331–339. <https://doi.org/10.1038/nature13001>.
- [29] A. Jimenez-Rosales, M.V. Flores-Merino, Tailoring Proteins to Re-Evolve Nature: A Short Review, *Mol Biotechnol* 60 (2018) 946–974. <https://doi.org/10.1007/s12033-018-0122-3>.
- [30] J.D. Bloom, P.A. Romero, Z. Lu, F.H. Arnold, Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution, *Biol Direct* 2 (2007) 17. <https://doi.org/10.1186/1745-6150-2-17>.
- [31] B. Kuhlman, D. Baker, Native protein sequences are close to optimal for their structures, *Proc. Natl. Acad. Sci. U.S.A.* 97 (2000) 10383–10388. <https://doi.org/10.1073/pnas.97.19.10383>.
- [32] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, J. Thompson, R. Jacak, K.W. Kaufman, P.D. Renfrew, C.A. Smith, W. Sheffler, I.W. Davis, S. Cooper, A. Treuille, D.J. Mandell, F. Richter, Y.-E.A. Ban, S.J. Fleishman, J.E. Corn, D.E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J.J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J.J. Gray, B. Kuhlman, D. Baker, P. Bradley, Rosetta3, in: *Methods in Enzymology*, Elsevier, 2011: pp. 545–574. <https://doi.org/10.1016/B978-0-12-381270-4.00019-6>.
- [33] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A.J. Ballard, J. Bambrick, S.W. Bodenstein, D.A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool,

- Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A.I. Cowen-Rivers, A. Cowie, M. Figurnov, F.B. Fuchs, H. Gladman, R. Jain, Y.A. Khan, C.M.R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E.D. Zhong, M. Zielinski, A. Židek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J.M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3, *Nature* 630 (2024) 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.
- [34] Z. Fang, H. Ran, Y. Zhang, C. Chen, P. Lin, X. Zhang, M. Wu, AlphaFold 3: an unprecedented opportunity for fundamental research and drug development, *Precis Clin Med* 8 (2025) pbaf015. <https://doi.org/10.1093/pcmedi/pbaf015>.
- [35] Chai Discovery, J. Boitreaud, J. Dent, M. McPartlon, J. Meier, V. Reis, A. Rogozhnikov, K. Wu, Chai-1: Decoding the molecular interactions of life, (2024). <https://doi.org/10.1101/2024.10.10.615955>.
- [36] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R.J. Ragotte, L.F. Milles, B.I.M. Wicky, A. Courbet, R.J. De Haas, N. Bethel, P.J.Y. Leung, T.F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A.K. Bera, N.P. King, D. Baker, Robust deep learning–based protein sequence design using ProteinMPNN, *Science* 378 (2022) 49–56. <https://doi.org/10.1126/science.add2187>.
- [37] F. Chiti, C.M. Dobson, Protein Misfolding, Functional Amyloid, and Human Disease, *Annu. Rev. Biochem.* 75 (2006) 333–366. <https://doi.org/10.1146/annurev.biochem.75.101304.123901>.
- [38] V.J. Hilser, J.O. Wrabl, C.E.F. Millard, A. Schmitz, S.J. Brantley, M. Pearce, J. Rehfus, M.M. Russo, K. Voortman-Sheetz, Statistical Thermodynamics of the Protein Ensemble: Mediating Function and Evolution, *Annual Review of Biophysics* 54 (2025) 227–247. <https://doi.org/10.1146/annurev-biophys-061824-104900>.
- [39] J.M. Sanchez-Ruiz, Protein kinetic stability, *Biophysical Chemistry* 148 (2010) 1–15. <https://doi.org/10.1016/j.bpc.2010.02.004>.

- [40] E. Shakhnovich, Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet, *Chem. Rev.* 106 (2006) 1559–1588. <https://doi.org/10.1021/cr040425u>.
- [41] J. Echave, S.J. Spielman, C.O. Wilke, Causes of evolutionary rate variation among protein sites, *Nat Rev Genet* 17 (2016) 109–121. <https://doi.org/10.1038/nrg.2015.18>.
- [42] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1751 (2005) 119–139. <https://doi.org/10.1016/j.bbapap.2005.06.005>.
- [43] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, *Nat Protoc* 1 (2006) 2876–2890. <https://doi.org/10.1038/nprot.2006.202>.
- [44] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.* 11 (2015) 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- [45] C. Tian, K. Kasavajhala, K.A.A. Belfon, L. Raguette, H. Huang, A.N. Migués, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution, *J. Chem. Theory Comput.* 16 (2020) 528–552. <https://doi.org/10.1021/acs.jctc.9b00591>.
- [46] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E.C. III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L.

- Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, P.A. Kollman, Amber 2021, University of California, San Francisco, 2021. <https://ambermd.org>.
- [47] C.M. Dobson, Protein folding and misfolding, *Nature* 426 (2003) 884–890. <https://doi.org/10.1038/nature02261>.
- [48] J.A. Brannigan, A.J. Wilkinson, Protein engineering 20 years on, *Nat Rev Mol Cell Biol* 3 (2002) 964–970. <https://doi.org/10.1038/nrm975>.
- [49] I. Yadid, D.S. Tawfik, Reconstruction of Functional β -Propeller Lectins via Homo-oligomeric Assembly of Shorter Fragments, *Journal of Molecular Biology* 365 (2007) 10–17. <https://doi.org/10.1016/j.jmb.2006.09.055>.
- [50] J. Ha, S.N. Loh, Protein Conformational Switches: From Nature to Design, *Chemistry A European J* 18 (2012) 7984–7999. <https://doi.org/10.1002/chem.201200348>.
- [51] Z. Shi, W. Xu, L. Wu, X. Yue, S. Liu, W. Ding, J. Zhang, B. Meng, L. Zhao, X. Liu, J. Liu, Z.-J. Liu, T. Hua, Structural and functional characterization of human sweet taste receptor, *Nature* 645 (2025) 801–808. <https://doi.org/10.1038/s41586-025-09302-6>.
- [52] P.A. Temussi, Natural sweet macromolecules: how sweet proteins work, *Cell. Mol. Life Sci.* 63 (2006) 1876–1888. <https://doi.org/10.1007/s00018-006-6077-8>.
- [53] R. Spadaccini, S. Leone, M.F. Rega, C. Richter, D. Picone, Influence of pH on structure and stability of the sweet protein MNEI, *FEBS Letters* 590 (2016) 3681–3689. <https://doi.org/10.1002/1873-3468.12437>.
- [54] J.C. Kapteyn, H. Van Den Ende, F.M. Klis, The contribution of cell wall proteins to the organization of the yeast cell wall, *Biochimica et Biophysica Acta (BBA) - General Subjects* 1426 (1999) 373–383. [https://doi.org/10.1016/S0304-4165\(98\)00137-8](https://doi.org/10.1016/S0304-4165(98)00137-8).

- [55] G. Lesage, H. Bussey, Cell Wall Assembly in *Saccharomyces cerevisiae*, *Microbiol Mol Biol Rev* 70 (2006) 317–343. <https://doi.org/10.1128/MMBR.00038-05>.
- [56] P. Montañó-Silva, O.A. Callejas-Negrete, A. Pereira-Santana, J. Verdín, Cell wall-resident proteins with internal repeats (PIR s) show an inverted architecture in *Neurospora crassa* , but maintain their role as wall stabilizers, *The FEBS Journal* 292 (2025) 2578–2601. <https://doi.org/10.1111/febs.70020>.
- [57] J. Breda, J. Jewell, A. Keller, The Importance of the World Health Organization Sugar Guidelines for Dental Health and Obesity Prevention, *Caries Res* 53 (2019) 149–152. <https://doi.org/10.1159/000491556>.
- [58] S. Firouzi, M. Gholami Parashkoochi, D.M. Zamani, I. Ranjber, An Investigation of the Environmental Impacts and Energy-Economic Analysis for Sugar Beet and Sugarcane Production Systems, *Sugar Tech* 24 (2022) 1851–1866. <https://doi.org/10.1007/s12355-022-01135-1>.
- [59] S. Lohner, I. Toews, J.J. Meerpohl, Health outcomes of non-nutritive sweeteners: analysis of the research landscape, *Nutr J* 16 (2017) 55. <https://doi.org/10.1186/s12937-017-0278-x>.
- [60] A.C. Sylvestsky, K.I. Rother, Nonnutritive Sweeteners in Weight Management and Chronic Disease: A Review, *Obesity* 26 (2018) 635–640. <https://doi.org/10.1002/oby.22139>.
- [61] R. Kant, Sweet proteins – Potential replacement for artificial low calorie sweeteners, *Nutr J* 4 (2005) 5. <https://doi.org/10.1186/1475-2891-4-5>.
- [62] M. Kohmura, T. Mizukoshi, N. Nio, E. Suzuki, Y. Ariyoshi, Structure-taste relationships of the sweet protein monellin, *Pure and Applied Chemistry* 74 (2002) 1235–1242. <https://doi.org/10.1351/pac200274071235>.
- [63] T.S. Novik, E.I. Koveschnikova, A.A. Kotlobay, L.P. Sycheva, K.G. Kurochkina, O.A. Averina, M.V. Belopolskaya, P.V. Sergiev, O.A.

- Dontsova, V.N. Lazarev, I.V. Maev, M.G. Kostyaeva, A.V. Ereemeev, S.I. Chukina, M.A. Lagarkova, Sweet-Tasting Natural Proteins Brazzein and Monellin: Safe Sugar Substitutes for the Food Industry, *Foods* 12 (2023) 4065. <https://doi.org/10.3390/foods12224065>.
- [64] R. Gutierrez, E. Fonseca, S.A. Simon, The neuroscience of sugars in taste, gut-reward, feeding circuits, and obesity, *Cell. Mol. Life Sci.* 77 (2020) 3469–3502. <https://doi.org/10.1007/s00018-020-03458-2>.
- [65] J.A. Joseph, S. Akkermans, P. Nimmegeers, J.F.M. Van Impe, Bioproduction of the Recombinant Sweet Protein Thaumatin: Current State of the Art and Perspectives, *Front. Microbiol.* 10 (2019) 695. <https://doi.org/10.3389/fmicb.2019.00695>.
- [66] J.R. Somoza, J.M. Cho, S.-H. Kim, The Taste-active Regions of Monellin, a Potently Sweet Protein, *Chem Senses* 20 (1995) 61–68. <https://doi.org/10.1093/chemse/20.1.61>.
- [67] F.M. Assadi-Porter, Brazzein, a Small, Sweet Protein: Effects of Mutations on its Structure, Dynamics and Functional Properties, *Chemical Senses* 30 (2005) i90–i91. <https://doi.org/10.1093/chemse/bjh128>.
- [68] T. Masuda, N. Kitabatake, Developments in biotechnological production of sweet proteins, *Journal of Bioscience and Bioengineering* 102 (2006) 375–389. <https://doi.org/10.1263/jbb.102.375>.
- [69] J.A. Morris, R. Martenson, G. Deibler, R.H. Cagan, Characterization of Monellin, a Protein That Tastes Sweet, *Journal of Biological Chemistry* 248 (1973) 534–539. [https://doi.org/10.1016/S0021-9258\(19\)44407-4](https://doi.org/10.1016/S0021-9258(19)44407-4).
- [70] C. Ogata, M. Hatada, G. Tomlinson, W.-C. Shin, S.-H. Kim, Crystal structure of the intensely sweet protein monellin, *Nature* 328 (1987) 739–742. <https://doi.org/10.1038/328739a0>.
- [71] Y.-H. Sung, H.-D. Hong, C. Cheong, J.H. Kim, J.M. Cho, Y.-R. Kim, W. Lee, Folding and Stability of Sweet Protein Single-chain Monellin,

- Journal of Biological Chemistry 276 (2001) 44229–44238. <https://doi.org/10.1074/jbc.M104239200>.
- [72] J. Thomas, The Miracle Fruit: It Miraculously Turns Sour Tastes Into Sweet, (2022). <https://metaefficient.com/food-and-drink/miracle-fruit-it-makes-sour-bitters-foods-sweet.html>.
- [73] J.R. Somoza, F. Jiang, L. Tong, C.-H. Kang, J.M. Cho, S.-H. Kim, Two Crystal Structures of a Potently Sweet Protein, Journal of Molecular Biology 234 (1993) 390–404. <https://doi.org/10.1006/jmbi.1993.1594>.
- [74] S.-H. Kim, C.-H. Kang, R. Kim, J.M. Cho, Y.-B. Lee, T.-K. Lee, Redesigning a sweet protein: increased stability and renaturability, Protein Eng Des Sel 2 (1989) 571–575. <https://doi.org/10.1093/protein/2.8.571>.
- [75] W.-F. Xue, O. Szczepankiewicz, M. C. Bauer, E. Thulin, S. Linse, Intra- versus Intermolecular Interactions in Monellin: Contribution of Surface Charges to Protein Assembly, Journal of Molecular Biology 358 (2006) 1244–1255. <https://doi.org/10.1016/j.jmb.2006.02.069>.
- [76] S. Leone, A. Pica, A. Merlino, F. Sannino, P.A. Temussi, D. Picone, Sweeter and stronger: enhancing sweetness and stability of the single chain monellin MNEI through molecular design, Sci Rep 6 (2016) 34045. <https://doi.org/10.1038/srep34045>.
- [77] S. Leone, D. Picone, Molecular Dynamics Driven Design of pH-Stabilized Mutants of MNEI, a Sweet Protein, PLoS ONE 11 (2016) e0158372. <https://doi.org/10.1371/journal.pone.0158372>.
- [78] R. Lucignano, R. Spadaccini, A. Merlino, D. Ami, A. Natalello, G. Ferraro, D. Picone, Structural insights and aggregation propensity of a super-stable monellin mutant: A new potential building block for protein-based nanostructured materials, International Journal of Biological Macromolecules 254 (2024) 127775. <https://doi.org/10.1016/j.ijbiomac.2023.127775>.
- [79] F. Donnarumma, A. Emendato, S. Leone, C. Ercole, G. D’Errico, D. Picone, Salt Modulated Fibrillar Aggregation of the Sweet Protein

- MNEI in Aqueous Solution, *J Solution Chem* 47 (2018) 939–949.
<https://doi.org/10.1007/s10953-018-0764-6>.
- [80] P. Arosio, T.P.J. Knowles, S. Linse, On the lag phase in amyloid fibril formation, *Phys. Chem. Chem. Phys.* 17 (2015) 7606–7618.
<https://doi.org/10.1039/C4CP05563B>.
- [81] S. Tzotzos, A.J. Doig, Amyloidogenic sequences in native protein structures, *Protein Science* 19 (2010) 327–348.
<https://doi.org/10.1002/pro.314>.
- [82] E. Žerovnik, V. Stoka, A. Mirtič, G. Gunčar, J. Grdadolnik, R.A. Staniforth, D. Turk, V. Turk, Mechanisms of amyloid fibril formation – focus on domain-swapping, *The FEBS Journal* 278 (2011) 2263–2282. <https://doi.org/10.1111/j.1742-4658.2011.08149.x>.
- [83] A.A. Kulikova, A.A. Makarov, S.A. Kozin, Roles of zinc ions and structural polymorphism of β -amyloid in the development of Alzheimer's disease, *Mol Biol* 49 (2015) 217–230.
<https://doi.org/10.1134/S0026893315020065>.
- [84] H. Puchtler, F. Sweat, M. Levine, ON THE BINDING OF CONGO RED BY AMYLOID, *J Histochem Cytochem.* 10 (1962) 355–364.
<https://doi.org/10.1177/10.3.355>.
- [85] T. Konno, K. Murata, K. Nagayama, Amyloid-like aggregates of a plant protein: a case of a sweet-tasting protein, monellin, *FEBS Letters* 454 (1999) 122–126. [https://doi.org/10.1016/S0014-5793\(99\)00789-9](https://doi.org/10.1016/S0014-5793(99)00789-9).
- [86] A.T. Sabareesan, J.B. Udgaonkar, Amyloid Fibril Formation by the Chain B Subunit of Monellin Occurs by a Nucleation-Dependent Polymerization Mechanism, *Biochemistry* 53 (2014) 1206–1217.
<https://doi.org/10.1021/bi401467p>.
- [87] O. Szczepankiewicz, To fold or to fibrillate: serendipity in stability studies, Department of Biophysical Chemistry, Lund University, Lund, 2011.
- [88] M. Delfi, S. Leone, A. Emendato, D. Ami, M. Borriello, A. Natalello, C. Iannuzzi, D. Picone, Understanding the self-assembly pathways of a single chain variant of monellin: A first step towards the design of

- sweet nanomaterials, *International Journal of Biological Macromolecules* 152 (2020) 21–29. <https://doi.org/10.1016/j.ijbiomac.2020.02.229>.
- [89] V. Esposito, F. Guglielmi, S.R. Martin, K. Pauwels, A. Pastore, R. Piccoli, P.A. Temussi, Aggregation Mechanisms of Cystatins: A Comparative Study of Monellin and Oryzacystatin, *Biochemistry* 49 (2010) 2805–2810. <https://doi.org/10.1021/bi902039s>.
- [90] M. Delfi, A. Emendato, S. Leone, E.A. Lampitella, P. Porcaro, G. Cardinale, L. Petraccone, D. Picone, A Super Stable Mutant of the Plant Protein Monellin Endowed with Enhanced Sweetness, *Life* 11 (2021) 236. <https://doi.org/10.3390/life11030236>.
- [91] R. Lucignano, A. Bologna, S. Gramazio, P.-H. Wang, C. Taxis, L.-O. Essen, D. Picone, R. Spadaccini, Unravelling the amyloid aggregation mechanism of the sweet protein Monellin: Insights from circular permuted mutants, *International Journal of Biological Macromolecules* 308 (2025) 142239. <https://doi.org/10.1016/j.ijbiomac.2025.142239>.
- [92] M.J. Bennett, M.P. Schlunegger, D. Eisenberg, 3D domain swapping: A mechanism for oligomer assembly, *Protein Science* 4 (1995) 2455–2468. <https://doi.org/10.1002/pro.5560041202>.
- [93] Y. Liu, D. Eisenberg, 3D domain swapping: As domains continue to swap, *Protein Science* 11 (2002) 1285–1299. <https://doi.org/10.1110/ps.0201402>.
- [94] N.M. Mascarenhas, S. Gosavi, Understanding protein domain-swapping using structure-based models of protein folding, *Prog Biophys Mol Biol* 128 (2017) 113–120. <https://doi.org/10.1016/j.pbiomolbio.2016.09.013>.
- [95] N. Nandwani, P. Surana, H. Negi, N.M. Mascarenhas, J.B. Udgaonkar, R. Das, S. Gosavi, A five-residue motif for the design of domain swapping in proteins, *Nat Commun* 10 (2019) 452. <https://doi.org/10.1038/s41467-019-08295-x>.

- [96] W. Fitriana, T. Sakai, L. Duan, K. Hengphasatporn, Y. Shigeta, T. Mashima, T. Uda, E. Hifumi, S. Hirota, Experimental and Computational Studies on Domain-Swapped Structure Stabilization of an Antibody Light Chain by Disulfide Bond Introduction, *J Med Chem* 67 (2024) 22313–22321. <https://doi.org/10.1021/acs.jmedchem.4c02570>.
- [97] M.J. Bennett, M.R. Sawaya, D. Eisenberg, Deposition Diseases and 3D Domain Swapping, *Structure* 14 (2006) 811–824. <https://doi.org/10.1016/j.str.2006.03.011>.
- [98] F. Rousseau, J. Schymkowitz, L.S. Itzhaki, Implications of 3D Domain Swapping for Protein Folding, Misfolding and Function, in: J.M. Matthews (Ed.), *Protein Dimerization and Oligomerization in Biology*, Springer New York, New York, NY, 2012: pp. 137–152. https://doi.org/10.1007/978-1-4614-3229-6_9.
- [99] J.A. Marsh, S.A. Teichmann, Structure, Dynamics, Assembly, and Evolution of Protein Complexes, *Annu. Rev. Biochem.* 84 (2015) 551–575. <https://doi.org/10.1146/annurev-biochem-060614-034142>.
- [100] A. Borgia, K.R. Kemplen, M.B. Borgia, A. Soranno, S. Shammash, B. Wunderlich, D. Nettels, R.B. Best, J. Clarke, B. Schuler, Transient misfolding dominates multidomain protein folding, *Nat Commun* 6 (2015) 8861. <https://doi.org/10.1038/ncomms9861>.
- [101] S. Kesgin-Schaefer, J. Heidemann, A. Puchert, K. Koelbel, B.A. Yorke, N. Huse, A.R. Pearson, C. Uetrecht, H. Tidow, Crystal structure of a domain-swapped photoactivatable sfGFP variant provides evidence for GFP folding pathway, *The FEBS Journal* 286 (2019) 2329–2340. <https://doi.org/10.1111/febs.14797>.
- [102] F. Sittel, A. Jain, G. Stock, Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates, *The Journal of Chemical Physics* 141 (2014) 014111. <https://doi.org/10.1063/1.4885338>.
- [103] O. Fleetwood, M.A. Kasimova, A.M. Westerlund, L. Delemotte, Molecular Insights from Conformational Ensembles via Machine

- Learning, *Biophysical Journal* 118 (2020) 765–780.
<https://doi.org/10.1016/j.bpj.2019.12.016>.
- [104] J. Krzywanski, M. Sosnowski, K. Grabowska, A. Zylka, L. Lasek, A. Kijo-Kleczkowska, Advanced Computational Methods for Modeling, Prediction and Optimization—A Review, *Materials* 17 (2024) 3521.
<https://doi.org/10.3390/ma17143521>.
- [105] V. Esposito, P.A. Temussi, Cystatins: a versatile family, *Bmc* 2 (2011) 95–102. <https://doi.org/10.1515/bmc.2011.001>.
- [106] N. Nandwani, P. Surana, J.B. Udgaonkar, R. Das, S. Gosavi, Amino-acid composition after loop deletion drives domain swapping, *Protein Science* 26 (2017) 1994–2002. <https://doi.org/10.1002/pro.3237>.
- [107] R.A. Staniforth, Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily, *The EMBO Journal* 20 (2001) 4774–4781.
<https://doi.org/10.1093/emboj/20.17.4774>.
- [108] R. Janowski, M. Kozak, E. Jankowska, Z. Grzonka, A. Grubb, M. Abrahamson, M. Jaskolski, Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping, *Nat. Struct Biol.* 8 (2001) 316–320. <https://doi.org/10.1038/86188>.
- [109] R. Janowski, M. Abrahamson, A. Grubb, M. Jaskolski, Domain Swapping in N-truncated Human Cystatin C, *Journal of Molecular Biology* 341 (2004) 151–160.
<https://doi.org/10.1016/j.jmb.2004.06.013>.
- [110] R. Kolodziejczyk, K. Michalska, A. Hernandez-Santoyo, M. Wahlbom, A. Grubb, M. Jaskolski, Crystal structure of human cystatin C stabilized against amyloid formation, *The FEBS Journal* 277 (2010) 1726–1737. <https://doi.org/10.1111/j.1742-4658.2010.07596.x>.
- [111] M. Orlikowska, E. Jankowska, R. Kołodziejczyk, M. Jaskólski, A. Szymańska, Hinge-loop mutation can be used to control 3D domain swapping and amyloidogenesis of human cystatin C, *Journal of Structural Biology* 173 (2011) 406–413.
<https://doi.org/10.1016/j.jsb.2010.11.009>.

- [112] B. Japelj, J.P. Waltho, R. Jerala, Comparison of backbone dynamics of monomeric and domain-swapped stefin A, *Proteins* 54 (2004) 500–512. <https://doi.org/10.1002/prot.10624>.
- [113] E. Žerovnik, Human stefin B: from its structure, folding, and aggregation to its function in health and disease, *Front. Mol. Neurosci.* 15 (2022) 1009976. <https://doi.org/10.3389/fnmol.2022.1009976>.
- [114] E. Dall, J.C. Hollerweger, S.O. Dahms, H. Cui, K. Häussermann, H. Brandstetter, Structural and functional analysis of cystatin E reveals enzymologically relevant dimer and amyloid fibril states, *Journal of Biological Chemistry* 293 (2018) 13151–13165. <https://doi.org/10.1074/jbc.RA118.002154>.
- [115] S. Yang, S.S. Cho, Y. Levy, M.S. Cheung, H. Levine, P.G. Wolynes, J.N. Onuchic, Domain swapping is a consequence of minimal frustration, *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004) 13786–13791. <https://doi.org/10.1073/pnas.0403724101>.
- [116] A. Lafita, P. Tian, R.B. Best, A. Bateman, Tandem domain swapping: determinants of multidomain protein misfolding, *Current Opinion in Structural Biology* 58 (2019) 97–104. <https://doi.org/10.1016/j.sbi.2019.05.012>.
- [117] Y. Yu, S. Lutz, Circular permutation: a different way to engineer enzyme structure and function, *Trends in Biotechnology* 29 (2011) 18–25. <https://doi.org/10.1016/j.tibtech.2010.10.004>.
- [118] D. Das, S.R.K. Ainavarapu, Protein engineering using circular permutation – structure, function, stability, and applications, *The FEBS Journal* 291 (2024) 3581–3596. <https://doi.org/10.1111/febs.17146>.
- [119] F.M. Klis, A. Boorsma, P.W.J. De Groot, Cell wall construction in *Saccharomyces cerevisiae*, *Yeast* 23 (2006) 185–202. <https://doi.org/10.1002/yea.1349>.
- [120] F.M. Klis, S. Brul, P.W.J. De Groot, Covalently linked wall proteins in ascomycetous fungi, *Yeast* 27 (2010) 489–493. <https://doi.org/10.1002/yea.1747>.

- [121] N.A.R. Gow, M.D. Lenardon, Architecture of the dynamic fungal cell wall, *Nat Rev Microbiol* 21 (2023) 248–259. <https://doi.org/10.1038/s41579-022-00796-9>.
- [122] S.M. Bowman, S.J. Free, The structure and synthesis of the fungal cell wall, *BioEssays* 28 (2006) 799–808. <https://doi.org/10.1002/bies.20441>.
- [123] H. Zlotnik, M.P. Fernandez, B. Bowers, E. Cabib, *Saccharomyces cerevisiae* mannoproteins form an external cell wall layer that determines wall porosity, *J Bacteriol* 159 (1984) 1018–1026. <https://doi.org/10.1128/jb.159.3.1018-1026.1984>.
- [124] F.M. Klis, P. Mol, K. Hellingwerf, S. Brul, Dynamics of cell wall structure in *Saccharomyces cerevisiae*, *FEMS Microbiol Rev* 26 (2002) 239–256. <https://doi.org/10.1111/j.1574-6976.2002.tb00613.x>.
- [125] M.S. Vogt, G.F. Schmitz, D. Varón Silva, H.-U. Mösch, L.-O. Essen, Structural base for the transfer of GPI-anchored glycoproteins into fungal cell walls, *Proc. Natl. Acad. Sci. U.S.A.* 117 (2020) 22061–22067. <https://doi.org/10.1073/pnas.2010661117>.
- [126] M. Ecker, R. Deutzmann, L. Lehle, V. Mersa, W. Tanner, Pir Proteins of *Saccharomyces cerevisiae* Are Attached to β -1,3-Glucan by a New Protein-Carbohydrate Linkage, *Journal of Biological Chemistry* 281 (2006) 11523–11529. <https://doi.org/10.1074/jbc.M600314200>.
- [127] R.A. Ribeiro, N. Bourbon-Melo, I. Sá-Correia, The cell wall and the response and tolerance to stresses of biotechnological relevance in yeasts, *Front. Microbiol.* 13 (2022) 953479. <https://doi.org/10.3389/fmicb.2022.953479>.
- [128] F.M. Klis, C.G. De Koster, S. Brul, A Mass Spectrometric View of the Fungal Wall Proteome, *Future Microbiol.* 6 (2011) 941–951. <https://doi.org/10.2217/fmb.11.72>.
- [129] M. Lozančić, B. Žunar, D. Hrestak, K. Lopandić, R. Teparić, V. Mrša, Systematic Comparison of Cell Wall-Related Proteins of Different Yeasts, *J Fungi (Basel)* 7 (2021) 128. <https://doi.org/10.3390/jof7020128>.

- [130] A. Toh-E, T. Oguchi, Y. Matsui, S. Yasunaga, H. Nisogi, K. Tanaka, Three yeast genes, *PIR1*, *PIR2* and *PIR3*, containing internal tandem repeats, are related to each other, and *PIR1* and *PIR2* are required for tolerance to heat shock, *Yeast* 9 (1993) 481–494. <https://doi.org/10.1002/yea.320090504>.
- [131] U.S. Jung, D.E. Levin, Genome-wide analysis of gene expression regulated by the yeast cell wall integrity signalling pathway, *Molecular Microbiology* 34 (1999) 1049–1057. <https://doi.org/10.1046/j.1365-2958.1999.01667.x>.
- [132] V. Mrša, W. Tanner, Role of NaOH-extractable cell wall proteins Ccw5p, Ccw6p, Ccw7p and Ccw8p (members of the Pir protein family) in stability of the *Saccharomyces cerevisiae* cell wall, *Yeast* 15 (1999) 813–820. [https://doi.org/10.1002/\(SICI\)1097-0061\(199907\)15:10A%253C813::AID-YEA421%253E3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0061(199907)15:10A%253C813::AID-YEA421%253E3.0.CO;2-Y).
- [133] J. Ruiz-Herrera, L. Ortiz-Castellanos, Analysis of the phylogenetic relationships and evolution of the cell walls from yeasts and fungi: Fungal wall evolution, *FEMS Yeast Research* 10 (2010) 225–243. <https://doi.org/10.1111/j.1567-1364.2009.00589.x>.
- [134] T. Tanaka, A. Kondo, Cell-surface display of enzymes by the yeast *Saccharomyces cerevisiae* for synthetic biology, *FEMS Yeast Res* (2014) n/a-n/a. <https://doi.org/10.1111/1567-1364.12212>.
- [135] F.M. Jofre, S.D.S. Queiroz, D.A. Sanchez, P.V. Arruda, J.C.D. Santos, M.D.G.D.A. Felipe, Biotechnological potential of yeast cell wall: An overview, *Biotechnology Progress* 40 (2024) e3491. <https://doi.org/10.1002/btpr.3491>.
- [136] N. Balasco, C. Diaferia, G. Morelli, L. Vitagliano, A. Accardo, Amyloid-Like Aggregation in Diseases and Biomaterials: Osmosis of Structural Information, *Front. Bioeng. Biotechnol.* 9 (2021) 641372. <https://doi.org/10.3389/fbioe.2021.641372>.
- [137] Y. Shen, A. Levin, A. Kamada, Z. Toprakcioglu, M. Rodriguez-Garcia, Y. Xu, T.P.J. Knowles, From Protein Building Blocks to Functional

- Materials, ACS Nano 15 (2021) 5819–5837. <https://doi.org/10.1021/acsnano.0c08510>.
- [138] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all, Nat Methods 19 (2022) 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
- [139] D.A. Case, H.M. Aktulga, K. Belfon, D.S. Cerutti, G.A. Cisneros, V.W.D. Cruzeiro, N. Forouzeshe, T.J. Giese, A.W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M.C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M.R. Machado, H.M. Nguyen, K.A. O’Hearn, A.V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T.E. Cheatham, D.R. Roe, A. Roitberg, C. Simmerling, D.M. York, M.C. Nagan, K.M. Merz, AmberTools, J. Chem. Inf. Model. 63 (2023) 6183–6191. <https://doi.org/10.1021/acs.jcim.3c01153>.
- [140] Bio-Rad Laboratories, ChromLab Software, (n.d.). <https://www.bio-rad.com/it-it/product/chromlab-software?ID=MFCVPXIVK>.
- [141] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr D Biol Crystallogr 60 (2004) 2126–2132. <https://doi.org/10.1107/S0907444904019158>.
- [142] E. Gasteiger, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Research 31 (2003) 3784–3788. <https://doi.org/10.1093/nar/gkg563>.
- [143] R. Anandakrishnan, B. Aguilar, A.V. Onufriev, H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, Nucleic Acids Research 40 (2012) W537–W541. <https://doi.org/10.1093/nar/gks375>.
- [144] OriginLab Corporation, OriginPro, (2008). <https://www.originlab.com>.
- [145] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson,

- J.S. Richardson, T.C. Terwilliger, P.H. Zwart, *PHENIX*: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr* 66 (2010) 213–221. <https://doi.org/10.1107/S0907444909052925>.
- [146] D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, B. Hintze, L.-W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, P.D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*, *Acta Crystallogr D Struct Biol* 75 (2019) 861–877. <https://doi.org/10.1107/S2059798319011471>.
- [147] Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8, (2015).
- [148] Schrödinger, LLC, The JyMOL Molecular Graphics Development Component, Version 1.8, (2015).
- [149] Schrödinger, LLC, The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8, (2015).
- [150] VectorBuilder, Sequence Alignment Tool, (n.d.). <https://en.vectorbuilder.com/tool/sequence-alignment.html>.
- [151] GSL Biotech LLC, SnapGene Viewer, (2025). <https://www.snapgene.com/snapgene-viewer>.
- [152] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, *J Comput Chem* 25 (2004) 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- [153] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, *Journal of Molecular Graphics* 14 (1996) 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).

- [154] H.C. Birnboim, J. Doly, A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Research* 7 (1979) 1513–1523. <https://doi.org/10.1093/nar/7.6.1513>.
- [155] GeneDireX, Inc., Plasmid DNA Miniprep Workflow, (n.d.). <https://www.genedirex.com/wp-content/uploads/2021/07/NA005-0100-01.jpg> (accessed January 31, 2026).
- [156] M. Mandel, A. Higa, Calcium-dependent bacteriophage DNA infection, *Journal of Molecular Biology* 53 (1970) 159–162. [https://doi.org/10.1016/0022-2836\(70\)90051-3](https://doi.org/10.1016/0022-2836(70)90051-3).
- [157] D. Hanahan, Studies on transformation of *Escherichia coli* with plasmids, *Journal of Molecular Biology* 166 (1983) 557–580. [https://doi.org/10.1016/S0022-2836\(83\)80284-8](https://doi.org/10.1016/S0022-2836(83)80284-8).
- [158] R.A. Warmack, B.B. Wenke, T. Spatzal, D.C. Rees, Anaerobic cryoEM protocols for air-sensitive nitrogenase proteins, *Nat Protoc* 19 (2024) 2026–2051. <https://doi.org/10.1038/s41596-024-00973-5>.
- [159] Cytiva, His SpinTrap™ Spin Columns – Purification Workflow, Fisher Scientific SAS (n.d.). <https://www.fishersci.fr/shop/products/his-spintrap-spin-columns-kit/10157094> (accessed February 1, 2026).
- [160] Interchim Blo, Troubleshooting His-tagged protein purification on Ni-NTA columns, Interchim (n.d.). <https://blog.interchim.com/troubleshooting-his-protein-ni-nta-columns/> (accessed February 1, 2026).
- [161] Tosoh Bioscience, Principle of Size-Exclusion Chromatography (SEC), (n.d.). <https://www.separations.eu.tosohbioscience.com/Image%20Library/TBG/GlobalNav/SOLUTIONS/HPLC%20-%20SEC/fig1-sec-columns.png> (accessed February 1, 2026).
- [162] U.K. Laemmli, Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4, *Nature* 227 (1970) 680–685. <https://doi.org/10.1038/227680a0>.
- [163] A. Micsonai, F. Wien, N. Murvai, M.P. Nyiri, B. Balatoni, Y.-H. Lee, T. Molnár, Y. Goto, F. Jamme, J. Kardos, BeStSel: analysis site for

- protein CD spectra—2025 update, *Nucleic Acids Research* 53 (2025) W73–W83. <https://doi.org/10.1093/nar/gkaf378>.
- [164] K. Gade Malmos, L.M. Blancas-Mejia, B. Weber, J. Buchner, M. Ramirez-Alvarado, H. Naiki, D. Otzen, ThT 101: a primer on the use of thioflavin T to investigate amyloid formation, *Amyloid* 24 (2017) 1–16. <https://doi.org/10.1080/13506129.2017.1304905>.
- [165] E. Arad, H. Green, R. Jelinek, H. Rapaport, Revisiting thioflavin T (ThT) fluorescence as a marker of protein fibrillation – The prominent role of electrostatic interactions, *Journal of Colloid and Interface Science* 573 (2020) 87–95. <https://doi.org/10.1016/j.jcis.2020.03.075>.
- [166] J.A. Gavira, Current trends in protein crystallization, *Archives of Biochemistry and Biophysics* 602 (2016) 3–11. <https://doi.org/10.1016/j.abb.2015.12.010>.
- [167] A. Bijelic, A. Rompel, Polyoxometalates: more than a phasing tool in protein crystallography, *ChemTexts* 4 (2018) 10. <https://doi.org/10.1007/s40828-018-0064-1>.
- [168] W. Kabsch, XDS, *Acta Crystallogr D Biol Crystallogr* 66 (2010) 125–132. <https://doi.org/10.1107/S0907444909047337>.
- [169] L. Potterton, J. Agirre, C. Ballard, K. Cowtan, E. Dodson, P.R. Evans, H.T. Jenkins, R. Keegan, E. Krissinel, K. Stevenson, A. Lebedev, S.J. McNicholas, R.A. Nicholls, M. Noble, N.S. Pannu, C. Roth, G. Sheldrick, P. Skubak, J. Turkenburg, V. Uski, F. Von Delft, D. Waterman, K. Wilson, M. Winn, M. Wojdyr, *CCP 4 i 2: the new graphical user interface to the CCP 4 program suite*, *Acta Crystallogr D Struct Biol* 74 (2018) 68–84. <https://doi.org/10.1107/S2059798317016035>.
- [170] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J Comput Chem* 25 (2004) 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- [171] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid

- water, *The Journal of Chemical Physics* 79 (1983) 926–935.
<https://doi.org/10.1063/1.445869>.
- [172] S. Izadi, R. Anandakrishnan, A.V. Onufriev, Building Water Models: A Different Approach, *J. Phys. Chem. Lett.* 5 (2014) 3863–3871.
<https://doi.org/10.1021/jz501780a>.
- [173] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics* 81 (1984) 3684–3690.
<https://doi.org/10.1063/1.448118>.
- [174] J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, B.O. Brandsdal, Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm, *Chemical Physics Letters* 384 (2004) 288–294. <https://doi.org/10.1016/j.cplett.2003.12.039>.
- [175] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems, *The Journal of Chemical Physics* 98 (1993) 10089–10092. <https://doi.org/10.1063/1.464397>.
- [176] R. Elber, A.P. Ruymgaart, B. Hess, SHAKE parallelization, *Eur. Phys. J. Spec. Top.* 200 (2011) 211–223.
<https://doi.org/10.1140/epjst/e2011-01525-9>.
- [177] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J. Chem. Theory Comput.* 9 (2013) 3084–3095.
<https://doi.org/10.1021/ct400341p>.
- [178] H. Nguyen, D.A. Case, A.S. Rose, NGLview-interactive molecular graphics for Jupyter notebooks, *Bioinformatics* 34 (2018) 1241–1242.
<https://doi.org/10.1093/bioinformatics/btx789>.
- [179] Q. Liu, L. Li, L. Yang, T. Liu, C. Cai, B. Liu, Modification of the Sweetness and Stability of Sweet-Tasting Protein Monellin by Gene Mutation and Protein Engineering, *BioMed Research International* 2016 (2016) 1–7. <https://doi.org/10.1155/2016/3647173>.
- [180] A. Pica, S. Leone, R. Di Girolamo, F. Donnarumma, A. Emendato, M.F. Rega, A. Merlino, D. Picone, pH driven fibrillar aggregation of the

- super-sweet protein Y65R-MNEI: A step-by-step structural analysis, *Biochimica et Biophysica Acta (BBA) - General Subjects* 1862 (2018) 808–815. <https://doi.org/10.1016/j.bbagen.2017.12.012>.
- [181] T. You, S. Shi, C. Chen, J. Li, J. Xu, M. Ma, P. Zhu, T. Qi, T. Dong, Y. Wang, S. Ye, S. Liu, Structural Basis for the Exceptional Thermal Stability of the Boiling-Resistant Sweet Protein MNEI, *J. Agric. Food Chem.* 73 (2025) 15959–15966. <https://doi.org/10.1021/acs.jafc.4c13158>.
- [182] C.M. Templeton, S. Ostovar Pour, J.R. Hobbs, E.W. Blanch, S.D. Munger, G.L. Conn, Reduced Sweetness of a Monellin (MNEI) Mutant Results from Increased Protein Flexibility and Disruption of a Distant Poly-(L-Proline) II Helix, *Chemical Senses* 36 (2011) 425–434. <https://doi.org/10.1093/chemse/bjr007>.
- [183] L. Yang, K. Zhu, H. Yu, X. Zhang, B. Liu, The Flexible Loop is a New Sweetness Determinant Site of the Sweet-Tasting Protein: Characterization of Novel Sweeter Mutants of the Single-Chain Monellin (MNEI), *Chemical Senses* 44 (2019) 607–614. <https://doi.org/10.1093/chemse/bjz057>.
- [184] S. Bliven, A. Prić, Circular Permutation in Proteins, *PLoS Comput Biol* 8 (2012) e1002445. <https://doi.org/10.1371/journal.pcbi.1002445>.
- [185] R. Spadaccini, O. Crescenzi, T. Tancredi, N. De Casamassimi, G. Saviano, R. Scognamiglio, A. Di Donato, P.A. Temussi, Solution structure of a sweet protein: NMR study of MNEI, a single chain monellin¹¹Edited by R. Huber, *Journal of Molecular Biology* 305 (2001) 505–514. <https://doi.org/10.1006/jmbi.2000.4304>.
- [186] J.R. Hobbs, S.D. Munger, G.L. Conn, Monellin (MNEI) at 1.15 Å resolution, *Acta Crystallogr F Struct Biol Cryst Commun* 63 (2007) 162–167. <https://doi.org/10.1107/S1744309107005271>.
- [187] W.-C. Lo, L.-F. Wang, Y.-Y. Liu, T. Dai, J.-K. Hwang, P.-C. Lyu, CPred: a web server for predicting viable circular permutations in proteins, *Nucleic Acids Research* 40 (2012) W232–W237. <https://doi.org/10.1093/nar/gks529>.

- [188] J.S. Klein, S. Jiang, R.P. Galimidi, J.R. Keefe, P.J. Bjorkman, Design and characterization of structured protein linkers with differing flexibilities, *Protein Engineering Design and Selection* 27 (2014) 325–330. <https://doi.org/10.1093/protein/gzu043>.
- [189] Understanding and applications of Ser/Gly linkers in protein engineering, in: *Methods in Enzymology*, Elsevier, 2021: pp. 1–22. <https://doi.org/10.1016/bs.mie.2020.12.001>.
- [190] S. Leone, F. Sannino, M.L. Tutino, E. Parrilli, D. Picone, Acetate: friend or foe? Efficient production of a sweet protein in *Escherichia coli* BL21 using acetate as a carbon source, *Microb Cell Fact* 14 (2015) 106. <https://doi.org/10.1186/s12934-015-0299-0>.
- [191] M.F. Rega, R. Di Monaco, S. Leone, F. Donnarumma, R. Spadaccini, S. Cavella, D. Picone, Design of sweet protein based sweeteners: Hints from structure–function relationships, *Food Chemistry* 173 (2015) 1179–1186. <https://doi.org/10.1016/j.foodchem.2014.10.151>.
- [192] A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, *Nucleic Acids Research* 46 (2018) W315–W322. <https://doi.org/10.1093/nar/gky497>.
- [193] P.C. Ke, M.-A. Sani, F. Ding, A. Kakinien, I. Javed, F. Separovic, T.P. Davis, R. Mezzenga, Implications of peptide assemblies in amyloid diseases, *Chem. Soc. Rev.* 46 (2017) 6492–6531. <https://doi.org/10.1039/C7CS00372B>.
- [194] Protein Aggregation, in: *Advances in Protein Chemistry and Structural Biology*, Elsevier, 2010: pp. 89–125. [https://doi.org/10.1016/S1876-1623\(10\)79003-9](https://doi.org/10.1016/S1876-1623(10)79003-9).
- [195] F. Donnarumma, S. Leone, M. Delfi, A. Emendato, D. Ami, D.V. Laurents, A. Natalello, R. Spadaccini, D. Picone, Probing structural changes during amyloid aggregation of the sweet protein MNEI, *The FEBS Journal* 287 (2020) 2808–2822. <https://doi.org/10.1111/febs.15168>.

- [196] H. Naiki, K. Higuchi, M. Hosokawa, T. Takeda, Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavine T, *Analytical Biochemistry* 177 (1989) 244–249. [https://doi.org/10.1016/0003-2697\(89\)90046-8](https://doi.org/10.1016/0003-2697(89)90046-8).
- [197] E. Žerovnik, R.A. Staniforth, D. Turk, Amyloid fibril formation by human stefins: Structure, mechanism & putative functions, *Biochimie* 92 (2010) 1597–1607. <https://doi.org/10.1016/j.biochi.2010.05.012>.
- [198] G. Gotte, E. Butturini, I. Bettin, I. Noro, A. Mahmoud Helmy, A. Fagagnini, B. Cisterna, M. Malatesta, Slow Evolution toward “Super-Aggregation” of the Oligomers Formed through the Swapping of RNase A N-Termini: A Wish for Amyloidosis?, *IJMS* 23 (2022) 11192. <https://doi.org/10.3390/ijms231911192>.
- [199] M. Blaber, Conserved buried water molecules enable the B-TREFOIL architecture, *Protein Science* 29 (2020) 1794–1802. <https://doi.org/10.1002/pro.3899>.
- [200] M. Blaber, Variable and Conserved Regions of Secondary Structure in the β -Trefoil Fold: Structure Versus Function, *Front. Mol. Biosci.* 9 (2022) 889943. <https://doi.org/10.3389/fmolb.2022.889943>.
- [201] M. Peydayesh, R. Mezzenga, Protein nanofibrils for next generation sustainable water purification, *Nat Commun* 12 (2021) 3248. <https://doi.org/10.1038/s41467-021-23388-2>.
- [202] S. Chowdhury, N. Sarkar, Exploring the potential of amyloids in biomedical applications: A review, *Biotech & Bioengineering* 121 (2024) 26–38. <https://doi.org/10.1002/bit.28569>.
- [203] M. Nabi Afjadi, F. Aziziyan, F. Farzam, B. Dabirmanesh, Biotechnological applications of amyloid fibrils, *Prog Mol Biol Transl Sci* 206 (2024) 435–472. <https://doi.org/10.1016/bs.pmbts.2024.04.001>.
- [204] J. Kim, S.-H. Oh, R. Rodriguez-Bobadilla, V.M. Vuong, V. Hubka, X. Zhao, L.L. Hoyer, Peering Into Candida albicans Pir Protein Function and Comparative Genomics of the Pir Family, *Front. Cell. Infect.*

- Microbiol. 12 (2022) 836632.
<https://doi.org/10.3389/fcimb.2022.836632>.
- [205] M. Alvarado, A.E. Moreno-Martínez, M. Micó, J.A. Gómez-Navajas, A. Blázquez-Abellán, V. Mixão, T. Gabaldón, E. Mateo, E. Valentín, P.W.J. De Groot, Wall incorporation of the β -1,3-glucan cross-linking protein Pir1 in the human pathogen *Candida albicans* is facilitated by the presence of two or more Pir repeat units, FEMS Yeast Research 25 (2025) foaf042. <https://doi.org/10.1093/femsyr/foaf042>.
- [206] N. Yang, Z. Yu, D. Jia, Z. Xie, K. Zhang, Z. Xia, L. Lei, M. Qiao, The contribution of Pir protein family to yeast cell surface display, Appl Microbiol Biotechnol 98 (2014) 2897–2905. <https://doi.org/10.1007/s00253-014-5538-5>.
- [207] K.V. Teymennet-Ramírez, F. Martínez-Morales, M.R. Trejo-Hernández, Yeast Surface Display System: Strategies for Improvement and Biotechnological Applications, Front. Bioeng. Biotechnol. 9 (2022) 794742. <https://doi.org/10.3389/fbioe.2021.794742>.
- [208] R.D. Teo, A. Migliore, D.N. Beratan, Mutation effects on charge transport through the p58c iron–sulfur protein, Chem. Sci. 11 (2020) 7076–7085. <https://doi.org/10.1039/d0sc02245d>.

7 Supplement

7.1 List of Publications

- I. S. Hasenjäger, **A. Bologna**, L.-O. Essen, R. Spadaccini, C. Taxis, C-terminal sequence stability profiling in *Saccharomyces cerevisiae* reveals protective protein quality control pathways, *Journal of Biological Chemistry* 299 (2023) 105166. <https://doi.org/10.1016/j.jbc.2023.105166>

Abstract

Protein quality control (PQC) mechanisms are essential for degradation of misfolded or dysfunctional proteins. An essential part of protein homeostasis is recognition of defective proteins by PQC components and their elimination by the ubiquitin-proteasome system, often concentrating on protein termini as indicators of protein integrity. Changes in amino acid composition of C-terminal ends arise through protein disintegration, alternative splicing, or during the translation step of protein synthesis from premature termination or translational stop-codon read-through. We characterized reporter protein stability using light-controlled exposure of the random C-terminal peptide collection (CtPC) in budding yeast revealing stabilizing and destabilizing features of amino acids at positions −5 to −1 of the C terminus. The (de)stabilization properties of CtPC-degrons depend on amino acid identity, position, as well as composition of the C-terminal sequence and are transferable. Evolutionary pressure toward stable proteins in yeast is evidenced by amino acid residues under-represented in cytosolic and nuclear proteins at corresponding C-terminal positions, but over-represented in unstable CtPC-degrons, and vice versa. Furthermore, analysis of translational stop-codon read-through peptides suggested that such extended proteins have destabilizing C termini. PQC pathways targeting CtPC-degrons involved the ubiquitin-protein ligase Doa10 and the cullin-RING E3 ligase SCF^{Das1} (Skp1-Cullin-F-box protein). Overall, our data suggest a proteome protection mechanism that targets proteins with unnatural C termini by recognizing a surprisingly large number of C-terminal sequence variants.

In the context of this work, my contribution was in the selection and computational preprocessing of the C-terminal sequences used for protein stability experiments. In particular, following a discussion with the authors regarding the need to quickly identify subgroups of sequences

with specific C-terminal characteristics, I offered to support the analysis using an automated approach.

I was provided with a dataset containing 6,969 peptide sequences derived from stop codon readthrough experiments in *Saccharomyces cerevisiae*, characterised by variable lengths and the explicit presence of the termination codon (*). The aim was to systematically identify sequences with a C-terminal length equal to or less than six amino acid residues and to generate modified versions of these sequences, in which the C-terminal region was standardised by introducing artificial residues (X), while preserving the correct position of the stop codon.

To this end, I developed a dedicated script in R language, designed to automatically filter and process the entire dataset, calculate the actual length of the sequences before the stop codon and, finally, generate alternative datasets according to predefined criteria.

In particular, I implemented three preprocessing variants in R:

1. forced normalisation of all sequences by adding six X residues to the C-terminal;
2. conditional normalisation applied exclusively to sequences with a length ≤ 6 residues, leaving longer sequences unchanged;
3. alternative variant in which the X residues were inserted upstream of the original sequence to evaluate the effect of the relative position of the artificial tail.

The sequences generated through this computational preprocessing were used by the other authors as a basis for selecting the experimental constructs, allowing the analysis to be restricted to a set of informative C-termini and significantly accelerating the subsequent experimental phase. My contribution therefore provided essential methodological support for the design of the experiments, translating a qualitative biological criterion into a reproducible and controlled computational procedure.

- II. **A. Bologna**, G. Graziano, *Remarks on the hydration entropy of polar and nonpolar species*, *Journal of Molecular Liquids* 391 (2023) 123437. <https://doi.org/10.1016/j.molliq.2023.123437>

Abstract

Around room temperature, experimental hydration entropy values are large negative, regardless of the polar or nonpolar nature of the solute. They cannot be rationalized by claiming that an enhancement of water tetrahedrality occurs upon solute insertion, because neutron diffraction measurements and computer simulations failed to detect such a structural enhancement in the hydration shell of several solutes. The negative hydration entropy values and their trend versus the solute molecular volume can be rationalized by recognizing that solute insertion causes a reduction in the number of configurations accessible to water molecules for simple geometric reasons. This entropy reduction is due to the solvent-excluded volume effect upon cavity creation, and can be estimated by means of analytic relationships provided by classic scaled particle theory. The agreement between the calculated estimates of the entropy change upon cavity creation and the experimental hydration entropy values is good. This result is discussed in depth.

In this work, the main contribution to the theoretical conception and interpretative approach was provided by Prof. Dr. Giuseppe Graziano. My contribution focused on quantitative support for the proposed analysis, particularly in the application of the analytical relationships of scaled particle theory and in the systematic comparison with experimental data available in the literature.

Specifically, I organised and processed the hydration entropy data for solutes of different nature and molecular size, implementing dedicated spreadsheets for the evaluation of the theoretical quantities associated with the creation of cavities in the solvent and for direct comparison with the experimental values. This work made it possible to verify the quantitative consistency between the entropic estimates derived from geometric considerations of excluded volume and the trend observed experimentally. I also contributed to the preparation of the tables and graphs used to illustrate the theory-experiment comparison and participated in the critical review of the manuscript, particularly regarding

the clarity of the presentation of the quantitative results and their physicochemical interpretation.

- III. P.-H. Wang, Y. Hosokawa, J. C Soares, H.-J. Emmerich, V. Fuchs, N. Caramello, S. Engilberge, **A. Bologna**, C.J. Rosner, M. Nakamura, M. Watad, F. Luo, S. Owada, T. Tosha, J. Kang, K. Tono, Y. Bessho, E. Nango, A.J. Pierik, A. Royant, M.-D. Tsai, J. Yamamoto, M. Maestre-Reyna, L.-O. Essen, *Redox-State-Dependent Structural Changes within a Prokaryotic 6-4 Photolyase*, *J. Am. Chem. Soc.* 147 (2025) 16084–16098. <https://doi.org/10.1021/jacs.4c18116>

Abstract

*Photolyases repair UV damage to DNA by using absorbed blue light. Within the photolyase/cryptochrome superfamily (PCSF), a major subgroup consists of prokaryotic (6-4) photolyases. These enzymes rely on flavin adenine dinucleotide (FAD) as a catalytic cofactor, besides an ancillary antenna chromophore, and a [4Fe-4S] cluster with yet unknown function. For the prokaryotic 6-4 photolyase of *Caulobacter crescentus*, we investigated structural changes associated with its different redox states by damage-free crystallography using X-ray free-electron lasers. EPR and optical spectroscopy confirmed redox-dependent structural transitions, including the formation of an oxidized [4Fe-4S]³⁺ cluster with the dynamic cleavage of a single iron-sulfur bond. Photoreduction to the catalytic FADH⁻ state alters the flavin binding site at the proximal aromatic pair Y390/F394 that is part of the electron transport pathway. Upon oxidation, the observable structural transitions of the protein matrix around the [4Fe-4S] cluster may affect DNA binding and are consistent with the much-debated role of the iron-sulfur cluster in DNA-binding proteins for quenching electron holes.*

My contribution to this work developed over a period of about six months and focused on computational modelling of the protein using molecular dynamics simulations, with the aim of investigating the structural effects associated with the different redox states of the FAD cofactor and the iron-sulphur cluster. This activity was carried out in close collaboration with Prof. Dr. Lars-Oliver Essen, in the context of an integrative analysis aimed at supporting the interpretation of structural and spectroscopic data.

Specifically, I focused on the parameterisation of the [4Fe-4S] cluster and the different redox states of the system, readapting and generalising scripts previously developed for the study of the effects of mutation on charge transport in proteins containing iron-sulphur clusters. [208] This work required a critical analysis of existing parameters and their adaptation to the specific context of *Caulobacter crescentus* photolyase (6-4), in order to ensure a consistent description of electronic and structural interactions under different redox conditions. Based on this parameterisation, I set up and performed full-atom molecular dynamics simulations for different redox states of the flavin cofactor, including the oxidised FAD, FAD^{•−}, FADH[−] and FADH[•] configurations, each representative of a distinct phase of the catalytic mechanism. The simulations were aimed at exploring the dynamic response of the active site and the surrounding protein environment to changes in redox state, with particular attention to local interactions between FAD and solvent molecules, including the possible modulation of the accessibility of the flavin N5 site.

In the initial submission of the manuscript, these molecular dynamics results were included to support the interpretation of the experimental data. However, during the review process, on the recommendation of a referee, it was considered that a quantitatively adequate description of the observed phenomena would require more advanced quantum treatment (QM or QM/MM), and that classical MD simulations, while informative from a qualitative point of view, were not sufficiently conclusive to be retained in the final version of the article. For this reason, the section on molecular dynamics was removed from the published manuscript.

Nevertheless, the work carried out contributed substantially to the exploratory and interpretative phase of the project, providing preliminary indications on the structural sensitivity of the flavin site and the surrounding microenvironment to changes in redox state, and

supporting the internal discussions that guided the final methodological choices of the work.

- IV. R. Lucignano, **A. Bologna**, S. Gramazio, P.-H. Wang, C. Taxis, L.-O. Essen, D. Picone, R. Spadaccini, *Unravelling the amyloid aggregation mechanism of the sweet protein Monellin: Insights from circular permuted mutants*, *International Journal of Biological Macromolecules* 308 (2025) 142239. <https://doi.org/10.1016/j.ijbiomac.2025.142239>

Abstract

Protein amyloid aggregates, once regarded solely as pathological hallmarks of human neurodegenerative diseases, have recently gained attention for their potential in biotechnological applications. Among others, MNEI and its variants, initially developed as single-chain derivatives of the sweet protein monellin, also serve as valuable models for studying protein fibrillary aggregation. In this work, we have characterized three circular permuted mutants of MNEI obtained joining the N- and C-termini of MNEI with linkers of different length and restoring the splitting of the polypeptide chain of native monellin. All proteins are well folded but have a different propensity to form oligomeric structures in solution and aggregation rates comparable to or faster than MNEI, as indicated by Thioflavin-T binding assays. Transmission Electron Microscopy (TEM) studies indicate that only Perm1, the mutant with the longest linker, forms fibrillar aggregates. X-ray structures of the mutants show that they crystallize as domain-swapped dimers. Molecular dynamics study highlights potential hot spots controlling the ordered aggregation process of Perm1. Our data support the idea that the formation of a domain-swapped dimer does not favour the formation of fibrillar aggregates and highlight circular permutation as a valuable tool to build nanostructured biomaterials.

This work constitutes the main subject of this doctoral thesis and is discussed in detail in the Results section.

- V. C. J. Rosner, F. Trunk, E. Seiler, **A. Bologna**, V. Fuchs, H.-J. Emmerich, M. Samir Kamel Bekhit, Y. Ibrahim, J. Wachtveitl, P.-H. Wang, L.-O. Essen, *Prokaryotic (6-4) photolyases depend on a one-way bridge for blue-light triggered electron transfer*. Submitted to *PNAS* (May 2026), under review.

Abstract

*Flavin-dependent enzymes, such as photolyases and cryptochromes, harness blue light to drive photoreduction for the repair of UV-damaged DNA and signaling. Here, we investigate the photoreduction dynamics and protonation pathways of the prokaryotic 6-4 photolyase from *Caulobacter crescentus* (Cc(6-4)) using ultrafast UV/Vis spectroscopy, X-ray crystallography and molecular dynamics. The Cc(6-4) absorption spectrum is dominated by its three cofactors: FAD, an [4Fe-4S]²⁺ cluster, and DLZ. FAD photoreduction proceeds via a super-exchange mechanism, with tyrosine serving as primary electron bridge, facilitating forward electron transfer from a tryptophan, the second residue in the highly conserved electron transfer triad. Protonation, crucial for stabilizing the reduced FAD^{•-}, is mediated by residue E402 and a nearby water molecule. Mutagenesis of this residue to glutamine blocks protonation, highlighting its critical role. Finally, the photochemical activity of the [4Fe-4S] cluster is reflected by sub-picosecond oxidation thus making the cluster to a second light-driven electron injector besides the fully reduced flavin cofactor. Given the absence of an ultimate electron acceptor, this cluster undergoes fast recombination within 1.5 ps. Overall, prokaryotic (6-4) photolyases depend for forward, but not backward electron transfer on a unique arrangement of the proximal aromatic residue allowing super-exchange, while protonation of the FAD^{•-} state requires a transient protonation pathway.*

To understand how the change in the redox state of FAD affects the dynamics of the active site in prokaryotic photolyases, I performed molecular dynamics simulations for the three functionally relevant states of the cofactor: the oxidised form (FAD_{ox}), the anionic semiquinone form (FAD^{•-}) and the neutral semiquinone form (FADH[•]). The aim was to observe how each electronic configuration modifies the overall conformational stability of the protein, the flexibility of the surrounding loops and, above all, the dynamic relationship between the flavin N5 atom and the conserved water molecule located in its immediate vicinity. In all three systems, the protein structure stabilises rapidly within the first 50-100 ns of simulation, converging on RMSD plateaus between 1.2 and 1.6 Å. This behaviour suggests that the protein maintains robust structural coherence even when the redox state of the cofactor changes: the “core” of photolyase does not undergo radical reorganisation,

suggesting that any functional differences arise from local changes rather than extensive restructuring.

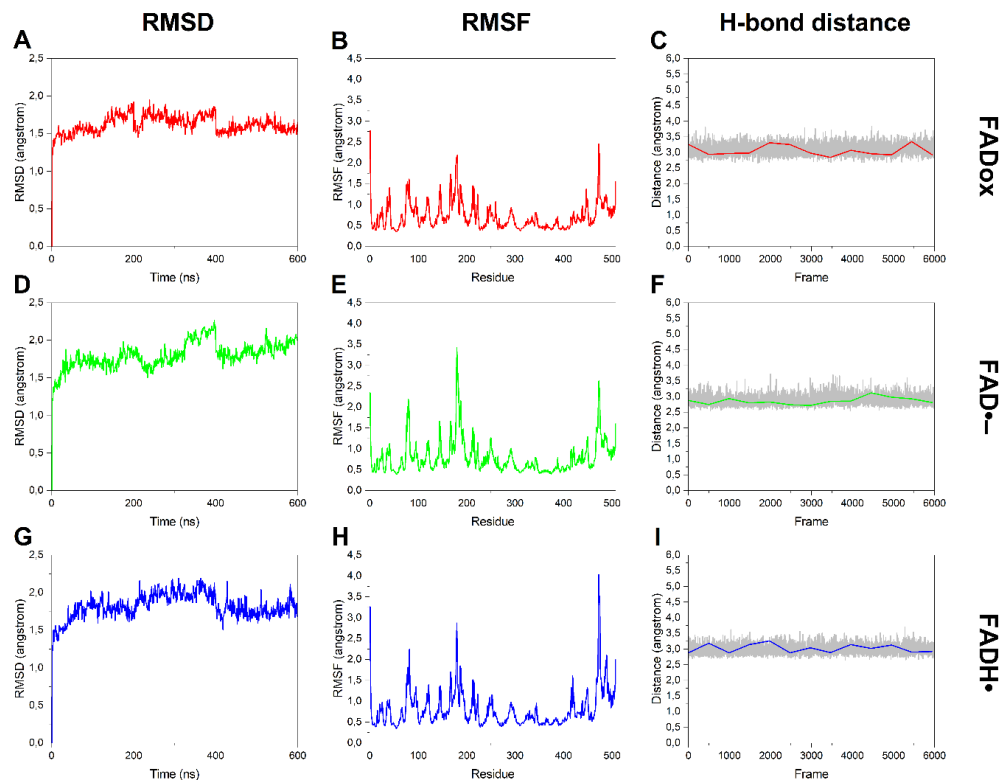


Figure 94. Conformational and dynamic analysis of hydrogen interactions in the three redox states of the FAD cofactor (FADox, FAD•⁻, FADH•) in 200 ns MD trajectories. (A-C) oxidized state (FADox); (D-F) anionic semiquinoid state (FAD•⁻); (G-I) neutral semiquinoid state (FADH•). Left column (A, D, G): root mean square deviation (RMSD) calculated from the initial structure after minimization and equilibration, indicating that all three systems reach a stable conformational plateau after 50-100 ns, with RMSD values between 1.2 and 1.6 Å. Middle column (B, E, H): root mean square fluctuation (RMSF) per residue, showing increased flexibility in N- and C-terminal regions (residues 1-10 and 500-508), and redox-dependent fluctuations in specific loop regions. In FADox, enhanced mobility is observed at residues 179-181 (AEN) and 472-474 (GNG); in FAD•⁻, this extends to 78-80 (PDN), 179-184 (AENRRK), and 472-475 (GDGR); in FADH•, fluctuations are more limited but still appreciable at residues 80 (G), 179-181 (AEN), and 472-475 (GDGR). Right column (C, F, I): time evolution of the dynamic distance between the N5 atom of FAD and a conserved water molecule (Water1) over the 6000 frames per trajectory. The gray line represents the global frame-by-frame trend, while the superimposed coloured line shows the moving average over

time windows of 495 frames, adopted to highlight significant redox-dependent interactions.

However, analysis of residue-by-residue fluctuations reveals a much more complex picture.

Some regions show marked sensitivity to the state of FAD, in particular three recurring segments: the loop around residues 179-184, the one proximal to 78-80, and the terminal region 472-475. In the oxidised system, these oscillations remain limited, while in the anionic form FAD^{•-} these loops become significantly more mobile, almost as if the charge distribution of the cofactor temporarily “loosens” the local constraints. In the neutral form FADH[•], this behaviour partially readjusts: the flexibility is no longer as extreme as in FAD^{•-}, although it shows signs of a livelier dynamic than the oxidised form.

Taken together, these differences paint a coherent picture: the loops that delimit the flavin site respond measurably to the redox state, suggesting a potential role for these micro-movements in the fine-tuning of electron transfer.

One of the most interesting points to emerge from the simulations concerns the dynamic distance between the N5 atom of FAD and a conserved water molecule, already indicated in the literature as a key element for proton activation and semiquinone stability. To track its behaviour over time, I analysed all 6,000 frames of each trajectory, then applied a moving window of 495 frames to isolate significant trends from background noise.

The average distance between N5 and water remains stable overall but shows small differences that align with the redox logic of FAD. In the oxidised form, water remains slightly more distant, as if the local environment were more open.

In the anionic semiquinone, on the other hand, the fluctuation increases and the water-N5 relationship becomes more dynamic: a behaviour compatible with the idea of a greater predisposition to proton exchange or micro-repositioning of the hydrogen network. Finally, in the neutral

form, the distance seems to “contract” slightly and, above all, stabilise, suggesting that the neutralisation of the semiquinone brings about a sort of “calming reorganisation” of the microenvironment.

Taken together, these results indicate that the dynamics of stored water is not a mere structural detail, but a potentially crucial component of the energy landscape governing the unidirectional electron transfer proposed in the manuscript.

- VI. **A. Bologna**, M. Alkheder Alahmad, M. Watad, P.-H. Wang, A. V. Constantin Nemeth, E. Seiler, H.-U. Mösch, R. Spadaccini, L.-O. Essen, *Insight into proteomics and the structural dynamics of PIR proteins in Saccharomyces cerevisiae*. Manuscript in preparation.

This work constitutes the second part of this doctoral thesis. Details are described in the Results section.

7.1.1 Substantial contributions to collaborative research not reflected in authorship

This section reports experimental contributions developed within collaborative research activities carried out during this doctoral project that were not reflected in the authorship of the corresponding publications.

In the context of scientific research, the attribution of authorship is intended to acknowledge substantial intellectual and experimental contributions. However, in collaborative settings, the alignment between the effective contribution provided and its formal recognition may not always be fully achieved. For this reason, it is important to ensure that all significant contributions remain clearly documented within the academic record.

The work presented in this section includes experimental activities personally performed by the author, encompassing the optimisation of biochemical protocols, the production and characterisation of protein samples, and the preparation of material subsequently employed in downstream analyses within collaborative projects.

The purpose of this section is therefore to provide a transparent and traceable account of these contributions, independently of their formal attribution in authorship, by reporting the experimental procedures, optimisation strategies, and results obtained.

- I. A. Just, N. Niemann, P. Gnau, D. Kock, T. Heimerl, L.-O. Essen, A. Batschauer, N. Morgner, *Reversible Lichtinduzierte Oligomerisierung von Arabidopsis-Cryptochrom 1 und Antagonismus durch BIC1: Einblicke durch Zeitaufgelöste Native Massenspektrometrie*, *Angew. Chem. Int. Ed.* (2026), <https://doi.org/10.1002/ange.202525792>

Abstract

Cryptochromes

Cryptochromes (CRYs) are blue-light photoreceptors that mediate light-dependent signalling in plants. Here, we uncover the molecular mechanism underlying blue-light activation of the *Arabidopsis thaliana* cryptochrome 1 photolyase homology region (CRY1-PHR) using time-resolved native mass spectrometry combined with kinetic modelling. This approach enables direct monitoring of light-driven complex formation with temporal and molecular resolution. We show that blue-light activation of CRY1-PHR follows a reversible assembly pathway in which monomers rapidly form dimers that further assemble into tetramers. A quantitative two-step kinetic model captures the dynamic interplay between light-induced oligomerization and thermal disassembly. Strikingly, ATP accelerates tetramer formation and stabilizes oligomers by tuning the underlying photochemistry of the flavin adenine dinucleotide (FAD) chromophore. In contrast, the Blue-light Inhibitor of Cryptochromes 1 (BIC1) acts as a potent antagonist. BIC1 binds to CRY1-PHR even in the dark, with significantly increased affinity under blue light, thereby inhibiting oligomerization and actively disassembling pre-formed tetramers. This disassembly is light-independent and occurs regardless of CRY's redox state. Together, these findings provide a kinetic and mechanistic framework for reversible blue-lighting by plant CRYs and highlight how opposing regulators precisely modulate photoreceptor activation at the molecular level.

The work presented in this section documents the experimental results I generated as part of a collaboration dedicated to the study of the AtCRY1 photoreceptor and its regulator BIC1. All the expression, purification and biochemical analysis experiments reported here were

performed by me personally and provided protein material and preliminary data that were used in the development of the project.

Although I personally generated the experimental data used in the project, I was not included among the authors of the currently available manuscript, in which the authorship of my results is not recognised. I am therefore including the experimental work I carried out in this section so that it remains clearly documented.

To begin the biochemical characterisation of AtCRY1, I expressed the recombinant protein in *E. coli* using two different strains, BL21(DE3) Gold and BL21(DE3) Star, to compare their yield and solubility. After cell lysis, I purified all samples by affinity chromatography on Ni-NTA, using an imidazole gradient with washes at 25 mM and 50 mM and subsequent elution at 250 mM. SDS-PAGE of the fractions (pellet, total lysate, soluble fraction and flow-through) confirmed (Figure 89) good protein expression and a net enrichment in the soluble component, an essential condition for proceeding with size exclusion chromatography.

Before SEC, I brought the samples to a concentration of approximately 0.714 mg/mL. During this phase, a particularly concentration-sensitive behaviour emerged: when approaching or exceeding 1 mg/mL, AtCRY1 tended to show evident oligomeric heterogeneity. To obtain monodisperse material, I therefore repeated four independent runs on a Superdex 200 Increase 16/600 column. I collected and analysed the fractions corresponding to the main peak by SDS-PAGE, confirming the presence of the monomeric form.

At the end of the four separations, I combined the monodisperse fractions, obtaining a total volume of 2.8 mL of AtCRY1 at a final concentration of 0.982 mg/mL. This material, in optimal purity conditions, was stored at 4 °C and prepared for shipment to the collaborating group responsible for the Cryo-EM experiments. (Figure 90)

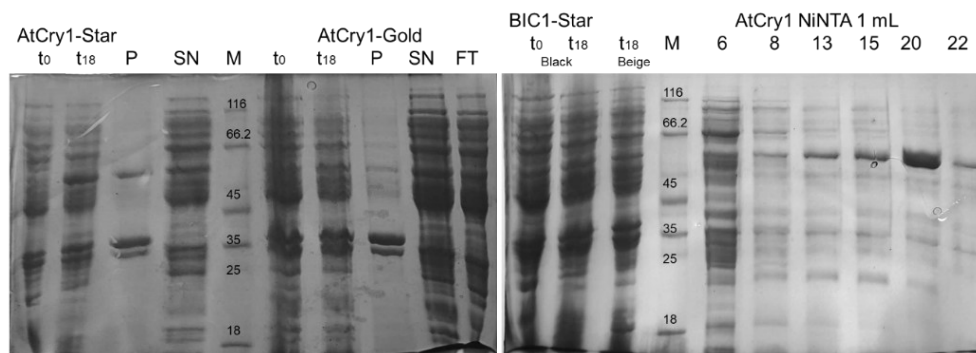


Figure 89. SDS-PAGE analysis of recombinant expression and Ni-NTA purification of AtCRY1 and BIC1. In the left panel, is shown AtCRY1 expressed in *E. coli* BL21(DE3) Star and BL21(DE3) Gold; lanes correspond to total lysate before induction (t0), induced culture after 18 h (t18), pellet (P), soluble fraction (SN), and, where indicated, flow-through (FT), with molecular weight markers (M, kDa). A distinct band at the expected molecular weight of AtCRY1 (~66 kDa) is clearly visible after induction and is enriched in the soluble fraction in both strains, indicating efficient expression and favourable solubility. In the right panel, BIC1 expression (t0 and t18; pellet colour indicated as black and beige) and representative Ni-NTA elution fractions of AtCRY1 (lanes 6-22) are shown. The target protein is predominantly recovered in the high-imidazole fractions, confirming specific binding to the resin and effective affinity purification, yielding material suitable for subsequent chromatographic and biochemical analyses.

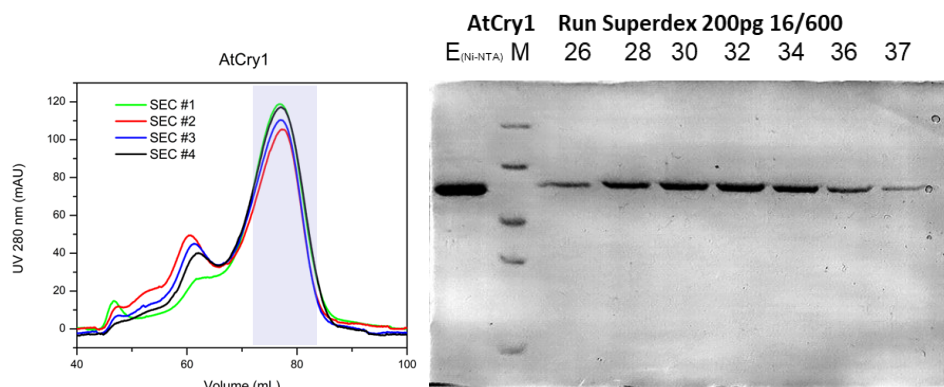


Figure 90. Size-exclusion chromatography (Superdex 200 Increase 16/600) profile of AtCRY1 and SDS-PAGE analysis of collected fractions. The left panel shows four independent SEC runs (SEC #1-#4) performed at ~0.7 mg/mL, displaying a reproducible main peak corresponding to the monomeric species (shaded area). The right panel reports SDS-PAGE of the Ni-NTA elution (E), molecular marker (M), and representative fractions across the main peak (26-37), confirming enrichment of the main peak.

monomeric AtCRY1 with high purity. Monodisperse fractions were pooled and concentrated for subsequent Cryo-EM analyses.

In parallel with AtCRY1, I also worked on the production of BIC1, expressed in *E. coli* BL21(DE3) Star. After lysis, I purified the protein using Ni-NTA, washing at 20 mM and 50 mM imidazole and final elution at 250 mM. SDS-PAGE showed good enrichment of the target protein, with a clean profile already after the affinity phase. (Figure 91)

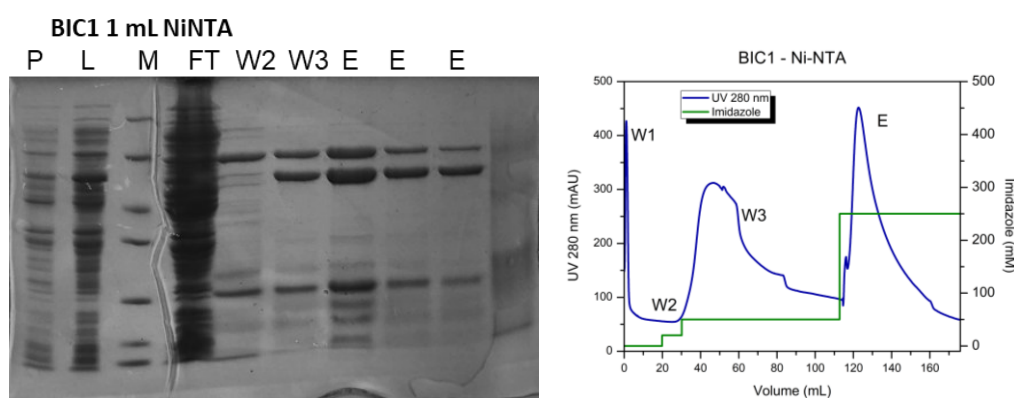


Figure 91. SDS-PAGE and chromatographic profile of BIC1 purified by Ni-NTA affinity chromatography. Lanes show pellet (P), lysate (L), molecular marker (M), flow-through (FT), washes at 20 mM and 50 mM imidazole (W2, W3) and elutions at 250 mM imidazole (E). A clear enrichment of the target protein is observed in the elution fractions, with limited contamination already after the washing steps. The corresponding UV 280 nm trace and imidazole gradient confirm efficient binding to the resin and selective recovery of BIC1 in the high-imidazole peak.

I then proceeded with size exclusion chromatography on a Superdex 75 pg 16/600 column. Due to the high sensitivity of BIC1 to loading concentration, I performed two independent SEC runs. The two separations yielded distinct elution profiles; from each, I collected the main fractions, which were then analysed by SDS-PAGE to verify homogeneity and purity. (Figure 92)

The two final pools reached concentrations of 1,159 mg/mL (Run #1, 1.11 mL) and 1,408 mg/mL (Run #2, 0.754 mL), respectively. After aliquoting into 50 μ L volumes, I froze approximately 70% of the material

in liquid nitrogen, keeping the remainder at 4 °C. The sample from Run #1 was then used for native mass spectrometry experiments at the collaborating laboratory in Frankfurt and Cryo-EM at the Academia Sinica (Taipei, Taiwan). (Figure 93)

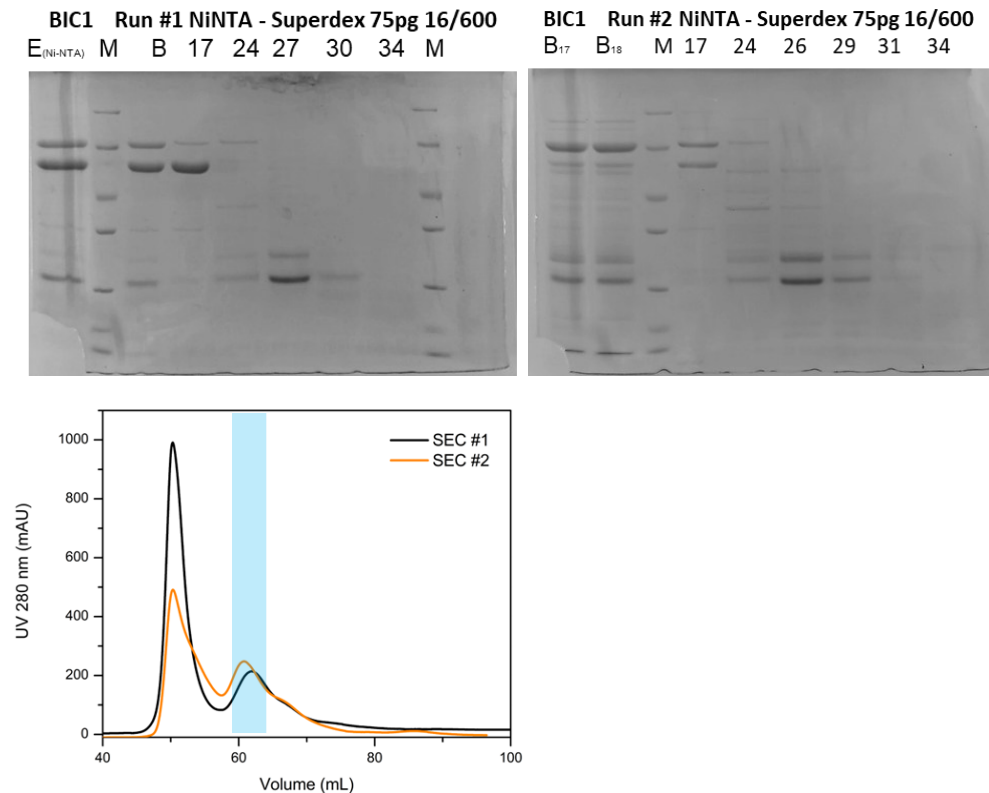


Figure 92. SDS-PAGE analysis and SEC chromatographic profiles of BIC1 following Ni-NTA purification and separation on a Superdex 75 pg 16/600 column. The upper panels (left and right) show the Ni-NTA elution (E), molecular marker (M) and selected fractions collected from two independent SEC runs, confirming the presence of BIC1 in the main elution fractions with comparable band patterns between runs. The lower panel reports the corresponding UV 280 nm chromatograms (SEC #1 and SEC #2), highlighting differences in peak shape and intensity and indicating the fractions selected (shaded area) for pooling based on monodispersity and purity.

All the activities described, from expression to purification, from optimisation to SEC runs, to the preparation of samples for Cryo-EM and native-MS, constitute the work I personally carried out as part of the collaboration on the AtCRY1-BIC1 system.

The protein materials and biochemical information I generated were effectively used in the subsequent phases of the project.

Pool BIC1 – AtCry1 for Taiwan

M BIC1 #1 #2 M AtCry1

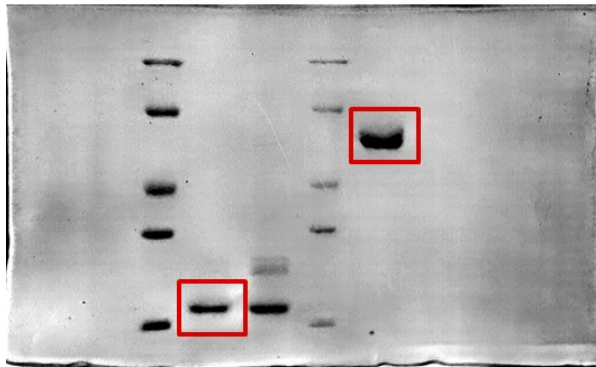


Figure 93. SDS-PAGE analysis of the final pooled fractions obtained from the two independent SEC runs (bottom panel in Figure 92). Molecular weight markers are shown alongside the concentrated pools, with the BIC1 and AtCRY1 bands highlighted. Both preparations show a predominant band at the expected molecular weight with minimal detectable contaminants, confirming high purity. Once aliquoted, the proteins were placed in a light-shielded box, frozen, and ready for subsequent analysis by native mass spectrometry and Cryo-EM.

7.2 Abbreviations

Abbreviation	Meaning
& co.	<i>And co-workers</i>
°C	<i>Degrees Celsius</i>
3D	<i>Three-Dimensional</i>
AcOH	<i>Acetic Acid</i>
AF2	<i>AlphaFold2</i>
AF3	<i>AlphaFold3</i>
Amp	<i>Ampicillin</i>
APS	<i>Ammonium Persulfate</i>
Au	<i>Absorption Units</i>
bp	<i>Base Pair</i>
Ca²⁺	<i>Calcium ion</i>
CD	<i>Circular Dichroism</i>
CI95	<i>95% Confidence Level</i>
CP	<i>Circular Permutation</i>
C-terminal	<i>Carboxy-Terminal</i>
CV	<i>Column Volume</i>
Da	<i>Dalton</i>
DAB	<i>3,3'-diaminobenzidine</i>
DNA	<i>Deoxyribonucleic Acid</i>
DSSP	<i>Define Secondary Structure of Proteins</i>
e.g.	<i>exempli gratia</i>

ECL	<i>Enhanced Chemiluminescence</i>
<i>et al.</i>	<i>and others</i>
EtOH	<i>Ethanol</i>
FAD	<i>Flavin Adenine Dinucleotide</i>
FAO	<i>Food and Agriculture Organization</i>
fu	<i>Fraction Unfolded</i>
g	<i>Gram</i>
GFP	<i>Green Fluorescent Protein</i>
GPI	<i>Glycosylphosphatidylinositol</i>
h	<i>Hour/s</i>
HIS₆	<i>Hexahistidine Tag</i>
HPC	<i>High-Performance Computing</i>
<i>i.e.</i>	<i>id est</i>
IPTG	<i>Isopropyl-β-D-1-thiogalactopyranoside</i>
K	<i>Kelvin</i>
K₂HPO₄	<i>Dipotassium Hydrogen Phosphate</i>
Kan	<i>Kanamycin</i>
kDa	<i>Kilodalton</i>
KH₂PO₄	<i>Potassium Dihydrogen Phosphate</i>
KOAc	<i>Potassium Acetate</i>
L	<i>Liter</i>
Lys	<i>Lysozyme</i>
M	<i>Molar</i>

MD	<i>Molecular Dynamics</i>
Mg²⁺	<i>Magnesium ion</i>
min	<i>Minute</i>
mM	<i>Millimolar</i>
MS	<i>Mass Spectrometry</i>
MSA	<i>Multiple Sequence Alignment</i>
NaOAc	<i>Sodium Acetate</i>
NGC	<i>Next-Generation Chromatography</i>
NMR	<i>Nuclear Magnetic Resonance</i>
N-terminal	<i>Amino-Terminal</i>
OD	<i>Optical Density</i>
PAE	<i>Predicted Aligned Error</i>
PAGE	<i>Polyacrylamide Gel Electrophoresis</i>
PCA	<i>Principal Component Analysis</i>
PDB	<i>Protein Data Bank</i>
PEG	<i>polyethylene glycol</i>
PIR	<i>Protein whit Internal Repetitions</i>
pLDDT	<i>predicted Local Distance Difference Test</i>
PMSF	<i>Phenylmethanesulfonyl Fluoride</i>
QM/MM	Quantum Mechanics / Molecular Mechanics
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SD	Standard Deviation

SDS	<i>Sodium Dodecyl Sulfate</i>
SEC	Size-Exclusion Chromatography
TEM	Transmission Electron Microscopy
TEV	Tobacco Etch Virus Protease
ThT	Thioflavin T
T_m	<i>Melting Temperature</i>
TRIS	Tris(Hydroxymethyl)Aminomethane
UV	Ultraviolet
V	Volt
v/v	Volume Per Volume
VMD	<i>Visual Molecular Dynamics</i>
vs.	<i>Versus</i>
w/v	<i>Weight per Volume</i>
w/w	<i>Weight per Weight</i>
WB	<i>Western blot</i>
WHO	<i>World Health Organization</i>
wt%	<i>Weight Percent</i>
αH	<i>Alpha-Helix</i>
Δ	<i>Change</i>
λ	<i>Wavelength</i>

Table 22. List of abbreviations used in this work.

7.3 List of Amino Acids

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table 23. Standard amino acid codes.

7.4 Curriculum Vitae – Academic career

PhD

- 12/2022-12/2025 PhD student in the research group of Prof. Roberta Spadaccini, Macro-Molecular Chemistry Laboratory, Department of Sciences and Technologies, University of Sannio in Benevento (Italy).
PhD project: “Design and characterisation of engineered proteins for biotechnological applications”.
- 08/2025-10/2025 Teaching and academic support activities for courses in Medical Genetics, Molecular Biology and Cell Biology, as part of the curriculum of the Faculty of Medicine and Surgery at the University of Naples Federico II in Naples (Italy), duplicate track at the University of Sannio in Benevento (Italy).
- 01/2023-03/2024 Visiting PhD student in the research group of Prof. Delia Picone, Structural Biology Laboratory, Department of Chemical Sciences, University of Naples Federico II in Naples (Italy).
- 04/2024-03/2025 Visiting PhD student in Prof. Lars-Oliver Essen’s research group, Laboratory of Structural Biochemistry, Department of Chemistry, Marburg University in Marburg an der Lahn (Germany).

Master's degree

10/2020-10/2022 Department of Sciences and Technologies, University of Sannio in Benevento (Italy), MSc in Genetic and Molecular Biotechnologies. MSc project in Molecular Modelling: "Structure and dynamics of circularly permuted proteins of the MNEI sweet protein".

Bachelor's degree

09/2016-07/2020 Department of Sciences and Technologies, University of Sannio in Benevento (Italy), BSc in Biotechnologies. BSc project in Biology: "The role of ubiquitination in the regulation of DNA damage".

8 Acknowledgements – Academic career

There are many people I'd like to thank for allowing me to reach this goal.

My first thoughts go to Prof. Giuseppe Graziano, whom I've known since the beginning of my studies in Biotechnologies back in 2016. He was the first person to truly encourage me to pursue a PhD, at a time when I didn't yet fully understand what a PhD meant. Over the years, he has remained a discreet yet constant presence, always available to resolve doubts, administrative issues, and matters related to scientific collaborations. Looking back, I recognize how much that initial suggestion inevitably helped change the course of my career.

I owe the concrete beginnings of this academic path to Prof. Roberta Spadaccini, my supervisor. She welcomed me first as a master's student and then as a doctoral researcher, giving me a kind of freedom that proved to be essential to my experimental work, but without imposing any specific lines of research or experimental activities on me, ensuring that the research project embraced as much knowledge, expertise and experience as possible. The work on molecular dynamics simulations of the MNEI protein and its circularly permuted mutants formed the first core of this thesis, but it was also the starting point from which I began to develop my approach to research. I thank her for believing in me, for following my progress even as it shifted between laboratories, cities, and different approaches, and for always encouraging me to expand my computational tools, even toward techniques that remain on the ever-growing list of things to learn.

A significant part of my academic training took place within the Structural Biology laboratory of Prof. Delia Picone. That period brought me face to face with experimental research in its most practical form, and it taught me many things: that a pipette is not just a pipette, that a protocol leaves no space for distractions, and that patience is, in all likelihood, an experimental technique in its own right. I would like to give a mention to Dr. Rosanna Lucignano for the work we shared in the lab and for everything I learnt, directly or indirectly,

during that time. It was not always an easy period, but precisely for that reason it played a significant role in my development.

My experience at Philipps-Universität Marburg, within the Structural Biochemistry laboratory of Prof. Lars-Oliver Essen, represented one of the most intense and formative phases of my PhD. Prof. Essen welcomed me into a research group where ideas could circulate freely, experiments could arise even from sudden inspirations, protocols could be optimised, and where errors were not mere accidents but part of the scientific process, if nothing else, useful for the statistics of negative experiments.

I owe a great deal to Petra Gnau for my practical survival in the Marburg laboratory. She welcomed me warmly from the very first days, even when my experimental debut was more memorable than flawless (a broken flask full of bacteria and a contaminated incubator remain, I believe, a fairly effective way of making oneself remembered), showing a patience that deserves to be listed among laboratory techniques. With her, I shared work, rules, advice, complicated experiments and also much sweeter moments, in the literal sense: the tiramisu we made together will remain among the most beautiful memories of that laboratory. A thought also goes to Kirsten Bubenheimer, for her help, her availability and all the organisational and administrative work that often goes unseen, but without which nothing would really run smoothly.

A special mention in these acknowledgements goes to Po-Hsun Wang, better known as Nigel. His presence was important not only from a scientific perspective, but also on a personal level. He believed in my computational skills, encouraged me to do more, and made me feel part of a larger scientific community. Many days began early, between the gym and the lab, and ended late, having started with the illusion of finishing early. Afterwards, however, there were dinners, conversations, ideas, discussions, trips together and that natural ability that certain people have to become such an important presence. I am truly honoured to have met him and grateful for

the relationship that has continued even after my time in Marburg came to an end.

Christian Joshua Rosner was one of the first people I looked for every day when I walked into the lab. Saying “good morning” and catching up on the experiments had become part of the routine and, of course, so had updates on my “appointments”, which by then had apparently taken on the status of an informal seminar. With him around, the lab felt lighter, the days were more enjoyable, and many evenings were truly memorable. Between afternoon beers, nights out, unlikely karaoke sessions and not always sober covers of *Don't Stop Believin'*, Joshua made Marburg not just a place of work, but a place where you could really feel less alone.

To Mohamed Watad, with whom I shared travels, conversations, days in the lab and many moments outside the lab that gave my time in Germany a deeper meaning. The pleasure trips to Berlin, Dublin, Malta, Madrid and the many other shared experiences were not merely breaks from work, but part of that journey that made Marburg an unforgettable stop and, despite everything, we continue to plan the next trip.

A warm thought goes out to Maisa Alkheder Alahmad, Elisa Seiler, Amon Victor Costantin Nemeth, Hagen Nußhär, Stefan Petschak, Ralf Pöschke, Tanja Ursula Hedderich and all my colleagues at the Essen lab. With some I shared experiments and scientific discussions; with others, lunches in the mensa, coffees, kebabs, barbecues, evenings out and little pieces of everyday life. I also owe Silvana Ruppenthal the sweet memory of little Elizabeth and of a laboratory that, at moments, took on an unexpected family atmosphere; something not to be taken for granted in an environment where it was usually easier to discuss proteins than babies. To all of them goes my gratitude for making that environment lively, welcoming and deeply human.

Even the students and trainees whom I did not mentor personally, but whom I met along the way, played an important role in my personal development, thanks to Angelique Petz, Lilly Briese, Marco Samir Kamel Bekhit, Farhan

Anjum, Schilan Sherin Salim, Leonie Stelzig. Working with them taught me that explaining a protocol is often the best way to realise you haven't understood it well enough, and that an experimental error can become a learning experience, especially when it doesn't ruin anything essential.

Between Benevento and Naples, I shared this journey with colleagues and friends who made many stages of my PhD feel less daunting. To Alessia Esposito and Michele Ginolfi I have to thank for messages, calls, frustrations, updates and that form of collective therapy not officially recognised by PhD regulations, but probably indispensable for making it to the end. To Giuseppina Forgione, Nancy Ferrentino and the other colleagues I've met over the years, I send a sincere thought for the life-saving templates, for the advice and the moments shared, both inside and outside the university.

My deepest thanks go to my family. I must thank my mother and father for everything that does not show up in a thesis but which makes it possible to write one: the departures, the returns, the trips to the airport, the meals prepared, the clothes washed and ironed, the shopping ready for my next departure, the car checked, and the financial, practical and emotional support. These are actions that might seem ordinary, but they are definitely not. They are the invisible framework that allows an individual journey to truly stand on its own. To my brother Luigi, to Sara and to uncle Oreste, I extend equally sincere thanks for the closeness, support and financial help shown over the years.

To my lifelong friends, I owe the evenings, the pizzas, the walks, the laughter and, above all, the rare privilege of conversations in which no one felt the need to mention DNA, protein purification, docking, molecular dynamics or operating protocols. At some moments, it was a form of mental well-being.

Finally, a thought goes out to myself as well. Not because of vanity, but because of honesty. To me, who began this adventure with enthusiasm, inexperience and many insecurities; to me, who has learnt that in a PhD, time follows its own physical laws, so that arriving at the lab very early in the

morning, whilst it is still dark, in order to finish early, can mysteriously lead to leaving late, when it is already dark outside again; to the one who has realised that saying “just one last quick experiment”, “I’ll finish early today” or “I’ve got nothing to do today” are often dangerous sentences; to me, who, despite everything, has remained curious enough to carry on. To me, who has passed through cities, laboratories, mistakes, departures, returns, disappointments, friendships and changes, somehow managing to make it this far.

In conclusion, I would like to express my sincere appreciation to all those who, in their own ways, have been part of this process of personal, scientific and academic growth.

9 Closing statements

This PhD thesis was written between December 2022 and December 2025 under the supervision of Prof. Dr. Roberta Spadaccini.

This PhD thesis was submitted on February 15th, 2026, to the PhD Coordinator, Prof. Dr. Marina Paolucci, and forwarded to the following examiners chose to assess the scientific quality of the work and to approve its admission to the final doctoral examination, in accordance with the University regulations:

- First reviewer: Prof. Dr. Christof Taxis
 - Health and Medical University Erfurt, Erfurt, Germany
- Second reviewer: Prof. Dr. Cameron Mackereth
 - University of Bordeaux, Bordeaux, France

Following the receipt of the reviewers' reports, all the requested revisions were carefully addressed and incorporated into the thesis.

The revised version was subsequently approved by the Supervisor and submitted to the doctoral board for the internal evaluation procedure, preceding the public doctoral defence.

The thesis was presented to the doctoral board on April 17th, 2026. The evaluation resulted in a positive outcome, and this PhD thesis was formally admitted to the public doctoral defence.

The Examination Committee for the final defence in the field of Chemistry-Biology was afterward established as follows:

- Prof. Dr. Federica Cioffi
 - University of Sannio, Benevento, Italy
- Prof. Dr. Daniela Baldantoni
 - University of Salerno, Fisciano, Italy
- Prof. Dr. Giada Lo Re
 - Chalmers University of Technology, Gothenburg, Sweden

The public doctoral defence was held on June 5th, 2026.