UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Bioquímica e Imunologia

VERONICA SILVA VALADARES

CHARACTERIZATION OF THE N-TERMINAL OF TETRACENOMYCIN AROMATASE/CYCLASE INTERACTIONS WITH INTERMEDIATE AND PRODUCT ANALOGS

BELO HORIZONTE
2024

VERONICA SILVA VALADARES

CHARACTERIZATION OF THE N-TERMINAL OF TETRACENOMYCIN AROMATASE/CYCLASE INTERACTIONS WITH INTERMEDIATE AND PRODUCT ANALOGS

Tese apresentada ao Programa de Pós Graduação em Bioquímica e Imunologia da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Ciências.

Orientador: Mariana Torquato Quezado de Magalhães, PhD

Coorientador: Adolfo Henrique de Moraes Silva, PhD

BELO HORIZONTE
2024

043 Valadares, Veronica Silva.
Characterization of the N-terminal of Tetracenomycin aromatase/cyclase interactions with intermediate and product analogs [manuscrito] / Veronica Silva Valadares. – 2024.
69 f. : il. ; 29,5 cm.

Orientador: Mariana Torquato Quezado de Magalhães. Coorientador: Adolfo Henrique de Moraes Silva.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Bioquímica e Imunologia.

1. Bioquímica e imunologia. 2. Biotecnologia. 3. Enzimas. 4. Tetraciclina. 5. Aromatase. 6. Policetídeos. I. Magalhães, Mariana Torquato Quezado de. II. Silva, Adolfo Henrique de Moraes. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 577.1

Ficha catalografica elaborada pela bibliotecária Fabiane C M Reis - CRB 6 2680

Universidade Federal de Minas Gerais
Curso de Pós-Graduação em Bioquímica e Imunologia ICB/UFMG
Av. Antônio Carlos, 6627 – Pampulha
31270-901 – Belo Horizonte – MG
e-mail: pg-biq@icb.ufmg.br (31)3409-2615

ATA DA DEFESA DA TESE DE DOUTORADO DE VERONICA SILVA VALADARES. Aos vinte e nove dias do mês de agosto de 2024 às 08:45 horas, reuniu-se no Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, a Comissão Examinadora da tese de Doutorado, indicada *ad referendum* do Colegiado do Curso, para julgar, em exame final, o trabalho intitulado "Characterization of the N-Terminal of Tetracenomycin Aromatase/Cyclase Interactions with Intermediate and Product Analogs", requisito final para a obtenção do grau de Doutor em Ciências: Bioquímica. Abrindo a sessão, a Presidente da Comissão, Profª. Mariana Torquato Quezado de Magalhães, da Universidade Federal de Minas Gerais, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: Dr. Ronaldo Alves Pinto Nagem (Universidade Federal de Minas Gerais), aprovada; Dr. Vinícius Gonçalves Maltarollo (Universidade Federal de Minas Gerais), aprovada; Dra. Cristiane Dinis Ano Bom (Universidade Federal do Rio de Janeiro), aprovada; Dra. Gisele Cardoso de Amorim (Universidade Federal do Rio de Janeiro), aprovada; Dr. Adolfo Henrique de Moraes Silva - Coorientador (Universidade Federal de Minas Gerais), aprovada; Dra. Mariana Torquato Quezado de Magalhães - Orientadora (Universidade Federal de Minas Gerais), aprovada. Pelas indicações a candidata foi considerada:

(X) APROVADA
( ) REPROVADA

O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 29 de agosto de 2024.

Dr. Ronaldo Alves Pinto Nagem (UFMG)

gov.br Documento assinado digitalmente
RONALDO ALVES PINTO NAGEM
Data: 25/10/2024 10:17:11-0300
Verifique em https://validar.iti.gov.br

Dr. Vinícius Gonçalves Maltarollo (UFMG)

gov.br Documento assinado digitalmente
VINICIUS GONCALVES MALTAROLLO
Data: 21/10/2024 12:20:39-0300
Verifique em https://validar.iti.gov.br

Dra. Cristiane Dinis Ano Bom (UFRJ)

gov.br Documento assinado digitalmente
CRISTIANE DINIS ANO BOM
Data: 18/10/2024 13:31:42-0300
Verifique em https://validar.iti.gov.br

Dra. Gisele Cardoso de Amorim (UFRJ)

gov.br Documento assinado digitalmente
GISELE CARDOSO DE AMORIM
Data: 24/09/2024 09:10:54-0300
Verifique em https://validar.iti.gov.br

Dr. Adolfo Henrique de Moraes Silva - Coorientador (UFMG)

Documento assinado digitalmente
gov.br
ADOLFO HENRIQUE DE MORAES SILVA
Data: 25/10/2024 17:42:52-0300
Verifique em https://validar.iti.gov.br

Dra. Mariana Torquato Quezado de Magalhães - Orientadora (UFMG)

Documento assinado digitalmente
gov.br
MARIANA TORQUATO QUEZADO DE MAGALHAE
Data: 30/10/2024 13:58:06-0300
Verifique em https://validar.iti.gov.br

AGRADECIMENTOS

Este trabalho foi desenvolvido em meio à realização de um sonho. Foi um processo repleto de planejamento, dedicação, aprendizado, discussões, experimentos, extensões, colaborações, viagens, desafios, estresses, ansiedades, dificuldades, superações e de muito trabalho. Sou grata e orgulhosa por eu ter tido a força e determinação de ter trabalhado e terminado este projeto do jeito que deu. Tudo isso tudo aconteceu graças às muitas trocas e vivências com pessoas especiais.

Agradeço ao meu orientador, professor e amigo Adolfo Moraes, por ter acreditado em mim, incentivado e guiado, fornecendo espaço e tempo para o meu desenvolvimento, me apresentado diversas oportunidades e dado suporte em todos os momentos do meu doutoramento.

Agradeço à professora Mariana Quezado por ter me acolhido e ajudado com o meu processo acadêmico.

Agradeço ao professor Haribabu Arthanari, por ter me recebido de maneira tão querida em seu laboratório em Harvard e ter me proporcionado experiências e momentos que impactaram positivamente a minha vida.

Agradeço a Ana Granja e ao Luan Carvalho por terem colaborado tão formidavelmente comigo, contribuindo significativamente nesse projeto. E aos amigos e colegas de laboratório, Diego, Pedro, Philipe, Gustavo, Yan, Brunno, Aline, prof Tiago e prof Amanda, Joana, Krishna, Ricarda, Thibault, Maxim, Sandra, Trey, Abhilash, Srdan e Sebastian, por terem divido tantos conhecimentos, experiências e momentos de vida.

À minha família, - pai, mãe, irmão e minha vó -, agradeço de coração todo o carinho e apoio durante essa longa caminhada acadêmica. Vocês compartilharam comigo o meu sonho e fizeram tudo o possível para ajudar. O suporte de vocês são fundamentais em minha vida.

À minha família de coração - Bruna, Augusto, tia Ana e tio Romário -, agradeço por terem me abrigado em Boston e por sempre me fazerem sentir em casa.

Agradeço aos meus amigos de longa data e aos amigos que fiz no mundo da dança, por fazerem parte da minha vida e me oferecerem ombro amigo em todas as situações durante esse tempo.

Agradeço por ter acesso à Universidade Pública no Brasil, e por todos que fazem parte dessas instituições e que me proporcionaram ensino de qualidade e excelência.

“E aprendi que se depende sempre de tanta muita diferente gente.
Toda pessoa sempre é as marcas das lições diárias de outras tantas pessoas.
E é tão bonito quando a gente entende que a gente é tanta gente onde quer que a gente vá.
E é tão bonito quando a gente sente que nunca está sozinho por mais que a gente pense estar.
É tão bonito quando a gente pisa firme nessas linhas que estão nas palmas de nossas mãos,
É tão bonito quando a gente vai à vida nos caminhos onde bate bem mais forte o coração.”
- Gonzaguinha, Caminhos do Coração

“Somewhere something incredible is waiting to be known.”
- Carl Sagan

“The only way to achieve the impossible is to believe that it is possible."
- Lewis Carroll, Alice in Wonderland

Apoio Financeiro

Este trabalho foi realizado com o auxílio financeiro da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradeço por ter tido acesso e suporte da infraestrutura do Laboratório Macromol do Departamento de Química da UFMG, do Laboratório de Ressonância Magnética Multiusuário (LAREMAR) na UFMG, do Centro Nacional de Ressonância Magnética Nuclear de Macromoléculas (CNRMN) na UFRJ, do Laboratório LBM do Departamento de Bioquímica e Imunologia da UFMG, do Laboratório Multiusuário do Departamento de Biofísica da UNIFESP, e da *Harvard Medical School Bio-molecular NMR Facility* e *Dana Faber Cancer Institute NMR Core*.

RESUMO

O N-terminal da Tetracenomicina aromatase/ciclase (TcmN) é uma enzima de *Streptomyces glaucescens* envolvida na ciclização e aromatização de policetídeos. Policetídeos aromáticos constituem uma importante classe de metabólitos secundários produzidos por certas bactérias, fungos e plantas, com diversas atividades biológicas. Exemplos incluem antibióticos, como as tetraciclinas, e fármacos anticâncer, como a doxorrubicina. Existem poucos ensaios experimentais que caracterizam a interação da TcmN com ligantes e, como a TcmN é uma enzima promissora para a produção *in vitro* de policetídeos, compreender o processo de reconhecimento molecular da enzima é fundamental para o entendimento de seu mecanismo catalítico. Este trabalho investiga o mecanismo de reconhecimento molecular da TcmN a substratos, intermediários e produtos, bem como o papel da dinâmica conformacional da TcmN em sua função enzimática. A TcmN foi produzida por expressão heteróloga. Dicroísmo circular (CD) e calorimetria diferencial de varredura (DSC) foram utilizados para avaliar a estabilidade térmica e temporal da enzima. A interação com ligantes foi estudada por experimentos de diferença de transferência de saturação em ressonância magnética nuclear (STD-RMN) de ^{1}H, perturbação de deslocamento químico por RMN de ^{1}H-^{15}N, experimentos de relaxação em RMN de ^{15}N e simulações de dinâmica molecular (MD) em escala de microssegundos. A temperatura de melting (Tm) da TcmN ligada à naringenina, medida por DSC, foi superior à da enzima livre. O sítio de ligação da naringenina, mapeado por experimentos de RMN, localizou-se na cavidade central da TcmN. Dinâmica conformacional na escala de μs–ms foi detectada por experimentos de relaxação em RMN para resíduos da TcmN na cavidade de ligação ao substrato quando ligada à naringenina. Duas conformações predominantes, representando estados de cavidade aberta e fechada, foram observados nas simulações de MD, e os resultados sugerem que a ligação de intermediários desloca a população conformacional da TcmN para o estado aberto. Os resultados fornecem insights sobre a base molecular do reconhecimento de substratos e auxiliam na elucidação do mecanismo catalítico da TcmN.

Palavras-chave: Enzimas, Tetracenomicina, TcmN, Biotecnologia, Aromatase/Ciclase, RMN, Policetídeo.

ABSTRACT

The N-terminal of Tetracenomycin aromatase/cyclase (TcmN) is an enzyme from *Streptomyces glaucescens* involved in polyketide cyclization and aromatization. Aromatic polyketides are an important class of secondary metabolites produced by certain bacteria, fungi, and plants with various biological activities. Examples include antibiotics, tetracycline, and anticancer drugs, such as doxorubicin. There are few experimental assays characterizing the interaction of TcmN with ligands, and because TcmN is a promising enzyme for the *in vitro* production of polyketides, knowing the enzyme's molecular recognition process is critical for understanding the TcmN's catalytic mechanism. This work investigates the mechanism of TcmN molecular recognition of substrates, intermediates, and products and the role of TcmN conformational dynamics on its enzymatic function. TcmN was produced by heterologous expression. Circular dichroism (CD) and differential scanning calorimetry (DSC) were used to evaluate the enzyme's thermal and temporal stability. Interaction with ligands was assessed by ^{1}H nuclear magnetic resonance (NMR) saturation-transfer difference (STD) experiments, ^{1}H-^{15}N NMR chemical shift perturbation, NMR ^{15}N relaxation experiments, and microsecond molecular dynamics (MD) simulations. The naringenin-bound TcmN melting temperature, *Tm*, measured from DSC experiments, was higher than the free enzyme one. The naringenin binding site, mapped by NMR experiments, was in the core TcmN cavity. Conformational dynamics on the µs-ms timescale were detected by NMR relaxation experiments for TcmN residues in the substrate-binding cavity for the naringenin-bound TcmN. Two predominant conformers representing opened and closed cavity states were observed in the MD simulations, and the results suggest that the binding of intermediate shifts the TcmN conformation population to an opened state. The results provide insights into the molecular basis of substrate recognition and help delineate the catalytic mechanism of TcmN.

Keywords: Enzymes, Tetracenomycin, TcmN, Biotechnology, Aromatase/Cyclase, NMR, Polyketyde.

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
3D	Three dimensional
Å	Angstrom
ACP	Acyl carrier protein
ARO	Aromatase
C	Carbon
CD	Circular Dichroism
CPMG	Carr-Purcell-Meiboom-Gill sequence
CSP	Chemical Shift Perturbation
CYC	Cyclase
D_2O	Deuterium Oxide
Da	Dalton
DSC	Differential scanning calorimetry
E. coli	*Escherichia coli*
FID	Free induction decay
FPLC	Fast protein liquid chromatography
H	Hydrogen
HCl	Hydrochloric acid
His-tag	Polyhistidine Sequence
Het-NOE	Heteronuclear NOE
HSQC	Heteronuclear Single Quantum Correlation
IPTG	Isopropyl β-D-1- thiogalactopyranoside
KR	Ketoreductases
KS-CLF	Ketosynthase-chain length factor
LB	Luria-Bertani
MD	Molecular dynamics
MHz	Megahertz
N	Nitrogen
NaCl	Sodium chloride
NaI	Sodium Iodide
NAR	Naringenin
NAR-R	R-isomer of Naringenin

NAR-S	S-isomer of Naringenin
NaPO4	Sodium phosphate
NOE	Nuclear Overhauser effect
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
PKS	Polyketide synthase
ppm	parts per million
R_1	Longitudinal relaxation
R_2	Transverse relaxation
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
TcmN	N-terminal of Tcm ARO/CYC protein
Tm*	Observable melting temperature
UV	Ultraviolet
Ala (A)	Alanine
Arg (R)	Arginine
Asn (N)	Asparagine
Asp (D)	Aspartic acid
Cys (C)	Cysteine
Gln (Q)	Glutamine
Glu (E)	Glutamic acid
His (H)	Histidine
Ile (I)	Isoleucine
Leu (L)	Leucine
Lys (K)	Lysine
Met (M)	Methionine
Phe (F)	Phenylalanine
Pro (P)	Proline
Ser (S)	Serine
Thr (T)	Threonine
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
Val (V)	Valine

Table of Contents

1 INTRODUCTION

1.1 Polyketide synthases

Polyketides are secondary metabolites produced by bacteria, fungi, and plants, with a wide range of potential biological activities and pharmaceutical properties. They are synthesized by polyketides synthases (PKSs) from repeated condensation of malonyl-CoA by acetyl coenzyme A, resulting in the formation of a molecular structure of alternating carbonyl and methylene groups with multiple β-hydroxyketone or β-hydroxyaldehyde functional groups. (MOSS; SMITH; TAVERNIER, 1995) (Figures 1 and 2).

Three types of PKSs have been described. Type I PKSs, or modular PKSs, are large multi-domain enzymes with active sites for catalyzing the various stages of polyketide synthesis. Type II PKSs are separate mono and bi-functional enzymes that interact iteratively during aromatic polyketide synthesis (GOMES; SCHUCH; LEMOS, 2013). Type III PKSs are homodimers of ketosynthases that catalyze the condensation of one or several molecules of extender substrate into a starter substrate through iterative decarboxylative Claisen condensation reactions (KATSUYAMA; OHNISHI, 2012).

The minimal functional composition of type II PKSs consists of a ketosynthase (KS or KSα), a chain length factor (CLF or KSβ), an acyl carrier protein (ACP), and a malonyl transferase (MAT). This enzyme complex synthesizes a polyketide chain (Figure 1) The CLF decarboxylates the ACP-bound malonyl and transfers it to the KS(BISANG et al., 1999). Once the starter unit is allocated on the KS and the ACP is reloaded with malonyl, decarboxylation activates the ACP-bound malonyl, allowing the resulting enolate to attach to the growing polyketide on the KS. The collapse of the tetrahedral intermediate frees the KS, resulting in an ACP-bound polyketide extended by one unit. The nascent polyketide, now extended by one acetate unit, is transferred back to the KS through transthioesterification, allowing the extension process to continue(CHEN; RE; BURKART, 2018; RIVERS; LOWELL, 2024). It is believed that the CLF stabilizes and regulates the length of the growing chain, and once the chain reaches the determined length, further transfer back to the KS is not possible, and the full-length ACP-bound polyketide dissociates from the KS-CLF dimer(KEATINGE-CLAY et al., 2004; RIVERS; LOWELL, 2024). Also, additional enzymes, such as cyclases, aromatases, and ketoreductases, act on this chain, producing aromatic and cyclic groups, and in the absence of these tailoring enzymes, the polyketide forms abnormal structures (Figure 2) (TANG et al.,

2003). After releasing the polyketide from the ACP, numerous other enzymes, such as oxidases, transferases, and hydrolases, may act on the chain to make further modifications (WEISSMAN, 2009). The versatility of type II PKSs and a broad array of post-PKS tailoring enzymes make the type II PKS-derived secondary metabolites one of the most structurally diverse groups (ROHR; HERTWECK, 2010). Several compounds produced by PKSs are used clinically, with well-known examples including tetracyclines as antibiotics and doxorubicin and mithramycin as anticancer drugs (WEISSMAN, 2009).

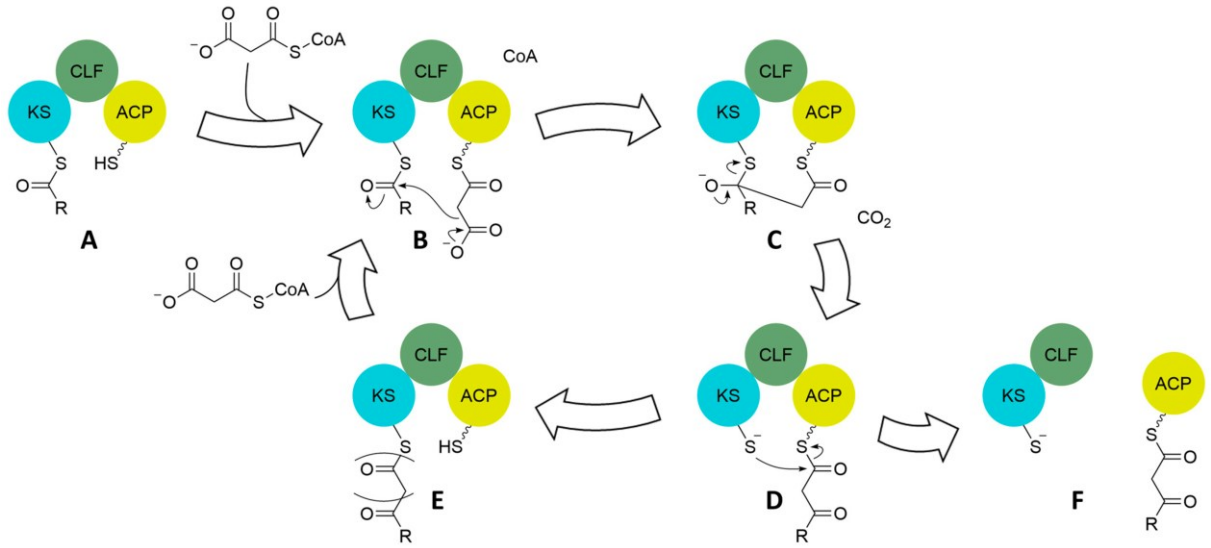

Figure 1 - Claisen-like condensation mechanism of the minimal functional composition of type II PKSs to grow the polyketide chain.

Once the KS is loaded with a starter unit (A) via one of several mechanisms, the ACP interacts with malonyl-CoA or transacylases from primary metabolism to obtain a malonyl for chain extension (B). Decarboxylation activates the malonyl as an enolate on the ACP, which reacts with the thioester on the KS to form a tetrahedral intermediate (C). The tetrahedral extension collapse involves transferring the extended polyketide back to the KS (E) and acquiring another malonyl by the ACP to reenter the extension cycle (E to B). Once the ACP-bound polyketide chain is at the appropriate length, it can no longer be accepted back onto the KS, causing dissociation of the ACP from the KS-CLF dimer (F) and enabling delivery of the polyketide chain to cyclases for tailoring to various ring systems. Adapted from Rivers, M.A.J.; Lowell, A.N. SynBio 2024, 2, 85-111.

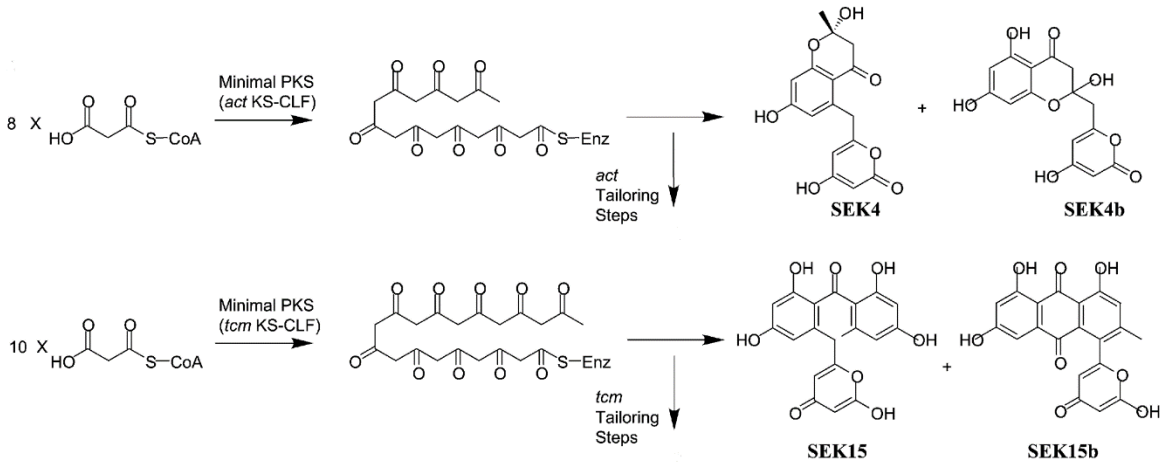

Figure 2 – Steps in the synthesis of polyketides by minimal PKSs.

Minimal PKSs consisting of actinorhodin (act) or tetracenomycin (tcm) KS-CLF can synthesize C16 or C20 polyketides from eight or ten malonyl CoAs, respectively. The minimal PKS initiates polyketide synthesis through decarboxylation of one malonyl-ACP. Additional malonyl ACPs are recruited and condensed with KS-CLF to elongate the polyketide chain until the desired length is reached. Without additional tailoring enzymes, SEK4/4B and SEK15/15B are formed by act and tcm minimal PKS, respectively. Adapted from Tang *et al*. Biochemistry, Vol. 42, No. 21, 2003.

The development of combinatorial biosynthesis by "mixing and matching" different PKS components, including the combinatorial manipulation of ketoreductases (KRs), aromatases (AROs), and cyclases (CYCs), in heterologous expression systems, has resulted in many synthetic polyketides (FITZGERALD et al., 2013; LEŠNIK et al., 2015; UEBERSCHAAR et al., 2013; WEISSMAN; LEADLAY, 2005). However, some challenges still need to be overcome for predictable and reliable polyketide biosynthesis in heterologous hosts. For example, it demands that large multienzyme assemblies be functionally expressed with correct posttranslational modification, their substrates be available *in vivo* in reasonable quantities, and the producer cell be protected against the toxicity of the biosynthetic products (LIU et al., 2019; PFEIFER; KHOSLA, 2001; ZHANG; PAN; TANG, 2017).

In vitro, cell-free platforms have been an attractive methodology for constructing new pharmaceuticals and bioactive molecules (LI; ZHANG; LIU, 2018). However, using the biosynthetic machinery *in vitro* to produce new natural products by controlling functional group formation and cyclization patterns requires detailed knowledge of how the individual enzymes work.

The poly-β-keto intermediate is a very reactive compound, and the type II PKS needs to specifically restrain the polyketide intermediate structure from avoiding strange cyclization

events. Previous studies have established the biosynthetic routes for aromatic polyketides in “reducing” (KR present) and “non-reducing” (KR absent) PKS systems (AMES, BRIAN D. et al., 2011; HERTWECK et al., 2007; RAWLINGS, 1999). In reducing systems, a KR first cyclizes the linear poly-β-ketone from C12 to C7, followed by a C9-carbonyl reduction, and then a di-domain ARO/CYC catalyzes the dehydration of the C9 hydroxyl, followed by first-ring aromatization (CALDARA-FESTIN et al., 2015; JAVIDPOUR et al., 2013). The growing poly-β-ketone intermediate is transported directly from the ketosynthase to the ARO/CYC in a non-reducing system. There are two types of ARO/CYCs in non-reducing systems (Figure 3): mono-domain ARO/CYCs, which are often associated with C9-C14 first-ring cyclization, and di-domain ARO/CYCs, which contain two repeats of the ARO/CYC domain for catalyzing C7-C12 first-ring cyclization (AMES, BRIAN D. et al., 2008; CALDARA-FESTIN et al., 2015; MCDANIEL, R., HUTCHINSON, C. R., KHOSLA, 1995).

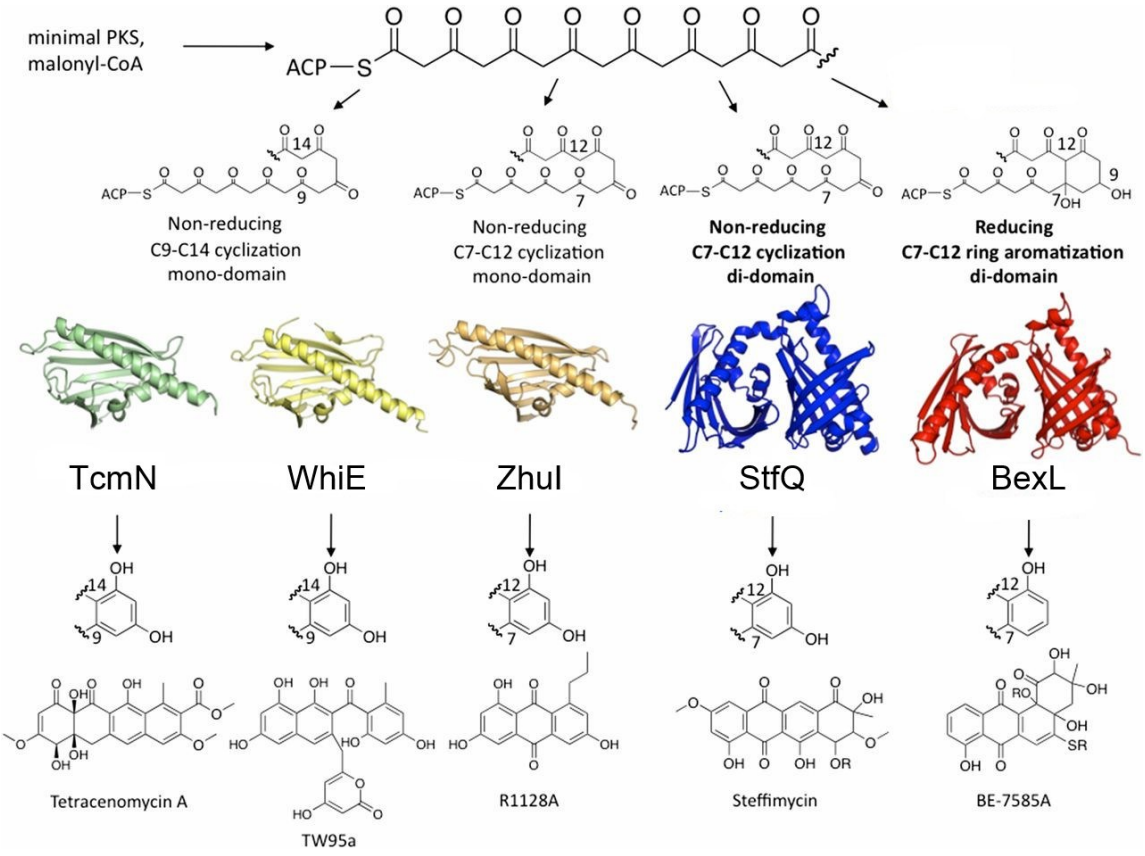

Figure 3 – Schematic diagram of ARO/CYC activity and cyclization specificity in representative type II PKSs.

The monodomain ARO/CYCs TcmN and WhiE act on unreduced polyketide intermediates to generate C9–C14 cyclized and aromatized products. The monodomain ARO/CYC ZhuI and di-domain ARO/CYC StfQ act on unreduced polyketide intermediates to generate C7–C12 cyclized and aromatized products. The di-domain ARO/CYC BexL acts on C9 reduced, C7–C12 cyclized intermediates and catalyzes the aromatization of the C7–12 cyclized ring by dehydration of the C9 hydroxyl group. They were adapted from Caldara-Festin, G. et al., Proceedings of the National Academy of Sciences Dec 2015, 112 (50) E6844-E6851.

1.2 N-terminal domain of Tetracenomycin ARO/CYC (TcmN)

In *Streptomyces glaucescens*, the deducted product of the tcmN gene is a multifunctional protein with two domains; the N-terminal portion acts as an aromatase cyclase, and the C-terminal portion acts as an O-methyltransferase (SUMMERS *et al.*, 1992). The N-terminal of Tetracenomycin aromatase/cyclase (TcmN) is a Bet v 1-like superfamily protein, like the lipid/sterol-binding StAR-related lipid transfer (START) proteins (TSUJISHITA; HURLEY, 2000) and PR-10 family of plant pathogenesis-related proteins (MORAES et al., 2018; RADAUER; LACKNER; BREITENEDER, 2008). The ubiquitous distribution of Bet v 1-related proteins among all superkingdoms suggests that a Bet v1-like protein was already

present in the last universal common ancestor. Bet v 1-like superfamily proteins have diverse biological functions with varying degrees of sequence similarity while maintaining the ancestral fold. For instance, TcmN shares 14% identity with Bet v 1 from *Betula verrucose*. Bet v 1-like proteins can bind several molecules, such as flavonoids, steroids, hormones, and lipids (MOGENSEN et al., 2002; MORAES et al., 2018). To bind to different compounds, transport, participate in catalytic activities, and release products, these Bet v1-like proteins undergo considerable conformational changes (AMES, BRIAN D. et al., 2008; KOFLER et al., 2012; MORAES et al., 2016).

TcmN possesses a β-sandwich fold consisting of a seven-stranded antiparallel β-sheet, a long C-terminal α-helix that runs down the center of the β-sandwich, and two small helices between β1 and β2, forming a helix–loop–helix motif that seals one end of the β-sandwich (AMES, BRIAN DOUGLAS et al., 2008) (Figure 4B). TcmN substrate nascent polyketide intermediates bind to residues within the interior pocket of TcmN, where regiospecificity cyclization and aromatization occur. The cavity has a ~20 Å deep interior with a hydrophobic and polar residues distribution that can accommodate an ACP-bound decaketide (20 carbons) (Figure 4A-D). The cavity has a narrow neck defined by residues W65, F88, L129, Y35, T132, T133, E34, and R82 (Figure 4C). Pocket residues W28, F32, W65, S67, R69, M91, and W95 are highly conserved among ARO/CYCs, while residues T54, L93, H128, T132, T133, and N136 are unique to TcmN. R69, T35, W108, and Q110 orient the polyketide chain inside the cavity, while S67 and R82 anchor the polyketide, and the size and shape of the cavity modulate chain folding for C9-C14 first-ring cyclization (Figure 4D). C9-C14 first-ring cyclization and interactions with residues in the narrow neck region orient C7 and C16 so that C7 carbonyl hydrogen bonds with R82. The C16 methylene is positioned near E34 and Y35, favoring C7-C16 second-ring cyclization (Figure 4C). The TcmN cavity might accommodate up to three linearly fused rings. The third ring cyclization likely happens spontaneously (Figure 4A). After that, the polyketide is transferred to TcmI, a cyclase that is structurally and functionally distinct from TcmN, where fourth-ring cyclization occurs to produce a compound that undergoes further enzymatic modification to yield the anthracycline antibiotic Tetracenomycin (AMES, BRIAN D. et al., 2008; LEE; AMES; TSAI, 2012).

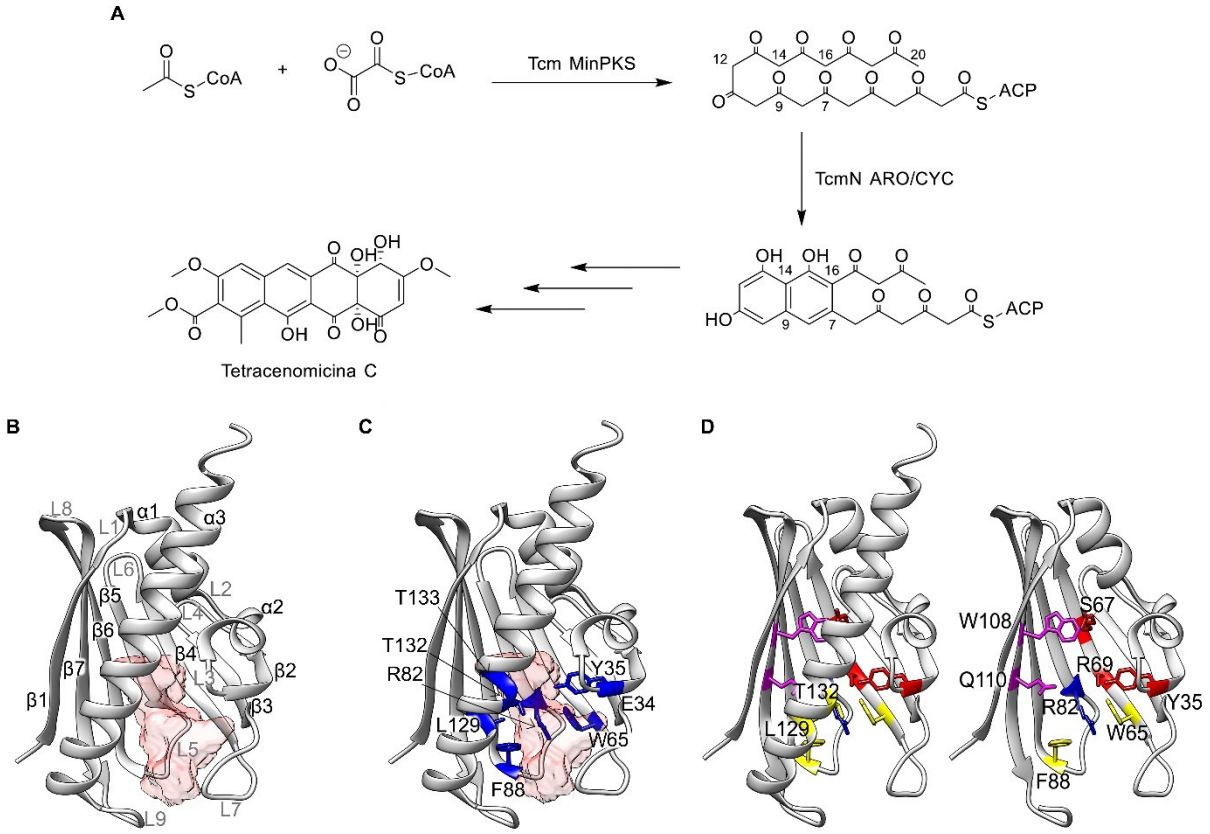

Figure 4 – Polyketide synthesis and TcmN structure.

(A) Simplified biosynthetic pathway for Tcm C. Adapted from Ames, B. et al., Proceedings of the National Academy of Sciences Apr 2008, 105 (14) 5349-5354; © 2008 by The National Academy of Sciences of the USA. (B) Ribbon representation of the TcmN crystal structure (PDB ID: 2RER). The TcmN cavity is identified by the surface being colored red. (C) Blue sticks represent sidechains of residues located at the narrow neck in the cavity; (D) Critical residues involved in the cyclization and aromatization of the polyketide chains are shown as sticks. Residues involved in the first and second cyclization, polyketide folding, and cavity neck are colored red, blue, magenta, and yellow, respectively. In the representation on the right, alpha-helix 3 has been removed for better visualization of the residues.

Conformational changes can be important for enzyme-substrate molecular recognition (MORAES et al., 2018). Sampling opened states is likely needed for substrate binding to TcmN and may also be critical for enzyme activity. The tcmN cavity comprises several hydrophobic residues that could trigger aggregation when exposed to solvents. Previously, our research group showed that TcmN presented higher thermal stability at low protein and higher salt concentrations. Thermal denaturation was irreversible, leading to amorphous but not amyloidogenic aggregates. NMR studies indicated that transient self-interaction may promote aggregation at higher protein concentrations. Analysis of conformational dynamics by NMR showed high flexibility of L5, L7, and L9, regulating the opening and closing of the catalytic cavity (VALADARES et al., 2021). There are few experimental assays characterizing the

interaction of TcmN with ligands, and understanding the molecular recognition process of the enzyme by substrates, intermediates, and products is critical for understanding the enzyme's catalytic mechanism. In the present study, the conformational changes involved in the molecular recognition of substrates are being investigated. The biochemical knowledge obtained from this work will be a step towards defining the molecular rules for the specificity of the cyclization of TcmN and other ARO/CYCs. It may help design mutants with optimal properties for biotechnological applications.

2 OBJECTIVES

This work aims to investigate the mechanism of TcmN molecular recognition of substrates, intermediates, and products and understand the role of TcmN conformational dynamics on its enzymatic function.

Specific objectives of the project included:

1. Optimize the expression and purification protocol, as well as the “His-tag” cleavage strategy;
2. Acquire and assign NMR spectra necessary to complete the backbone resonance assignment of TcmN;
3. Characterize structurally and thermodynamically the interaction of TcmN to compounds analogous to the substrate and product;
4. Characterization of TcmN conformational states by MD simulations.

3 MATERIALS AND METHODS

3.1 Expression and purification of TcmN

The sequence of N-terminally His6-tagged TcmN, coding for residues 1-173, was commercially synthesized and cloned by GenScript (Piscataway, USA) in a pET28a plasmid. *E. coli* BL21 (DE3) cells were transformed with the plasmid and cultivated at 37°C in 50 mL Luria Bertani (LB) medium with 50 $\mu g.mL^{-1}$ kanamycin for 16 h under shaking at 200 rpm. Cells were harvested by centrifugation and used to inoculate 500 mL of LB. M9 minimal media containing ^{15}N-ammonium chloride and ^{13}C-glucose was used to produce ^{15}N and ^{15}N-$_{13}C$-labeled samples (Sigma-Aldrich, USA). Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when the OD at 600 nm was 0.8. Cultures were incubated for 16 h at 22 °C under shaking at 200 rpm. Cells were centrifugated for 40 min at 4°C and 4500 rpm in a Thermo Scientific TX-750 Swinging Bucket Rotor (Thermo Fisher, USA). Harvested cells were resuspended in 50 mM sodium phosphate pH 8, containing 300 mM NaCl and 10 mM imidazole. The cell suspension was lysed and then centrifuged at 11000 rpm for 40 min at 4°C. The protein was purified from the soluble fraction by Ni^{2+} affinity chromatography using a HisTrap HP Ni-NTA column (GE Healthcare, USA) in 50 mM sodium phosphate pH 8, containing 300 mM NaCl and 40 mM imidazole, and eluted with an imidazole gradient up to 500 mM. His 6-tag was cleaved by thrombin using and following the protocol of the Thrombin CleanCleave Kit (Sigma-Aldrich, USA). The purified protein was dialyzed against 20 mM sodium phosphate pH 7, containing 150 mM NaCl. Purity was determined by SDS–PAGE, and concentration was estimated by UV-vis spectroscopy at 280 nm, using an extinction coefficient of 37470 $M^{-1}.cm^{-1}$, obtained from the TcmN primary sequence using protparam (https://web.expasy.org/protparam/). The primary amino acid sequence of the TcmN is:

MAARTDNSIV VNAPFELVWD VTNDIEAWPE LFSEYAEAEI LRQDGDGFDF RLKTRPDANG RVWEWVSHRV PDKGSRTVRA HRVETGPFAY MNLHWTYRAV AGGTEMRWVQ EFDMKPGAPF DNAHMTAHLN TTTRANMERI KKIIEDRHRE GQRTPASVLP TELHAQQLLL

The protein was heterologously expressed with an N-terminal His-tag (sequence: HHHHHHLVPRGSH). After cleavage of the His-tag, three additional residues (GSH) remained at the N-terminus of the protein.

3.2 CD spectroscopy

CD measurements were performed on a J-815 spectropolarimeter (JASCO Inc., USA) equipped with a Peltier temperature controller, located at the Centro de Laboratórios Multiusuários do Instituto de Ciências Biológicas – CELAM-ICB of the Federal University of Minas Gerais. Spectra were acquired using a 0.01 cm path length cuvette, from 190 to 260 nm, using a scanning speed of 100 nm*min-[1], wavelength increment of 0.5 nm, and three scans. Thermal denaturation was assessed by measuring the CD intensity at 208 nm over temperatures varying from 25 to 90°C, with increments of 0.5°C and an equilibration period of 5 s. The normalized intensity, I_N, was estimated by:

$$I_N = \frac{I_{CD} - I_{CD,U}}{I_{CD,F} + I_{CD,U}},$$

Where I_{CD}, $I_{CD,\,U}$, and $I_{CD,\,F}$ are the molar residue ellipticity (MRE) values at 208 nm at the measurement temperature, 25°C (folded state), and 90°C (unfolded state), respectively. The apparent melting temperature, Tm*, analysis was performed with the server CalFitter (MAZURENKO et al., 2018). The two-state N⇌D model of protein denaturation was used, where N and D are the native and denatured states. Fitting of the experimental denaturation curves was performed using:

$$Signal_{spec}(T) = f_N(T)x_N(T) + f_D(T)x_D(T)$$

where $f_N(T)$ and f_D(T) are the coefficients for pre- and post-transition baseline linear dependency over temperature:

$$f_N(T) = B_{N0} + B_{N1}T$$
$$f_D(T) = B_{D0} + B_{D1}T$$

where B_{N0} and B_{N1} are the pre-transition and B_{D0} and B_{D1} the post-transition fitting constants; x_N and x_D are the fractions of TcmN in native and denatured states, respectively, given by:

$$x_N = \frac{1}{1+K}$$
$$x_D = \frac{K}{1+K},$$

Where K is the equilibrium constant, and T_m is the melting temperature at which K = 1 (MAZURENKO et al., 2018).

3.3 DSC

TcmN DSC thermograms were recorded using a MicroCal PEAQ differential scanning calorimeter (Malvern, UK), located at the Laoratório Multiusuário do Departamento de Biofísica of the Federal University of São Paulo. DSC thermograms were acquired using 211 μM TcmN in 20 mM sodium phosphate pH 7, 150 mM NaCl. The thermal stability of naringenin-bound TcmN was accessed by acquiring DSC thermograms of TcmN with 1 and 2 mM of Naringenin. Since the naringenin stock solution was prepared in DMSO, 4 and 8% of DMSO were added to the TcmN final solution. To monitor the temperature increment rate on TcmN thermal resistance, the free protein solutions were heated from 20°C to 90°C at constant rates of 0.3°C $\cdot min^{-1}$, 1.5°C $\cdot min^{-1}$ and 2°C $\cdot min^{-1}$ inside a 300 μL capillary cell of the calorimeter. The naringenin binding impact on TcmN thermal resistance was monitored by performing the DSC thermograms with the naringenin-bound TcmN with constant rates of 1°C $\cdot min^{-1}$. The reference cell was filled with the same buffer solution containing the same concentrations of Naringenin and DMSO as in the measurement cell. The reversibility of the thermal transitions was assessed by reheating the sample immediately after the cooling step. The DSC data analysis was carried out using the Origin 2018 software (Originlab Corporation, USA). The thermal stability of the proteins was described by the temperature of the maximum thermal transition (Tm).

3.4 ITC

ITC measurements were performed at 25°C in a buffer containing 20 mM sodium phosphate pH 7.0 and 150 mM NaCl with 4% DMSO using a Affinity ITC (TA Instruments, New Castle, DE) equipped with an autosampler. A concentration of 140 μM protein solution or buffer (control) in the calorimetric cell was titrated by injecting 2.5 μL of 2 mM ligand solution in 350-second intervals with stirring speed at 125 rpm. Resulting isotherms were subtracted against buffer runs and fitted with a single site model to yield thermodynamic parameters of ΔH, ΔS, stoichiometry, and Kd using NanoAnalyze software (TA Instruments).

3.5 NMR spectroscopy

3.5.1 - Protein backbone assignment

NMR experiments were performed at 298 K using Bruker 900 MHz Avance Neo and 800 MHz Avance III spectrometers equipped with TXI 5 mm triple-resonance probe

located at the Jiri Jonas National Center for Nuclear Magnetic Resonance UFRJ, a Bruker Avance Neo 600 MHz spectrometer equipped with a PABBO 5 mm SmartProbe located at the Laboratório de Ressonância Magnética de Alta Resolução (LAREMAR UFMG), and a Bruker Avance III 800 MHz equipped with a TXO cryogenic probe located at the Harvard Medical School Bio-molecular NMR Facility and the Dana-Farber Cancer Institute NMR Core. For the TcmN backbone resonance assignment, the following NMR triple-resonance experiments were acquired and analyzed: ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, HNCO, HN(CA)CO, HNCA, HNCACB, and CBCA(CO)NH. 2D experiments were acquired using the traditional full acquisition schedule, and 3D experiments using non-uniform Poisson Gap Sampling (HYBERTS et al., 2012; HYBERTS; TAKEUCHI; WAGNER, 2010). The spectra were processed using Topspin 3.7 (Bruker Corporation, USA) and NMRpipe (DISEASES; BAY; KONG, 1995) and analyzed using CCPN analysis 2.4.2 (VRANKEN et al., 2005).

3.5.2 – NMR relaxation dispersion experiments

Backbone dynamics of TcmN were characterized by monitoring longitudinal, R_1, and transversal, R_2, relaxation rates measured on a Bruker Avance Neo 600 MHz spectrometer. The overall protein correlation time, τ_c, was calculated from the R_2/R_1 ratios. Dynamics on the μs-ms timescale were monitored by performing NMR relaxation-compensated ^{15}N single-quantum Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments (LORIA; RANCE; PALMER, 1999). CPMG experiments were acquired at 298 K using Bruker Avance Neo 600 MHz spectrometer, equipped with a 5 mm dual-channel multinuclear Smart Probe, located at the Nuclear Magnetic Resonance Laboratory – LAREMAR of the Federal University of Minas Gerais. Experimental R_2 effective rates, $R_{2,eff}$ (s^{-1}), were obtained from ^{1}H-^{15}N correlation spectra intensities using CCPN (VRANKEN et al., 2005), and applying the following equation:

$$R_{2,eff} = \frac{1}{T_{relax}} \ln\left(\frac{I_0}{I_{CPMG}}\right)$$

Where T_{relax} is the relaxation delay of 30 ms, I_{CPMG} is the intensity measured in ^{1}H-^{15}N spectra acquired with CPMG frequency, ν_{CPMG}, ranging from 50 to 1000 Hz. I_0 is the peak intensity obtained with a T_{relax} of 0 ms. Experimental errors were calculated from the standard deviation of noise to the resonance intensities following Loria *et al*. (LORIA; RANCE; PALMER, 1999).

3.4.3 - NMR protein interaction

The interaction of 80 µM of ^{15}N-labeled TcmN was monitored by chemical shift perturbation on the ^{1}H-^{15}N HSQC spectra upon titration with up to 560 µM Naringenin. For analysis of the chemical shift perturbations of ^{1}H and ^{15}N backbone resonances, a weighted average chemical shift change was calculated:

$$CSP_i = \sqrt{\Delta H_i{}^2 + 0.1\Delta N_i{}^2},$$

Where ΔH_i and ΔN_i denote the observed changes of the ^{1}H and ^{15}N chemical shifts for residue i.

1D ^{1}H STD spectra were acquired at 25 °C on a Bruker Avance Neo 600 MHz spectrometer, equipped with a 5 mm dual-channel multinuclear Smart Probe, located at the Nuclear Magnetic Resonance Laboratory – LAREMAR of the Federal University of Minas Gerais. The 600 µL samples were prepared with 20 µM TcmN in sodium phosphate 20 mM, NaCl 150 mM, Naringenin 1 mM, DMSO-d6 4% (v/v), D2O 10% (v/v), and pH 7, corresponding to a protein/ligand concentration ratio of 1/20. The STD build-up curves were obtained using saturation times of 0.25, 0.75, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 s, and data were acquired using a train of 50 ms Gaussian pulses (MAYER; MEYER, 2001). The on- and off-resonance frequencies were 0 and -40 ppm, respectively. STD peak intensity differences were extracted for each experiment with different saturation times to calculate STD-Amplification Factors (ANGULO; NIETO, 2011). STD-Amplification Factors were calculated using the following equation:

$$STDamp = \left(\frac{STD\ diff\ resonance\ intensity}{STD\ off\ resonance\ intensity}\right) \times\ ligand\ excess$$

3.5.3 – Titration experiments

The interaction of TcmN with Naringenin was monitored through titration experiments, wherein the ligand was added to the isotopically ^{15}N-enriched recombinant protein sample. For each titration point, a ^{1}H-^{15}N HSQC spectrum was acquired using the hsqcetgpsi sequence for the experiment acquired at 600 MHz and the hsqcetfpf3gp sequence at 800 MHz. Titration experiments were conducted at 298 K on a Bruker Avance Neo 600 MHz spectrometer, equipped with a 5 mm multinuclear PABBO dual-channel Smartprobe, located at the Laboratório de Ressonância Magnética de Alta Resolução (LAREMAR), and on a Bruker Avance III 800 MHz spectrometer equipped with a TXO cryogenic probe, located at the Harvard Medical School Bio-molecular NMR Facility and the Dana-Farber Cancer Institute NMR Core. The samples for the 600 MHz experiment were in 20 mM phosphate buffer, pH 7,

with 100 mM NaCl. The samples for the 800 MHz experiment were in 20 mM phosphate buffer, pH 7, with 150 mM NaCl and 10% D2O (v/v). The concentration of TcmN and the racemic mixture of naringenin solubilized in DMSO-d^6 at each titration point for the experiments acquired at 600 MHz and 800 MHz are specified in Table 1 and Table 2, respectively.

Table 1 - Concentration of TcmN and Naringenin at each titration point used in the ^1H-^{15}N HSQC experiment acquired at 600 MHz. The dilution of the protein due to the addition of the ligand solution in DMSO-d^6 was considered.

TcmN concentration (μM)	Naringenin concentration (μM)
80.0	0
79.87	41.6
79.74	83.2
79.61	124.8
79.35	208
78.72	416
78.28	561.6

Table 2 - Concentration of TcmN and naringenin at each titration point used in the ^1H-^{15}N HSQC experiment acquired at 800 MHz. The dilution of the protein due to the addition of the ligand solution in DMSO-d^6 was considered.

TcmN concentration (μM)	Naringenin concentration (μM)
50.00	0
49.98	25
49.95	50
49.9	100
49.8	200
49.6	400
49.31	1000
48.83	2000

3.5.4 – TITAN analysis

The NMR signal shape analysis was performed using the TITAN program (WAUDBY et al., 2016) They were implemented in the MATLAB programming language and platform. The analyses were conducted according to the methodology described by Waudby et al.

(WAUDBY et al., 2020). This analysis was done in collaboration with Ana Clara Granja, who, by the time, was an undergraduate student mentored by Professor Adolfo de Moraes and me.

3.6 Molecular dynamics simulations

3.5.1 – Free TcmN MD simulations

Independent replicates (total = 6) of 1 μs MD simulations of free TcmN were carried out using GROMACS 5 (ABRAHAM et al., 2015), utilizing the CHARMM22* (PIANA; LINDORFF-LARSEN; SHAW, 2011) force field and standard TIP3P water model (HEINZELMANN; HENRIKSEN; GILSON, 2017). The TcmN crystal structure (PDB ID: 2RER) (AMES, BRIAN DOUGLAS et al., 2008) was used with protonation states of titratable residues predicted by PropKa (OLSSON et al., 2011) at pH 7. Na^+ and Cl^- were added to achieve charge neutrality and simulate a physiological concentration of 0.15 M. After steepest-descents minimization, ten replicates were initiated with different velocity distributions and propagated for 1 μs each in the NPT ensemble at 298 K and 1 bar, using previously described parameters (KARTTUNEN; CHOY; CINO, 2018). GROMACS analysis tools and in-house scripts were employed to analyze the trajectories. Principal component analysis (PCA) was performed using Python and scikit libraries on the pooled frames from the six replicas with least-squares fitting to the Cα atoms of the initial structure (CINO; CHOY; KARTTUNEN, 2013). Due to high flexibility, the PCA analysis did not consider residues K141 to P155 from the C-terminal. Representative structures of the closed and open states were taken from the centermost frames of the energy minima. Bidimensional histograms were prepared by binning the trajectory projections on the first two principal components and expressed in energy units using Boltzmann's factor (CINO; CHOY; KARTTUNEN, 2016). Hydration variations in the TcmN cavity were analyzed in the following manner: water molecules were classified as located inside the cavity if they were <4.0 Å from the side chains of group A (T22, F32, Y35, L52, L54, T82, F88, L93, T95, W108, Q110, L129, and T133), and outside the cavity, if they were >3 Å from the side chains of group B (R4, D6, F29, R55, H81, T85, Y90, N92, and R134). Water molecules were counted using VMD version 1.9.3 (HUMPHREY; DALKE; SCHULTEN, 1996). The distance between loops L5 and L9 was evaluated by measuring the distance between the Cα of P119 (L9) and R61 (L5). Interpolation of structures to represent the motion along PC1 was performed using gmx tools (ABRAHAM et al., 2015). One hundred interpolated frames were generated and analyzed. Gaussian fitting of the histograms of

the L5-L9 distances and the number of water molecules inside the TcmN cavity was carried out using Scipy (VIRTANEN *et al.*, 2020).

3.5.2 – Bound TcmN MD simulations

To select poses to seed MD simulations to study TcmN and intermediate and product analog interaction, docking simulations were performed using High Ambiguity Driven biomolecular DOCKing (HADDOCK) version: 2.4. TcmN residues E34, Y35, S67, R89, W95, H128, and T133 (predicted by Ames *et al.* 2008 as pocket residues that make contact with ligands) and residues A80, M91, W108, Q110, I140, T132, L129 AND F112 (mapped by CSP experiment with titration of Naringenin to TcmN) were selected as HADDOCK restraints and defined as the protein active and passive residues, respectively, in the rigid-body docking (it0). The related residues were depicted as passive in the semi-flexible docking (it1) and explicit solvent refinement(water)). 10000 poses were generated in it0 and 1000 poses in it1. The best-docked complexes of top-ranked clusters were selected and visualized using Discovery Studio Visualizer (BIOVIA, Dassault Systèmes, San Diego, USA, 2022) and UCSF Chimera (PETTERSEN *et al.*, 2004). Ligand poses from Haddock was aligned using GROMACS tools (gmx trjconv) using TcmN Cα for alignment. Pairwise RMSD, RMSD to the best scored posed, and ligand center-of-mass were calculated with GROMACS tools. Each ligand pairwise RMSD matrix was embedded into a bi-dimensional space using T-distributed Stochastic Neighbor Embedding (t-SNE) (VAN DER MAATEN; HINTON, 2008). Bidimensional data was then visually inspected and clustered with Density-based spatial clustering of applications with noise (DBSCAN) (ESTER *et al.*, 1996) using ε=13 and a minimum of 75 samples to define a cluster. Cluster elements were visually inspected and selected by considering the complementarity to the binding pocket, polar interactions, if any, and the pose score. The TcmN crystal structure (PDB ID: 2RER) (AMES, BRIAN DOUGLAS *et al.*, 2008) and the selected poses of Naringenin and INT12-bound TcmN were prepared in the same way as the simulations described above. Each system had 10 independent replicates of 1 μs MD simulations using GROMACS 5 (ABRAHAM *et al.*, 2015), utilizing the CHARMM36 (HUANG; MACKERELL, 2013) force field. GROMACS analysis tools and in-house scripts were employed to analyze the trajectories.

4 RESULTS AND DISCUSSION

4.1 Free TcmN stability

The previous work done by our research group, cited in section 1.2, was carried out with uncleaved protein, that is, with 6His-tag in the N-terminal of TcmN. However, it was observed that the protein was relatively unstable when stored, and at concentrations above 200 μM it began to show considerable changes in the ^{1}H-^{15}N HSQC NMR spectra, making it challenging to acquire more extended NMR experiments and thus contributing to the limited quality of the 3D experiments obtained for protein signaling. The heterologous expression and purification of TcmN were performed, and the cleavage protocol was followed. TcmN cleavage was completed successfully (Figure 5). The typical yield per 500 mL culture of the TcmN production was 7 mg.

CD and DSC measurements were performed to assess and compare the cleaved TcmN thermal resistance. CD spectra of the protein in the range of 190–250 nm showed the most negative signal at 208 nm. Therefore, the signal at 208 nm was selected to monitor changes in ellipticity during the thermal denaturation experiment, allowing determination of the protein melting temperature (Tm). TcmN thermal plot shows that TcmN is more stable without the 6His-tag, with a Tm* (observable melting temperature) of 54.7°C, compared to the uncleaved TcmN Tm* of 50.5°C (Figure 6A). DSC measurements were carried out at different heating rates to monitor the effect of the temperature rise rate on the Tm of the TcmN. The thermogram analysis shows that at the heating rates of 0.3°C/min, 1.5°C/min, and 2°C/min, the Tm* were 47.5°C, 49°C, and 49.7°C, respectively (Figure 6B). Tm* values increase with the heating rate, and the denaturation process is irreversible. The DSC data did not indicate that aggregation occurs, as there are no signs of the formation of soluble aggregates on the thermogram, and perhaps there is immediate precipitation of the protein after unfolding, as precipitated protein could be seen.

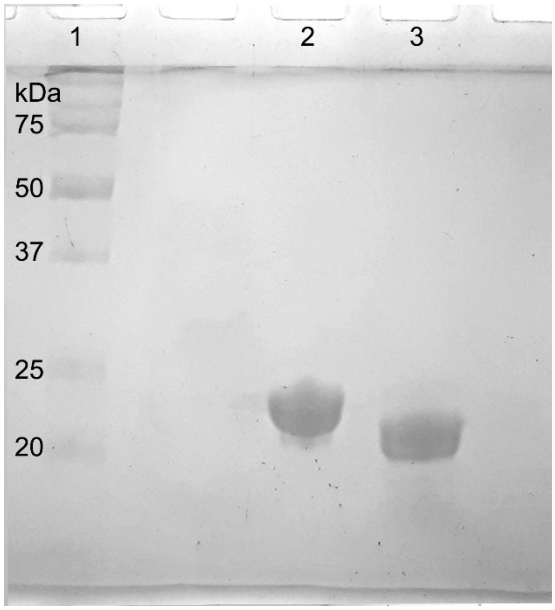

Figure 5 - 12% SDS-PAGE of TcmN cleavage.

Lane 1: molecular weight standard. Lane 2: purified, not cleaved TcmN. Lane 3: TcmN after cleavage.

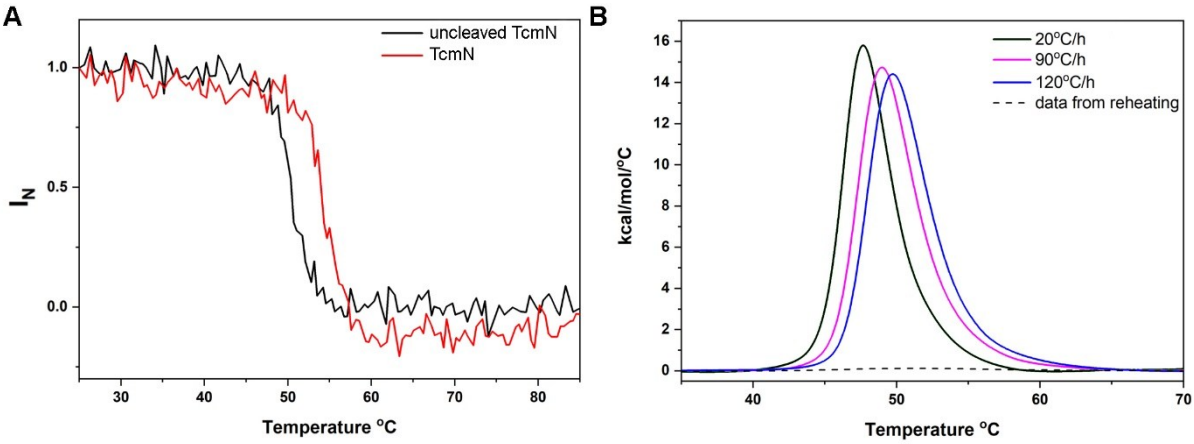

Figure 6 – Thermal stability of TcmN.

(A) Normalized intensity, I_N, as a function of the uncleaved and cleaved 100 µM TcmN temperature in 20 mM sodium phosphate pH 7, 150 mM NaCl estimated from the CD intensities at 208 nm. (B) Overlay of experimental thermograms of 211 µM TcmN in 20 mM sodium phosphate pH 7 and 150 mM NaCl at different temperature rates. The dashed black profile was acquired after heating the sample to 90 °C.

To ensure that the cleaved TcmN structure does not show considerable structural differences compared with the uncleaved one, ^{1}H-^{15}N HSQC NMR spectra of both proteins were acquired and compared. Figure 7 shows the ^{1}H-^{15}N HSQC NMR spectrum of the uncleaved ^{15}N-labeled TcmN (black) overlaid with the ^{15}N-labeled TcmN (red). The cleaved TcmN spectrum has a good chemical shift dispersion, showing that the protein is well-folded and similar to the uncleaved one. Different peaks of the uncleaved TcmN are related to residues of the cleaved region that comprise the His-Tag sequence and the thrombin cleavage site.

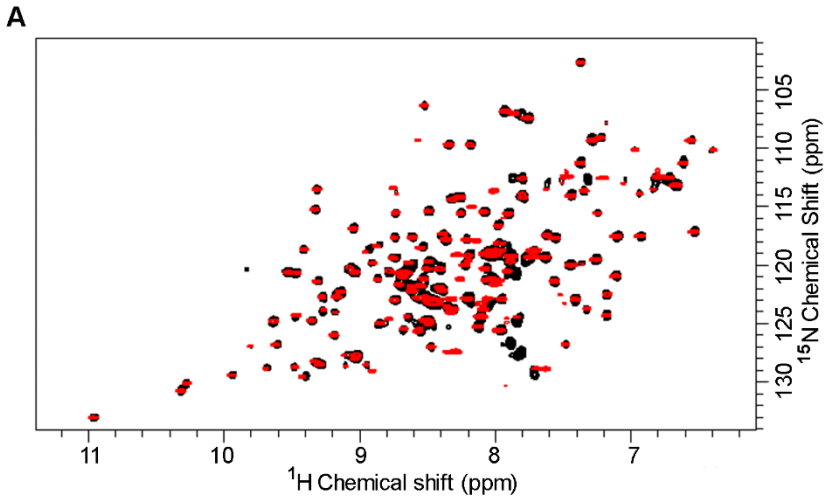

Figure 7 – TcmN NMR overlay cleavage.

(A) Overlay of ^{1}H-^{15}N HSQC spectra of uncleaved 170 µM TcmN in 20 mM sodium phosphate pH 7, 150 mM NaCl in black and cleaved 200 µM TcmN in 20 mM sodium phosphate pH 7, 150 mM NaCl in red.

^{15}N and ^{13}C enriched TcmN was expressed in M9 media and purified as previously described. NMR triple-resonance experiments were acquired and analyzed to assign the TcmN backbone resonances. Figure 8 shows the sequential assignment of residues 86-90 using CBCACONH and HNCACB spectra. The remainder of the backbone was assigned using the same procedure.

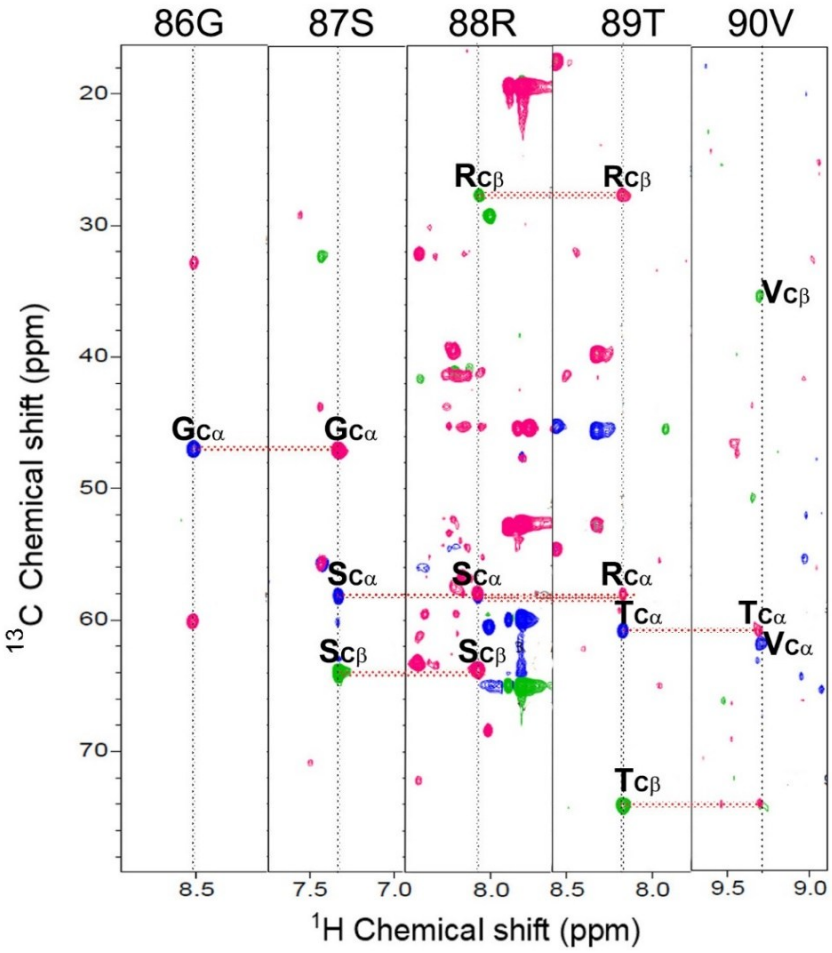

Figure 8 – Overlapping of the CBCACONH spectra (pink) with the HNCACB (blue and green).
At HNCACB, the blue (positive) peaks are Cα and green (negative) peaks are Cβ. The chemical shifts of ^{1}H and ^{13}C planes are shown. The residues and their numbers in the primary sequence are shown at the top of the boxes. The dotted lines link the HNCACB peaks of one system with its peaks on the CBCACONH.

Figure 9A shows the assigned ^{1}H-^{15}N-HSQC TcmN spectrum. The spectrum has a good chemical shift dispersion, showing the protein is well-folded. 238 peaks were expected in this spectrum, including the 66 HN side chain peaks. 89% of the backbone atoms have been assigned. The residues missing assignment are located on β-sheet 3 (residues R51, L52, T54), β-sheet 4 (residues V66, S67, H68, R69, V70), β -sheet 5, and near loop 7 (residues A80, H81, R82, V83, E84, T85), β -sheet 6 (residues A89, M91), and part of the α-helix 3 (residues K141, E145, D146) (Figure 7B). Among these residues, R69, S67, and R82 participate in the catalytic activity of TcmN and might be in conformational exchange since residues in intermediate conformational exchange (i.e., dynamics on the μs-ms timescale and constant rate in the order of the chemical shift difference ($\Delta\delta$) often shows NMR peaks with very low intensities (MORAES *et al.*, 2018; PALMER, 2004). Other residues surrounding the cavity and those near

the loops are also suggested to be of higher flexibility, which can influence the intensity of the resonance peaks. While most of these residues are likely present in the ^{1}H-^{15}N-HSQC spectrum, they could have relatively low intensity in the HNCA and other triple-resonance spectra used to assign the backbone.

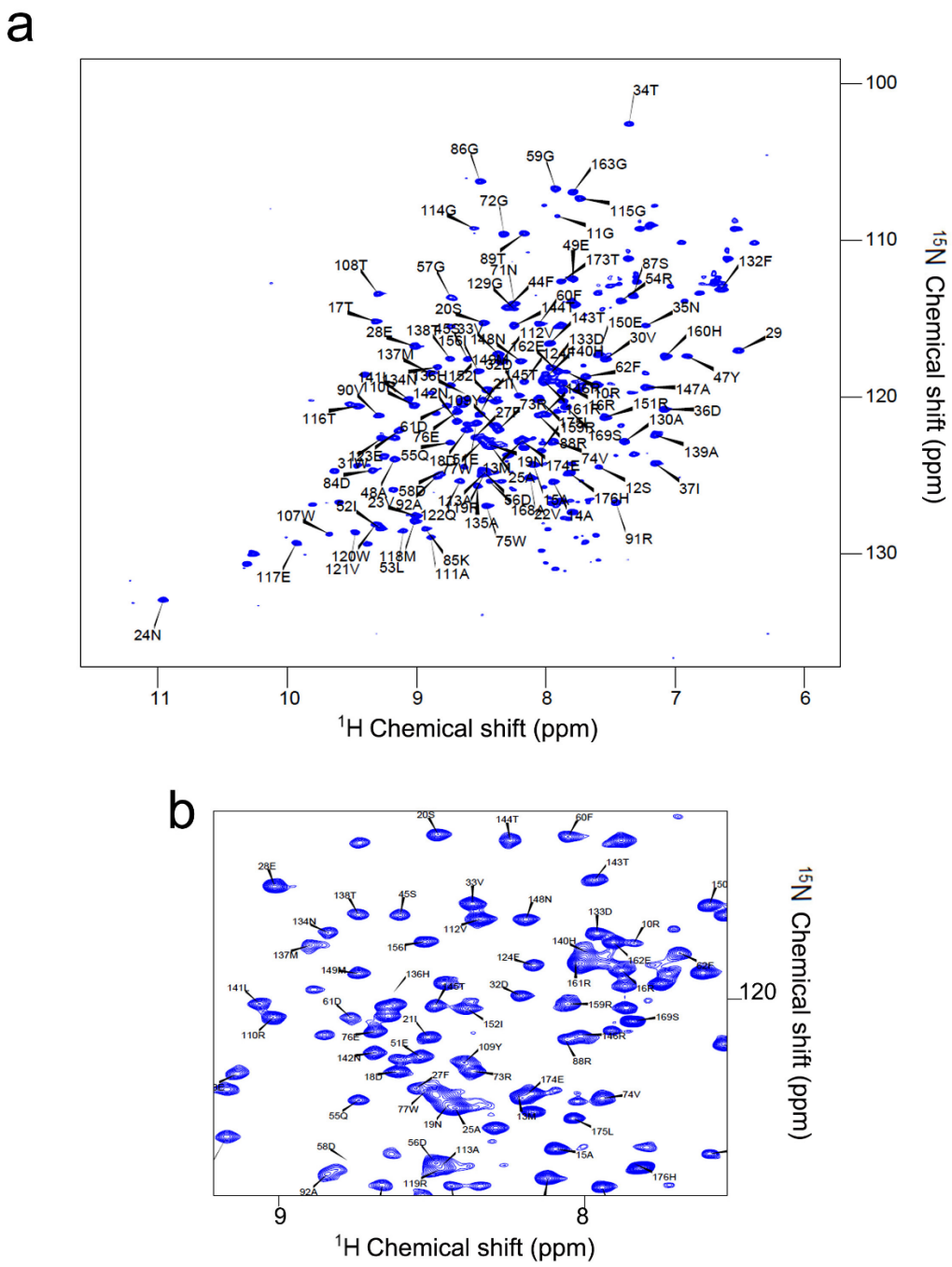

Figure 9 – TcmN NMR assignment.

(a) ^{1}H-^{15}N HSQC spectrum showing the assignments and (b) zoom region. In this experiment, each peak contains information on the frequency of ^{1}HN and ^{15}N of a specific residue, except for prolines.

4.2 Conformational dynamics of free TcmN

Aiming to build an atomistic picture of TcmN conformational equilibrium, 6 × 1 μs MD simulations of free TcmN were performed and analyzed. Principal Component Analysis (PCA) was employed to examine the main contributions to TcmN conformational variance. The first two components, PC1 and PC2, represented 28% and 18% of the Cα conformational variance, respectively (Figure 10A). These components correspond primarily to movements of the L3, L5, L7, and L9 segments located at the neck of the central cavity. The PC1-PC2 probability-based free-energy landscape shows that the two most populated regions correspond to closed and open conformations, respectively, and residues comprising the L3, L5, L7, and L9 loops located at the entrance of the central cavity were major contributors to the conformational variance of the PC1 and PC2 modes (Figure 10B). PC1 correlated strongly with L5-L9 distance (r2 = 0.82) (Figure 10C). Time traces of PC1 and PC2 evolution illustrate that the two minima were sampled by multiple replicas (Figure 11). Movements of L5 and L9 could be necessary to allow the internalization of a linear polyketide chain into the cavity. The PC1 projection of each frame and the distance between loops L5 and L9 (Figure 12A) showed bimodal distributions and were analyzed with two Gaussian curves to find the center of each peak. Two populations were observed with L5-L9 distances of 22.8Å and 12.6 Å, corresponding to open and closed states, respectively. In the TcmN crystallographic structure (PDB ID: 2RER) (AMES, BRIAN DOUGLAS *et al.*, 2008), the distance between L5-L9 is 9.1 Å, corresponding to a closed conformation. Changes in cavity hydration were assessed by measuring the water content inside the cavity over the trajectories. A strong correlation (r^2=0.61) was obtained between the L5-L9 distance and the number of water molecules inside the cavity (Figure 12B). The histograms revealed two clusters: one related to the closed state with an L5-L9 distance of 12.6 Å and ~28 water molecules, and another corresponding to the open state, with an L5-L9 distance of 22.8 Å and ~36 water molecules. The higher number of water molecules inside the cavity in the open conformation may impose an entropic penalty on this state by decreasing the solvent entropy, as the cavity is lined with hydrophobic residues, which could contribute to the higher free energy of the open state.

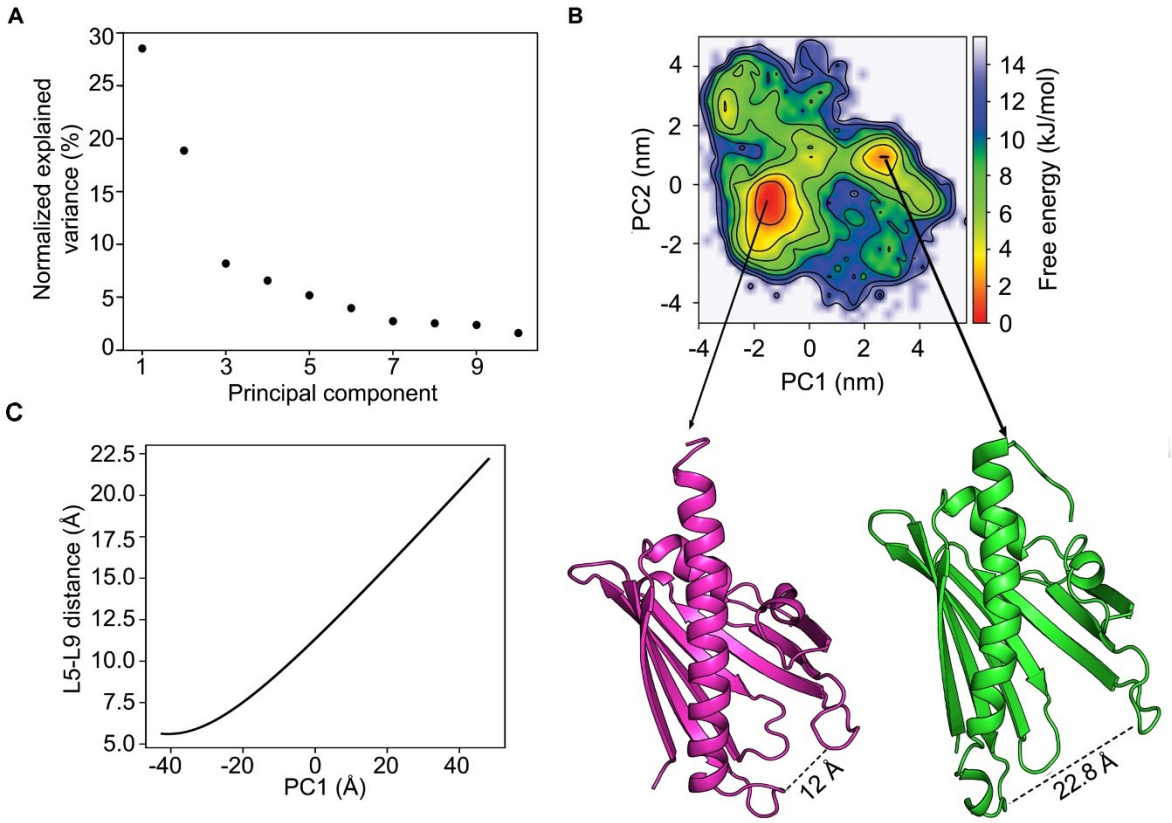

Figure 10 - Principal Component Analysis of TcmN MD simulation.

(A) Variance is explained by the first ten principal components. (B) PCA correlation plot of the projections of Cα variance on the two principal eigenvectors, PC1, and PC2, and representative "closed" (pink) and "open" (green) structures with the R61 (L5) - P119 (L9) distance indicated. The Gibbs free energy was estimated using the Boltzmann equation and represented as a landscape built up using a color gradient. (C) L5-L9 distance along PC1.

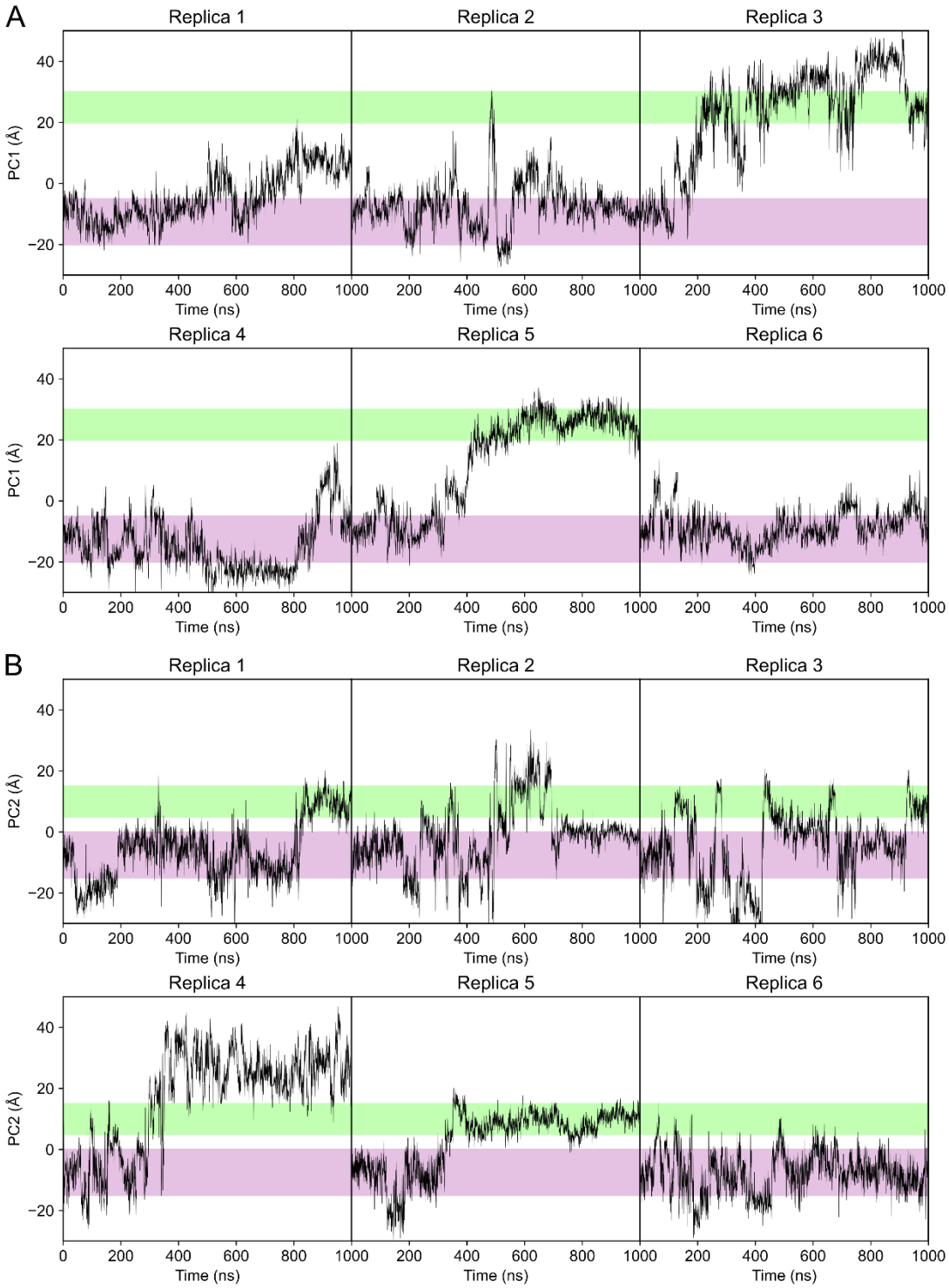

Figure 11 - PC1 and PC2 evolution throughout the TcmN MD trajectories.

(A) and (B) show the projections of the trajectories onto PC1 and PC2, respectively. The purple and green regions correspond to the closed and open minima in Figure 10.

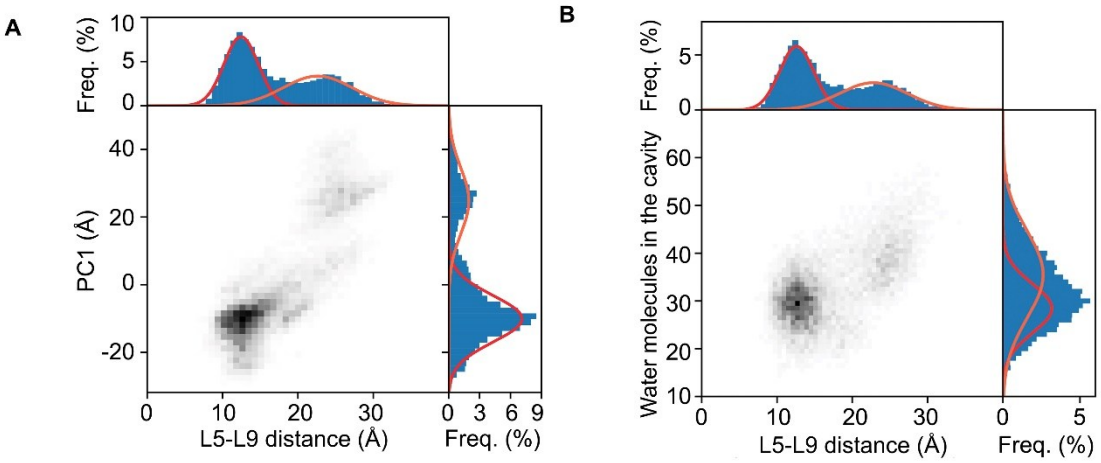

Figure 12 – Principal Component Analysis of free TcmN MD simulations.

(A) 1D (blue) and bidimensional (greyscale) histograms of L5-L9 distance vs. a projection of the structures along PC1. (B) 1D (blue) and bidimensional (greyscale) histograms of the distance between loops L5 and L9 and the number of water molecules in the TcmN cavity. The 1D histograms show two fitted Gaussians, red and orange.

A hydration tuning mechanism might have evolved to control the opening and closing of Bet v 1-like proteins. Like TcmN, the higher-energy, open state of Fag s 1 is entropically unfavorable due to the hydrophobic cavity's exposure to solvent (MORAES *et al.*, 2016, 2018). It is likely that this entropic penalty pressure TcmN and Fag s 1 to return to lower energy, closed states. Similar results were observed by Oroguchi and Nakasako (OROGUCHI; NAKASAKO, 2016a), who investigated the modulation of the relative dynamics between two domains of glutamate dehydrogenase by changes in local hydration in the hydrophobic cleft between the two domains. They formulated a model where the wetting/drying of a hydrophobic pocket worked as a switch for interdomain motions (OROGUCHI; NAKASAKO, 2016b). Functional protein movements might be coupled with changes in hydration structure (MATSUOKA *et al.*, 2015; OROGUCHI; NAKASAKO, 2016b). Thus, modulating the conformational equilibrium of Bet v 1-like proteins may be possible by changing the polar and apolar residue distributions within the main cavity.

4.3 The interplay of conformational equilibrium, function, and aggregation

Proteins are dynamic molecules and may sample conformational states at risk of misfolding and aggregation. The trade-off between function vs. aggregation leads to an evolutionary pressure that creates competition between increased binding efficiency and reduced protein stability, balancing the ability to populate multiple states without creating an unacceptable propensity for aggregation (DEPRISTO; WEINREICH; HARTL, 2005; FERROLINO *et al.*, 2013; GERSHENSON *et al.*, 2014). Inspired by the work of Ferrolino et

al. (FERROLINO *et al.*, 2013) with the β-rich cellular retinoic acid-binding protein 1, a scheme for describing the fine-tuned interplay between conformational flexibility, function, and aggregation was proposed (Figure 13). Using PC1 as a reaction coordinate and plotting it as a function of Gibbs free energy, we built an energy landscape profile of TcmN (Figure 13A). In the model, TcmN is in conformational equilibrium between closed and open conformational ensembles, with a preference for the former. The open state is needed for substrate binding but, at the same time, is more prone to amorphous aggregate formation (Figure 13B). In the case of TcmN, it seems plausible that evolution solved the function vs. aggregation problem by controlling the open-state population and kinetically modulating its lifetime.

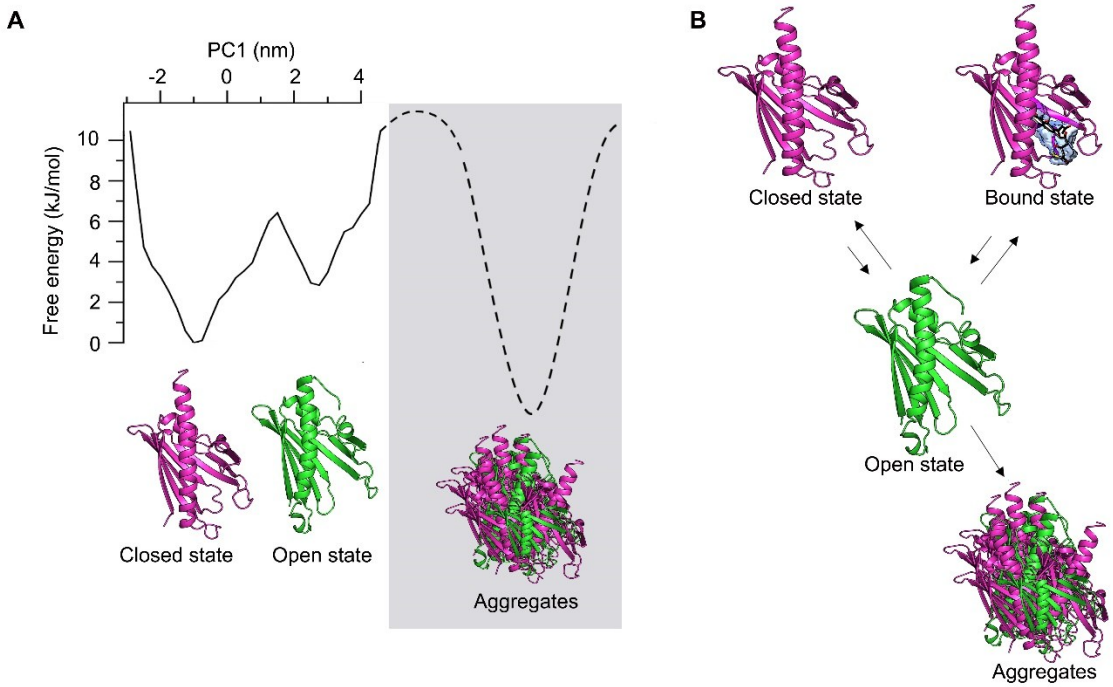

Figure 13 – The proposed interplay between conformational flexibility and aggregation of TcmN.
(A) The energy landscape of TcmN under native conditions comprises a delicate balance between protein conformational flexibility and aggregation. (B) Schematic representation of TcmN's conformational transitions shows that the opened state (aggregation-prone conformation) is required for ligand binding and can lead to aggregation.

4.4 TcmN interactions

TcmN promotes C9-C14 first ring, C7-C16 second-ring cyclization, and aromatization of unreduced linear polyketides, and the poly-β-keto intermediate is a very reactive compound. Because of that, it is impossible to experimentally study the interaction of TcmN with substrate

compounds. To overcome this problem, we investigated the interaction of TcmN with three aromatic compounds, which possess two rings, hydroxyl and carbonyl groups. Those compounds should resemble the structure of the cyclic polyketide intermediate in the TcmN-catalyzed aromatization and cyclization reactions (Figure 4). Three molecules with the described characteristics were selected to study the interaction with TcmN (Figure 14), Naringenin, Taxifolin, and 8-Anilinonaphthalene-1-sulfonic acid (ANS).

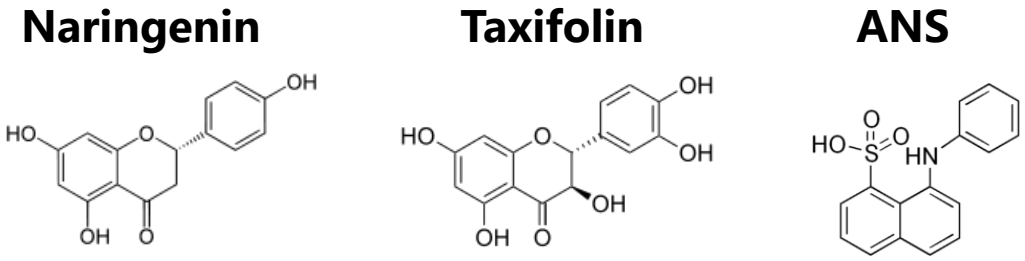

Figure 14 – Selected TcmN's reaction intermediate and product analogs.
Chemical structure of Naringenin, Taxifolin, and ANS.

NMR spectroscopy evolved as a powerful tool to investigate protein-ligand binding. The STD experiments showed that TcmN recognizes Naringenin and ANS but not taxifolin (data not shown). Due to the greater similarity to the TcmN product, having the first aromatic ring with two hydroxyl groups, Naringenin was chosen to describe the subsequent experiments. The ^{1}H NMR spectra of 25 mM of racemic Naringenin in DMSO-D6 were analyzed and assigned. An STD experiment involves subtracting a spectrum in which the protein has been selectively saturated by irradiation from a spectrum without protein saturation. The protein saturation is rapidly propagated through a spin diffusion effect, and when a low molecular weight ligand binds to the protein, the saturation is also transferred to the bound ligand by a cross-relaxation process at the ligand-protein interface (ANGULO; NIETO, 2011). In the difference spectrum, only the ligand signals that received the protein saturation transfer will remain (Figure 15B). Stronger STD signals were observed for the aromatic protons H_1 and H_5 (Figure 15C-D), showing that the structure of two fused rings of Naringenin interact more with the protein, suggesting that they are positioned further inside the cavity.

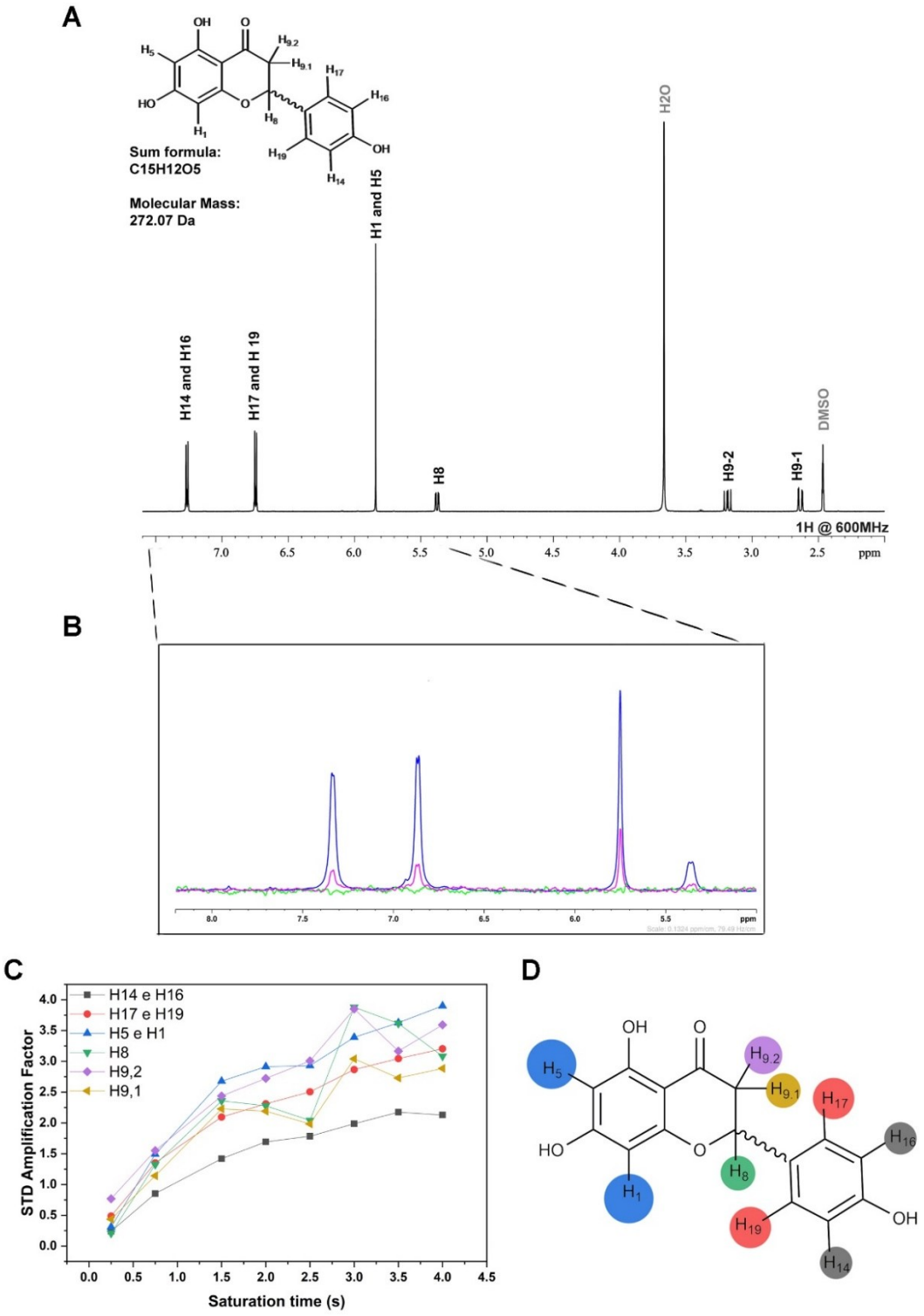

Figure 15 – STD NMR.

(A) ^{1}H NMR assigned spectra of 25mM of racemic Naringenin. (B) Spectra overlay of the aromatic spectral region of 1D STD NMR spectra recorded at a single saturation time (tsat=2 s, 25°C). The off-resonance STD spectrum of 1 mM Naringenin binding to 20 µM TcmN in blue, the difference resonance STD spectrum of 1 mM Naringenin binding to 20 µM TcmN in pink, and the difference resonance STD spectrum of 1 mM Naringenin in green; (C) Built-up curves of STD amplification factor as a function of the saturation time at 298 K and 600 MHz (^{1}H nucleus) for binding of Naringenin to 20 µM TcmN in 20 mM sodium phosphate pH 7, 100 mM NaCl. (D) Chemical structure of Naringenin with assigned protons labeled.

Next, the binding affinity by ITC and bound TcmN stability by DSC measurements were evaluated. Utilizing a two-state model for the fitting, a binding affinity of 135 µmol/L was observed on the ITC experiments of Naringenin titration onto 140 µM of TcmN (Figure 16). DSC thermograms were used to evaluate bound TcmN stability. The sample of Naringenin was solubilized with DMSO, and a control experiment with TcmN and DMSO was also acquired. The concentration of 1 mM and 2 mM of Naringenin had an equivalent concentration of 4% and 8% DMSO in the final sample. Tm* values of TcmN with Naringenin were 2°C higher than the control, suggesting that the interaction with the ligand increases TcmN stability (Figure 17). Protein concentration also affects TcmN thermal stability; it decays as protein concentration increases (VALADARES et al., 2021), and as the concentration of this experiment is much higher than the CD experiment done with free TcmN (Figure 6A), both cannot be compared.

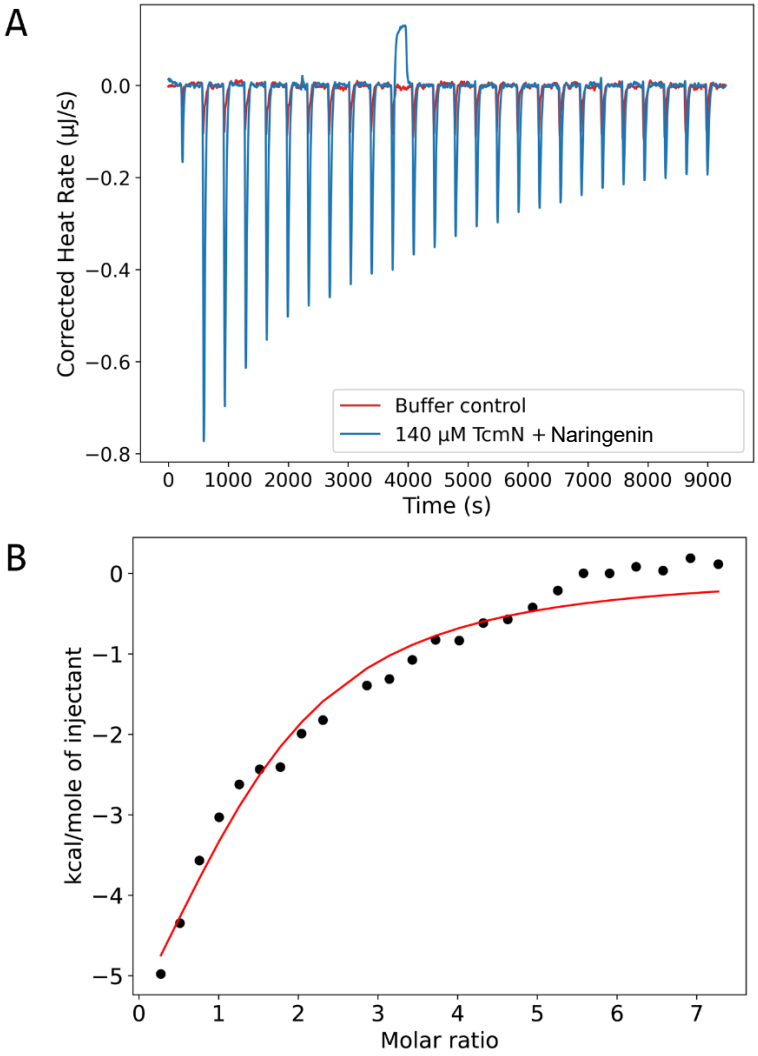

Figure 16 – ITC experiment TcmN's reaction intermediate and product analogs.

Calorimetric titration of Naringenin into a sample cell containing TcmN. The titration was performed at 25 °C in 20 mM sodium phosphate pH 7, 100 mM NaCl, and 4% DMSO

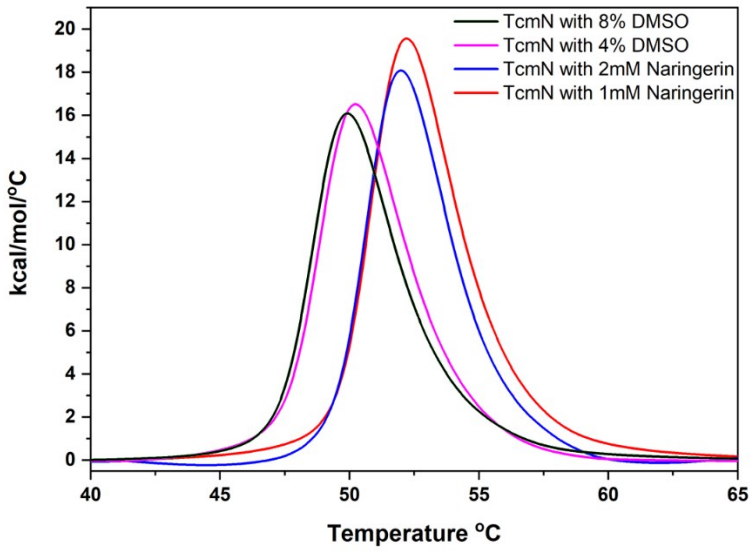

Figure 17 - Effect of Naringenin and DMSO on TcmN thermostability.

Overlayed experimental thermograms of 211 μM TcmN in 20 mM sodium phosphate pH 7, 100 mM NaCl in the presence of 1 mM and 2 mM of Naringenin, and with 4 and 8% of DMSO.

Having mapped the binding affinity, thermostability, and interaction position of the ligand, the next step was identifying how the protein behaves structurally upon binding with Naringenin. The ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled TcmN titrated with Naringenin were acquired, and structural changes were identified by mapping the chemical-shift perturbations (CSP). Figure 18 shows the overlay of free TcmN with TcmN in a complex with different concentrations of Naringenin, and spectral changes can be observed. Significant chemical shift perturbation can be observed in residues T5, V62, W63, T96, Q110, F120, T126, A127, T131, and N136. These residues surround the TcmN cavity and the alpha-helix 3. Residues F120 and W63 near the cavity entrance in loops 5 and 9 showed high CSP. These loops, identified as regulators of cavity opening and closing movements necessary for substrate binding and product release, are crucial for TcmN function (VALADARES et al., 2021). One can imply that residue W63 could be identified as a potential gatekeeper, akin to its homolog residue in WhiE-ORFVI, suggesting its involvement in controlling ligand access to the binding site (LEE; AMES; TSAI, 2012). Interestingly, residues Y35, L129, and T132, which form part of the bottleneck region, have the NMR signal broadened upon Narigenin titration, implying variation in cavity volume and shape, a behavior documented in structurally similar proteins Bet v1 and Fag s1 (ASAM et al., 2014; MORAES et al., 2018). Additionally, curved signal migration

patterns in the titration experiment indicate a complex interaction model between TcmN and Naringenin, suggesting the possible binding of multiple ligands to the enzyme or from conformational changes in the protein upon ligand binding (Figure 18 and Figure 23).

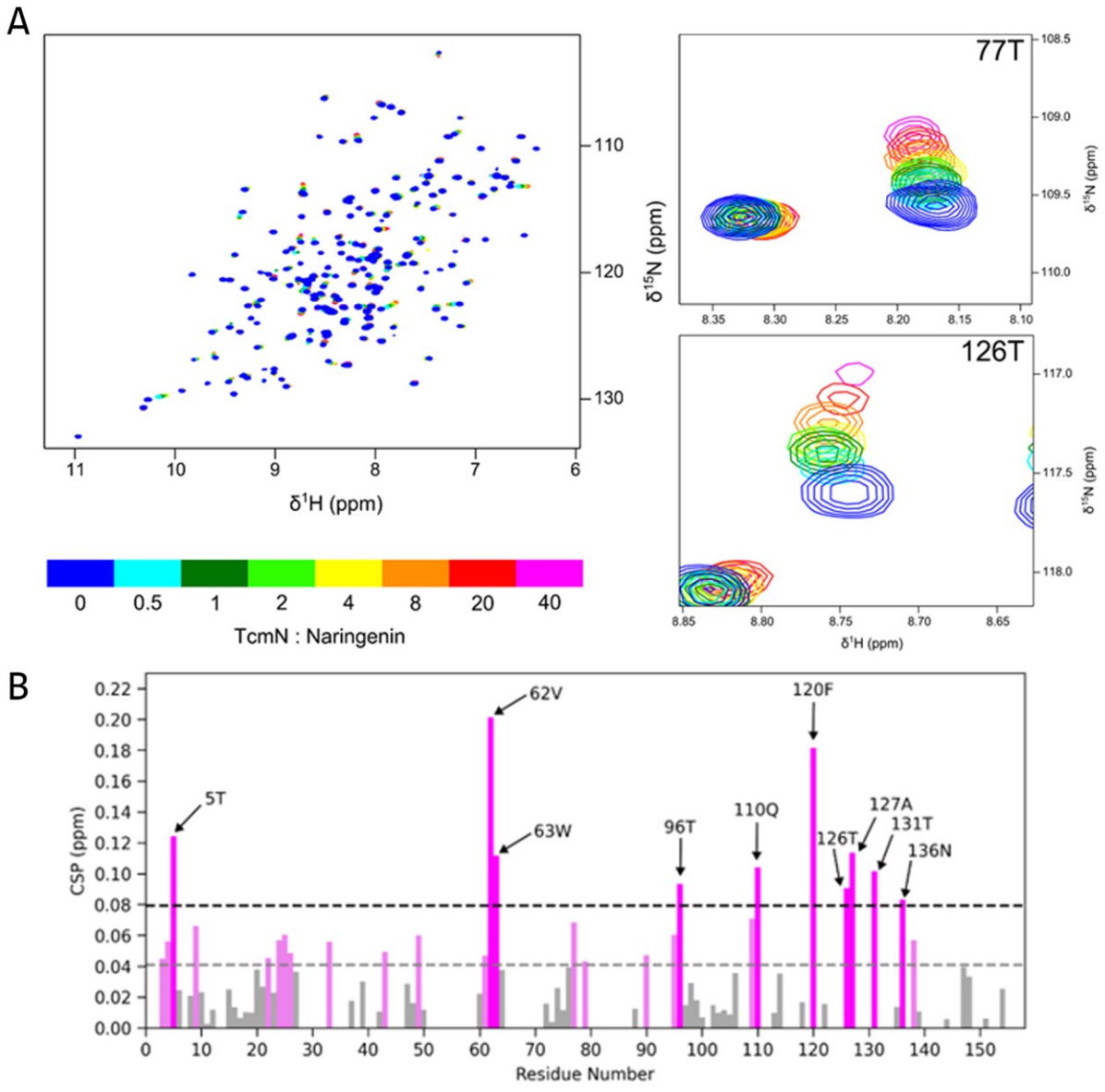

Figure 18 – CSP effect on TcmN upon Naringenin titration.

A) ^{1}H-^{15}N HSQC spectra overlay of TcmN titrated with Naringenin and selected regions of ^{1}H-^{15}N HSQC showing examples of chemical shift variation of residues T77 and T126 on the side. The color scale indicates the concentration ratio of TcmN to Naringenin. B) CSP as a function of TcmN residue number.

Bound TcmN dynamics were investigated using NMR relaxation measurements. Results obtained from the NMR ^{1}H relaxation rates R_1 and R_2 data can be used to calculate the rotational correlation time (τ_c). The τ_c value is crucial as it reflects the size and oligomeric state of the protein. τ_c value of the free protein is higher than the bound protein, suggesting that the

protein in its bound form has a more compact conformation. However, this data should be considered with caution since the τ_c values of free and bound TcmN fall within the measurements' confidence interval. Relaxation rates and τ_c values for free and bound TcmN are shown in the table below:

TcmN state	R_1 (s^{-1})	R_2 (s^{-1})	$\boldsymbol{\tau}_c$ (ns)
Free	1.01 ± 0.03	7.82 ± 0.89	8.21 ± 0.69
Bound	0.96 ± 0.08	6.79 ± 0.62	7.80 ± 0.81

Carr-Purcell- Meiboom-Gill (CPMG) relaxation dispersion NMR spectroscopy (CARR; PURCELL, 1954; MEIBOOM; GILL, 1958) was measured to characterize residues of TcmN that might experience conformational exchange. $R_{2,effs}$ (s^{-1}) values obtained from ^{15}N CPMG experiments at 298 K of the free TcmN (Figure 19A) showed no significant difference between the $R_{2,eff}$ measured with CPMG frequency of 50 and 1000 Hz. $R_{2,effs}$ (s^{-1}) values obtained from the bound TcmN showed an extra contribution in the experiment at 50 Hz that was attenuated at 1000 Hz, decreasing it to a value similar to the experiment carried out with the free TcmN (Figure 19B). ΔR_2 of bound TcmN was calculated and compared to CSP. The CSP reflects changes in residues that directly participate in the protein-ligand; ΔR_2 possibly demonstrates the balance between two protein conformations, the majority conformation of the bound protein and a minority conformation of the free protein.

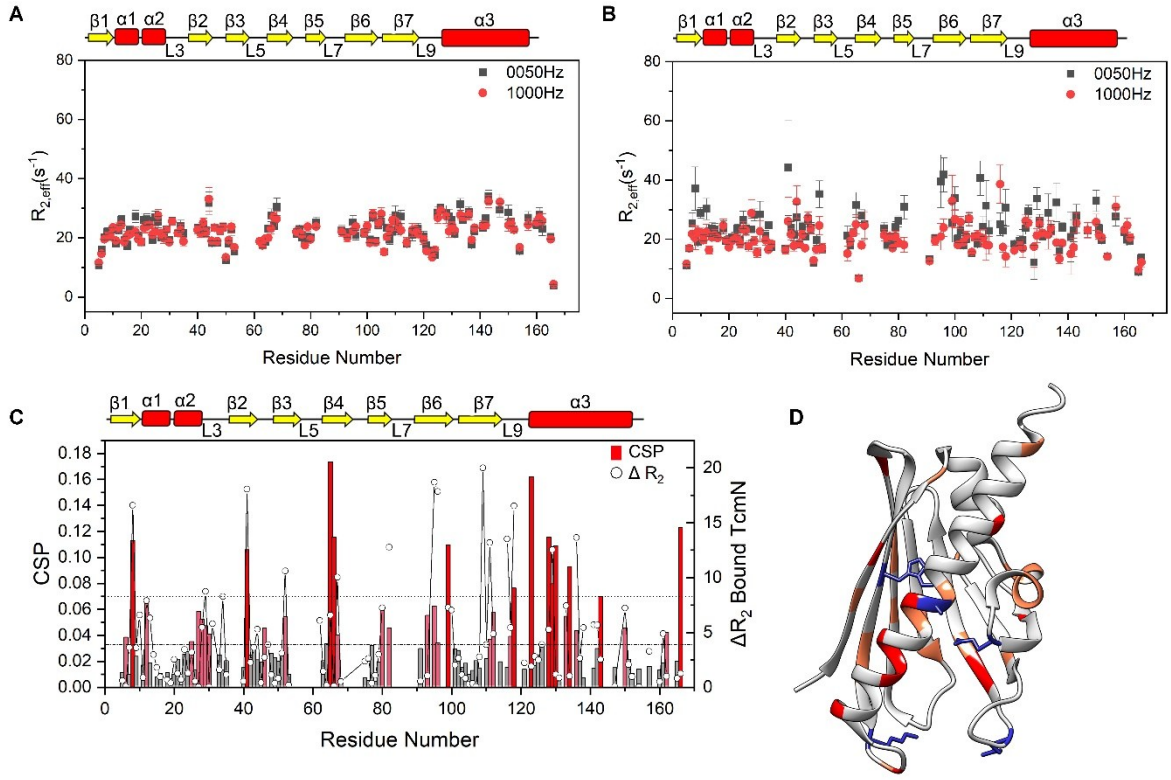

Figure 19 – Backbone dynamics of TcmN by NMR.

(A) $R_{2,eff}$ (s^{-1}) rate as a function of TcmN residue number of 130 µM TcmN in 20 mM sodium phosphate pH 7, 100 mM NaCl. (B) $R_{2,eff}$ (s^{-1}) rate as a function of TcmN residue number of 126 µM of TcmN titrated with 1.2 mM of Naringenin. Black squared, and red circles refer to ^{15}N-RC-R_2 CPMG relaxation dispersion experiments acquired with CPMG frequencies of 50 Hz and 1000 Hz, respectively. (C) Chemical shift perturbation index, CSP, and $\Delta R_{2,eff}$ (s^{-1}) of bound TcmN as a function of TcmN residue number. CSP was calculated from chemical shift perturbations obtained in NMR spectra shown in Figure 10A, and $\Delta R_{2,eff}$ (s^{-1}) was calculated from data shown in (B). (D) TcmN structure (PDB ID: 2RER) with side chains of residues in which CSP values were higher than the average value plus the standard deviation of all sets is colored in red, and residues that had a high ΔR_{ex} (s^{-1}) and low CSP are shown in blue.

Further experiments assessing the interaction affinity between TcmN and naringenin were performed. NMR lineshape analysis was done using the Titan software. $^{1}H^{15}N$ HSQC experiments acquired at 600 MHz with naringenin titration up to 561 µM, and $^{1}H^{15}N$ HSQC experiments acquired at 800 MHz with naringenin titration up to 2 mM were utilized. The residues selected for the signal shape analysis in both $^{1}H^{15}N$ HSQC experiments were those with CSP above the mean plus standard deviation: residues T5, V62, T96, Q110, F120 and A127 in the 600 MHz experiments, and residues T5, V62, W63, T96, F120, T126, A127, T131 and N136 in the 800 MHz experiments. Data analysis of the spectra acquired at 600 MHz

revealed that the simulation using the lock-and-key model best represented the system and matched experimental data, giving a K_D value of 111± 5 µM.

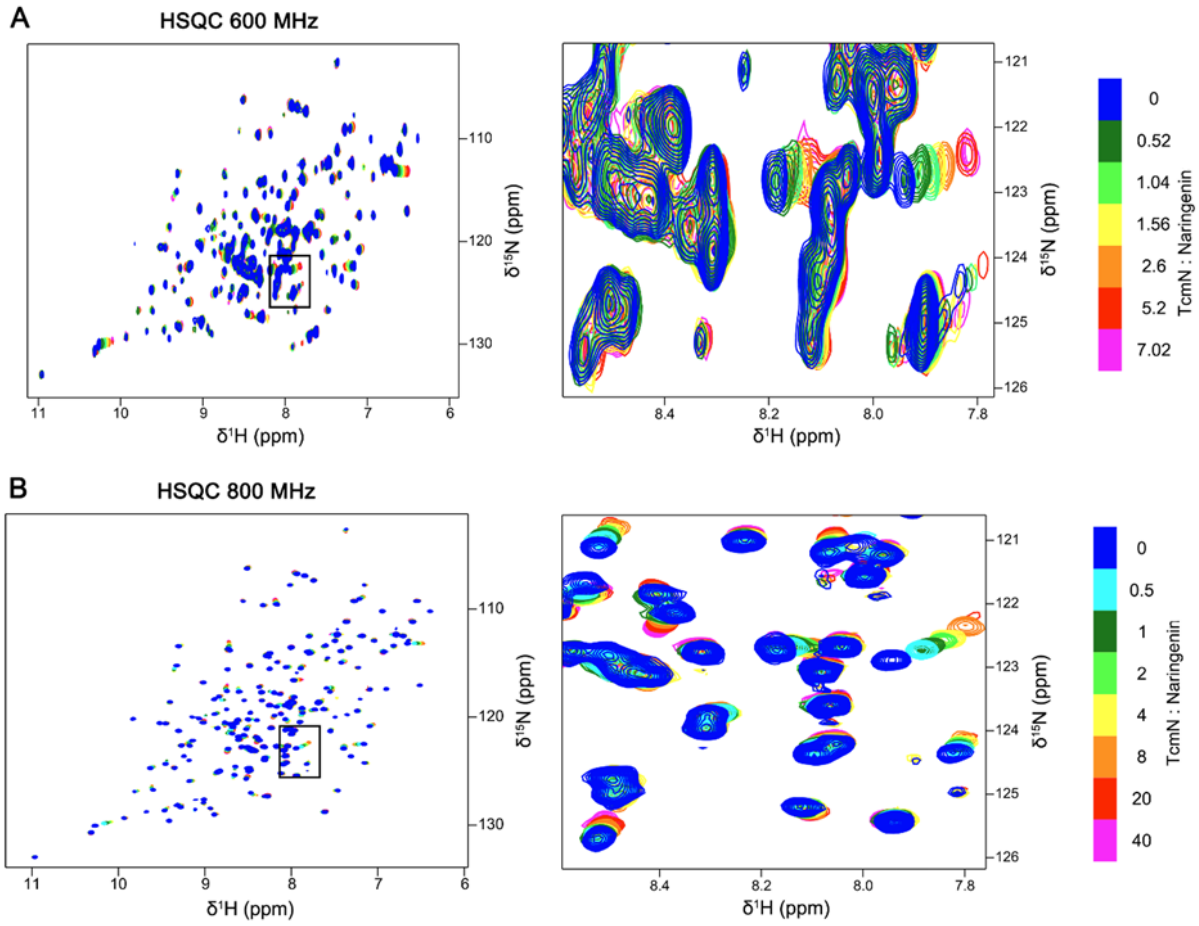

Figure 20 – HSQC overlay of TcmN and Naringenin titration

A) ^{1}H-^{15}N HSQC spectra overlay of TcmN titrated with Naringenin experiment acquired at 600 MHz. The color scale indicates the ratio of TcmN to Naringenin. B) ^{1}H-^{15}N HSQC spectra overlay of TcmN titrated with Naringenin experiment acquired at 800 MHz.

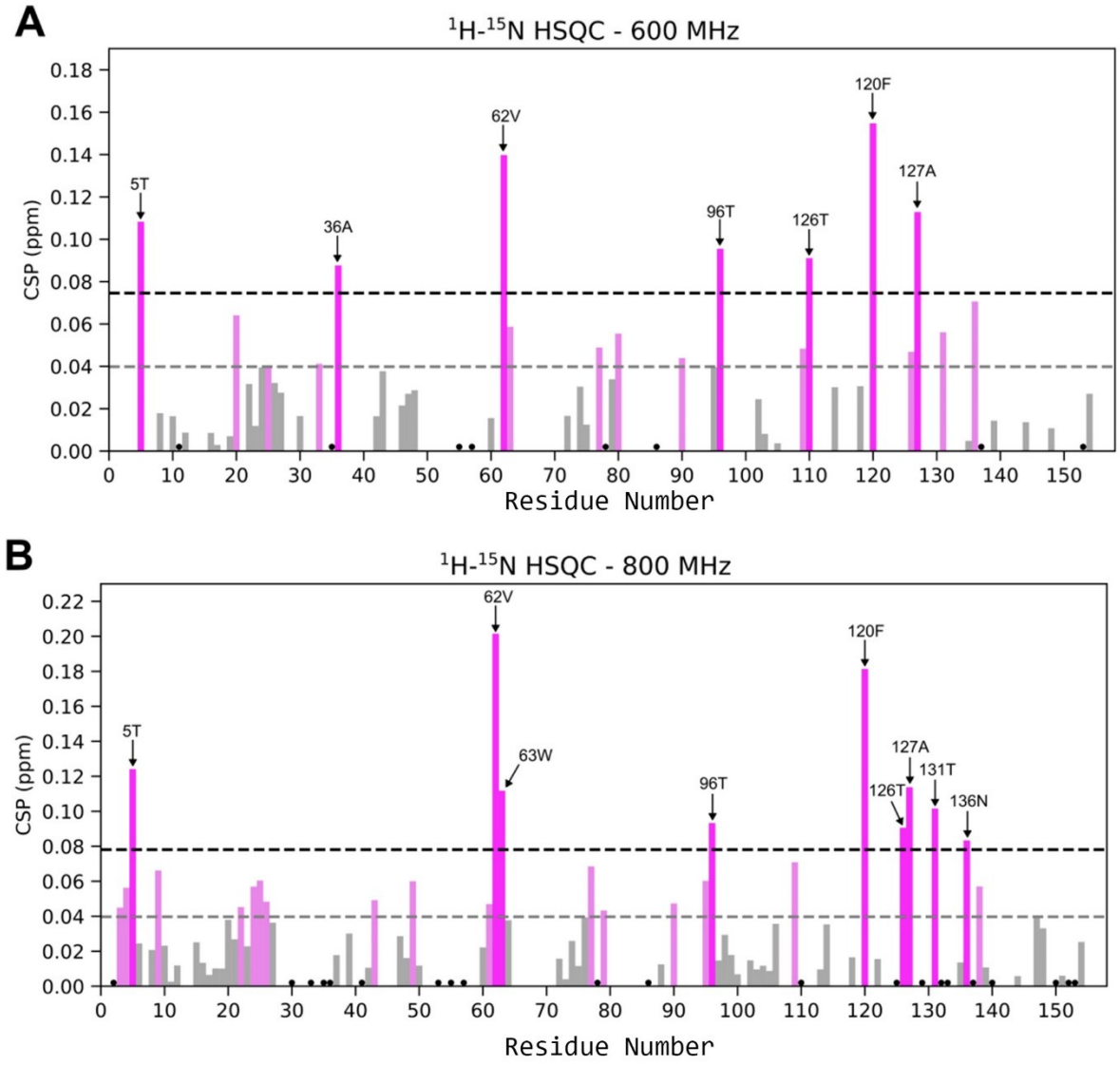

Figure 21 – CSP as a function of TcmN residue number.

A) CSP as a function of TcmN residue number of the titration with Naringenin of the experiments acquired at 600 MHz. B) CSP as a function of TcmN residue number of the titration with Naringenin of the experiments acquired at 800 MHz.

The experiments acquired at 800 MHz provided better resolution and were done with more titration points. In the Titan analysis for these experiments, it was observed that only the model of sequential interactions could reproduce the curved migration behavior and intensity loss seen on the spectra. Thus, this model was considered the most suitable for representing the system. It can be inferred that the change in slope observed during the migration of some signals is due to three distinct states: free TcmN, TcmN interacting with one naringenin, and TcmN interacting with two naringenins. The two sites may have different affinities for the ligand, and

the second interaction site only begins to be populated after the first is saturated. The presence of curved migration in specific residues becomes apparent in the spectrum at a concentration of approximately 400 μM of naringenin, which could indicate the onset of occupation of the second interaction site. Subsequent data adjustment and spectrum simulation were carried out for the lock-and-key model up to a titration of 200 μM of naringenin. The similarity between the spectrum simulation and experimental data suggests that, until this point of the titration, the interaction of TcmN with the first naringenin predominates significantly. Lee et al. analyzed the co-crystal structure of TcmN and Taxifolin. In this structure, taxifolin is situated proximal to the cavity entrance, accompanied by conformational changes in residues R82, H128, and F120, as well as loops 5 and 7. Furthermore, aromatic interactions occur with W63, W65, F88, and F120, while hydrophobic interactions involve the side chains of M125 and P87, along with hydrogen bonding interactions with R82 and D57(LEE; AMES; TSAI, 2012). It can be assumed that the second naringenin interacting with TcmN in the sequential interaction model may be in a position similar to that of taxifolin since both molecules have a similar structure with the same patterns of hydroxyl substitution of the first ring. The higher CSP than the mean plus standard deviation observed for residues F120 and W63, both in the entrance loops of the cavity, may be indicative of this interaction, as well as the broadment of the signals from D57 and M125. In the titration experiment acquired at 600 MHz, possibly due to the maximum naringenin concentration of 561 μM, the first site may not have been completely saturated. At this concentration, the second site may not be yet populated. In this study, the kinetic and thermodynamic constants obtained using the TITAN analysis cannot be quantitatively analyzed due to the limited number of titration points.

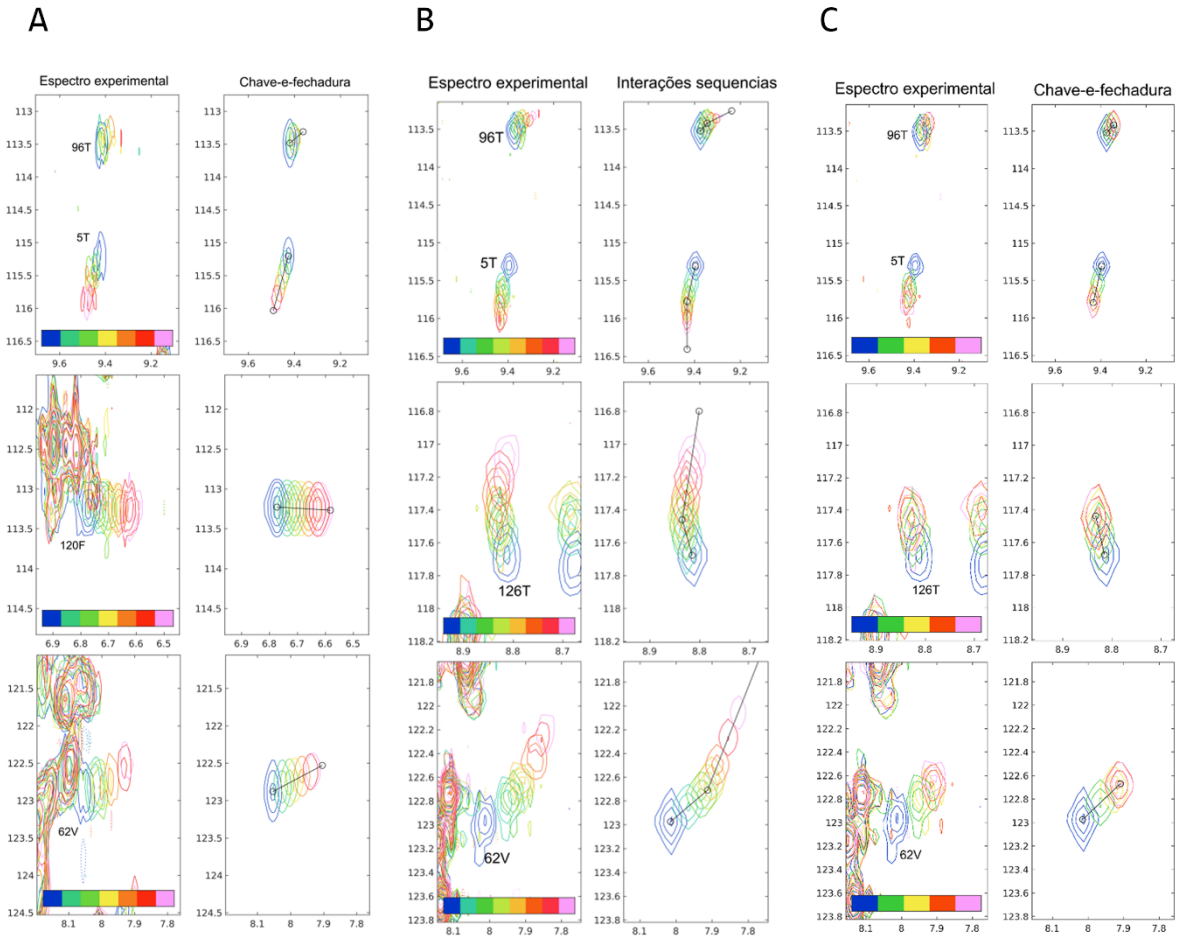

Figure 22 – Comparison of the spectra from the ^{1}H-^{15}N HSQC experiments and the simulations performed by TITAN.

A) Comparison of the spectra from the ^{1}H-^{15}N HSQC experiments acquired at 600 MHz and the simulations performed by TITAN for the lock-and-key model. The signal assignments are indicated in the figure. B) Comparison of the experiments acquired at 800 MHz and the simulations performed by TITAN for the sequential interactions model. C) Comparison of the experiments acquired at 800 MHz up to the titration of 200 µM of Naringenin and the simulation performed by TITAN for the lock-and-key model.

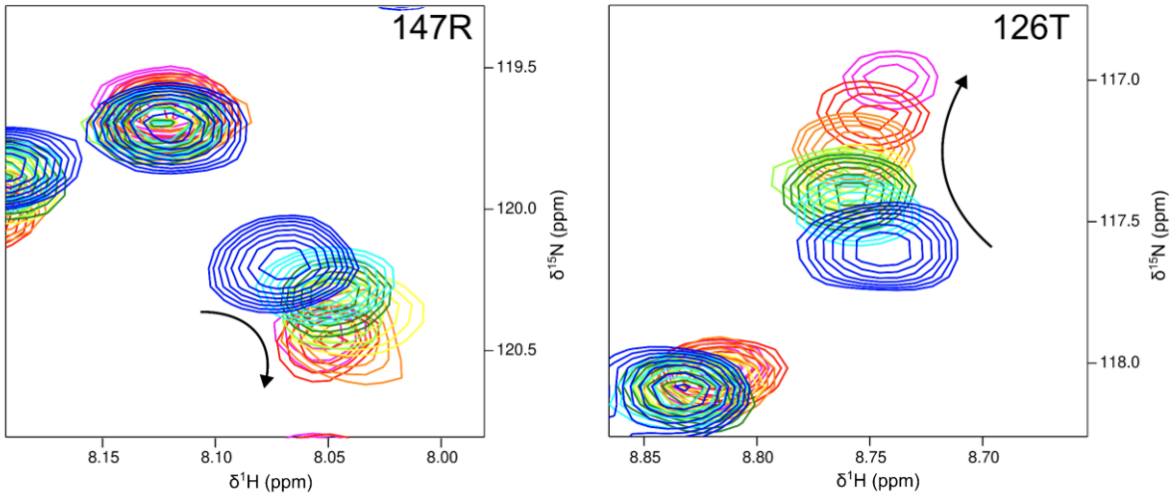

Figure 23 – Examples of signals exhibiting curved behavior during the titration of naringenin in TcmN. ^{1}H-^{15}N HSQC spectra overlay of selected residues of TcmN titrated with Naringenin on experiments acquired at 800 MHz.

4.5 Conformational dynamics of bound TcmN

Ames *et al.* studied the possible role of TcmN pocket residues by docking simulations of all possible putative substrates or products; it was seen that the substrate makes contact with residues that enable catalyzing the formation of the first and second rings of a linear polyketide substrate. That pocket can accommodate up to three linearly fused rings, but not the fourth. It was also suggested that third-ring cyclization is likely a spontaneous process (AMES, BRIAN DOUGLAS *et al.*, 2008). Aiming to investigate the mechanism of TcmN molecular recognition of substrates, intermediates, and products and to identify and characterize the different conformational states of TcmN, molecular docking, and MD simulations were made. From the possible intermediates in first and second-ring cyclization made by TcmN identified by Ames *et al.*, the intermediate 12 (INT12) with two aromatic linear rings was selected for our studies. The TcmN product analog Naringenin was also chosen for computational analyses. Since NMR interaction experiments were performed with a racemic sample of Naringenin, docking, and MD simulations were done with the crystal structure of TcmN (PDB ID: 2RER) and R and S isomers Naringenin. Docking simulations were performed using HADDOCK v 2.4. The center of mass of the ligand in each resultant pose was calculated to check the quality of docking results and the position of the ligand. Most of the NAR-R and all NAR-S poses were inside the TcmN's cavity. However, none of the INT12 poses were docked inside the TcmN's cavity (Figure 24A). For docking made with NAR-R and NAR-S, the poses with the lowest score (best

interaction predictions) were used as a reference to calculate the ligand RMSD from all other poses. Then, all poses were aligned using GROMACS, and clusterization was done. Clusters and poses from NAR-S had lower energy scores than NAR-R poses. The centroid pose from each cluster was visualized, and the best poses were chosen according to STD-NMR results and docking scores. From NAR-S docking, a pose named 461w with a score of -15.87 was selected, and from NAR-R, two poses, 71w and 568w, with respective scores of 23.94 and -8.16 were chosen (Figure 24B). Since no pose of INT12 was located inside TcmN's catalytic cavity, the software Autodock Vina was used for docking this molecule. The generated poses from Autodock Vina contained the ligand buried in the catalytic cavity, and the highest-scoring pose was chosen. INT12 docking probably did not work with HADOOCK because it is a large and flexible molecule, and the HADDOCK algorithm might not be efficient for this case. Docking simulation poses used to seed the molecular dynamic simulations of NAR-S e INT12 showed that NAR-S interacts with residues T132, T133, and R82. INT12 was predicted to interact with residues T132, T133, R82, F88, T35, S67, W65, and M91.

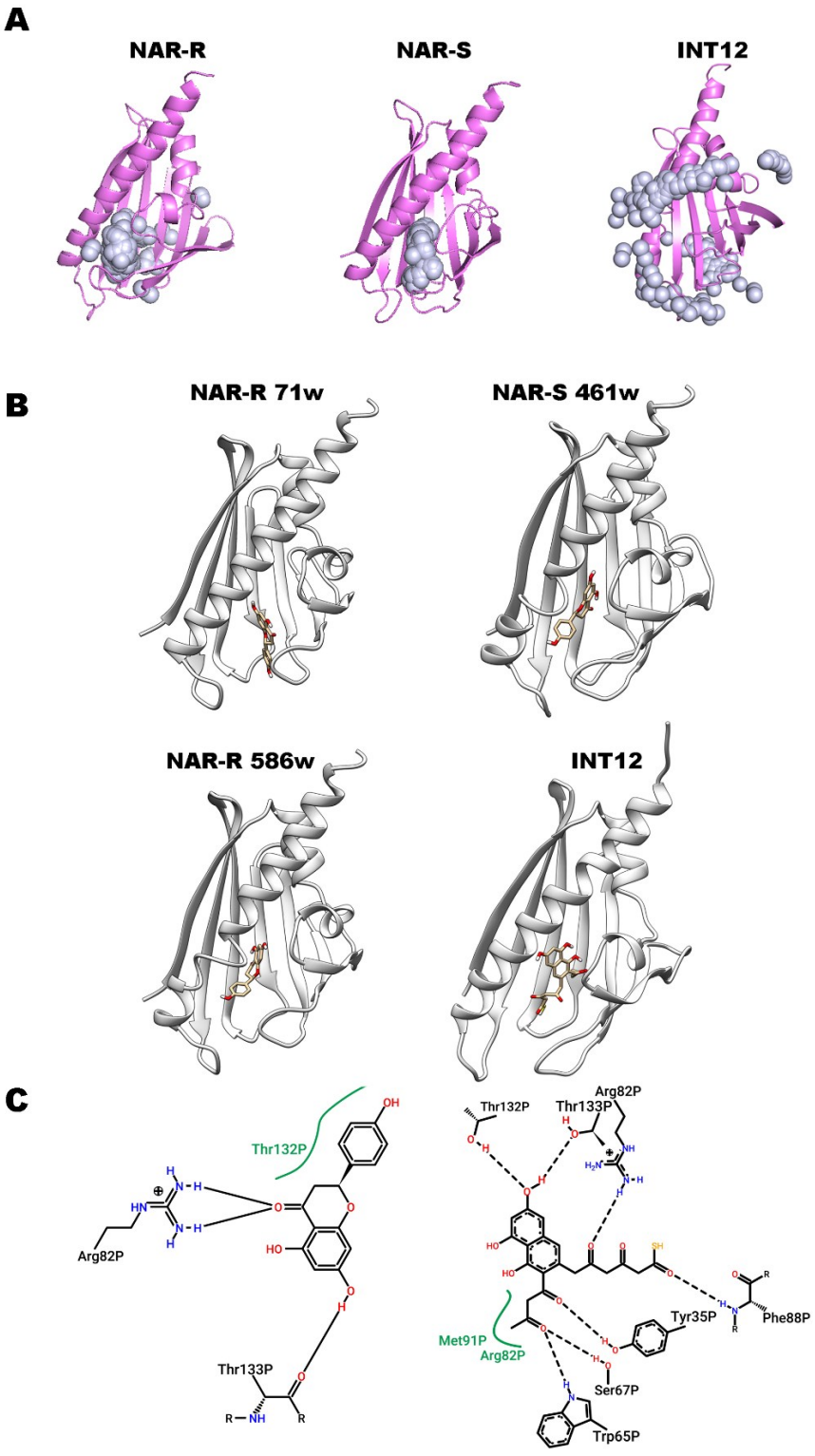

Figure 24 – Docking analysis.

(a) Ribbon representation of TcmN with the position of the centers of mass docking poses represented in gray dots (b) Naringenin-bound TcmN poses from Haddock and INT12-bound TcmN pose from Autodock chosen to seed MD simulation. C) 2D diagram of the output docking simulation showing the predicted binding site residues formed between TcmN and ligands.

Each selected pose of the NAR-R, NAR-S, and INT12-bound TcmN was used to seed ten independent replicates of 1 μs of MD simulation, and ten replicates of 1 μs of MD simulation for the free TcmN were carried out using the same force field to allow a comparison of the results. The results of all replicates of TcmN bound with NAR-R and NAR-S, as well as the trajectories of free TcmN and INT12-bound TcmN, were analyzed. General visualization of the trajectories of MD simulation results showed that NAR-R leaves the catalytic cavity of TcmN more times than NAR-S, and INT12 remains within the cavity in all replicates. The distance between the center of mass of TcmN and NAR-R and NAR-S was calculated. By visual inspection of the MD trajectories, it was stipulated that when this distance exceeds 20Å, the ligand has left the catalytic cavity of the enzyme. Ligand diffuses away from the TcmN pocket in 19% of the frames of all replicas with the NAR-R 71w pose, 14% of the frames with the NAR-R 586w pose, and 4% of the frames with the NAR-S 461w pose (Figure 25). This result suggests that the interaction of TcmN with the R isomer of Naringenin is weaker than with the S isomer. Because of these results, only the simulations performed with the NAR-S 461w pose were considered for further analysis.

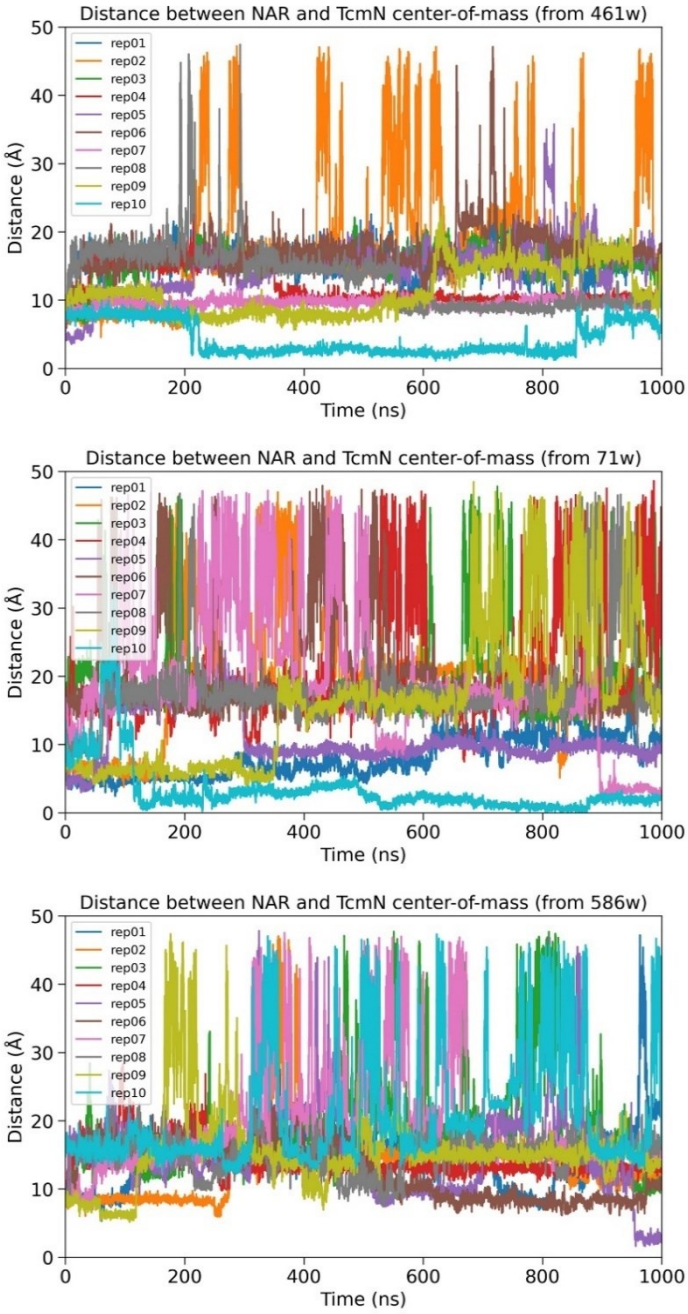

Figure 25 – Distance between the center of mass of TcmN and Naringenin in MD simulations.

To evaluate the conformational dynamics of free and bound TcmN, histograms of the trajectories of all MD simulation replicas were produced. The RMSD of INT12-bound TcmN is higher than the RMSD of the free and Naringenin-bound TcmN, meaning that the structure of the TcmN varied more from the initial pose when bound to INT12 (Figure 26A). The gyrate radius shows that INT12-bound TcmN assumes a larger volume than free and Naringenin-bound TcmN (Figure 26B). Cα RMSF values were calculated to evaluate the atomic-level dynamics of TcmN. The result showed elevated flexibility at L5 and L9 for all simulations but

higher flexibility of L3 in INT12-bound TcmN (Figure 26C). The distance between the L5 and L9 loops, which control the opening and closing of TcmN's cavity, was calculated, and it was seen that TcmN assumes two populations, closed and opened (Figure 26D). The opened TcmN population in the simulations done with free and Naringenin-bound TcmN is much less frequent, which shows that INT12 shifts the TcmN conformation to an opened loop population. Naringenin is a smaller molecule than INT12, which may suggest that TcmN does not have to undergo a significant conformational change to accommodate Naringenin. PCA was used to identify dominant motions from the MD trajectories. The first two components, PC1, and PC2, represented 43 and 9% of the Cα conformational variance, respectively (Figure 26E). Residues comprising the L3, L5, L7, and L9 loops located at the entrance of the central cavity, were significant contributors to the conformational variance of the PC1 and PC2 modes (Figure 26G). PC1 correlated strongly with L5-L9 distance (Figure 26H). TcmN's conformational plasticity is crucial for its enzymatic function because, depending on the stage of the chemical reaction, the enzyme needs to adopt significantly different conformations. Free TcmN assumes a more closed conformation; to interact with the linear polyketide and INT12, the protein becomes more open to accommodate the ligand.

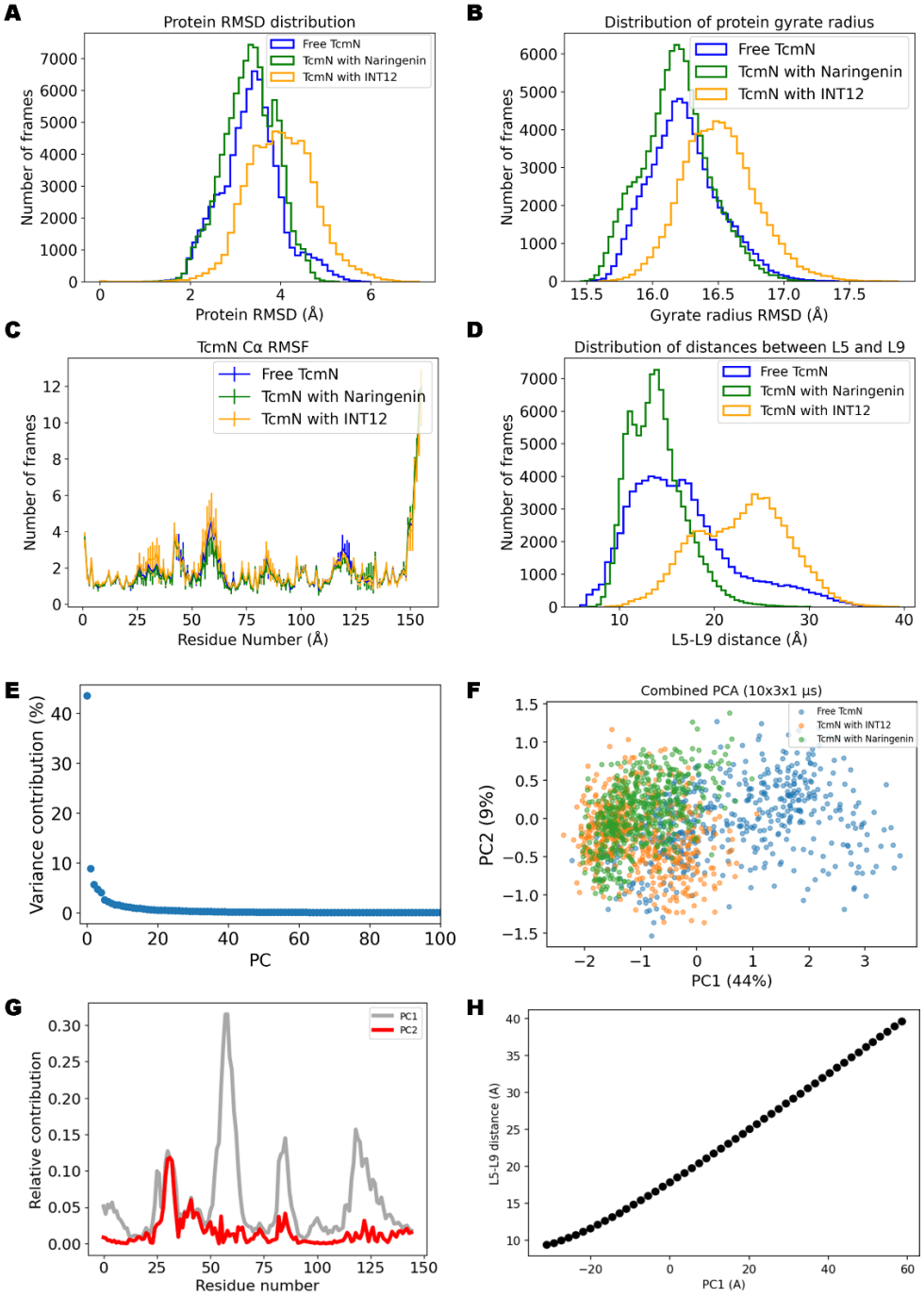

Figure 26 – Distribution of TcmN conformations.

A) Histogram representing the TcmN RMSD distribution from MD simulations. B) Histogram representing the TcmN gyrate radius distribution from MD simulations. C) Average RMSF calculated from Cα coordinates in MD simulations as a function of TcmN residue number. D) Histogram representing the distribution of distances between R61(L5) - P119(L9) from MD simulations. E) Variance explained by the principal components. F) Combined PCA of Free TcmN, TcmN with Naringenin, and TcmN with INT12. G) Per residue contributions to PC1 and PC2 in Å. H) L5-L9 distance along PC1.

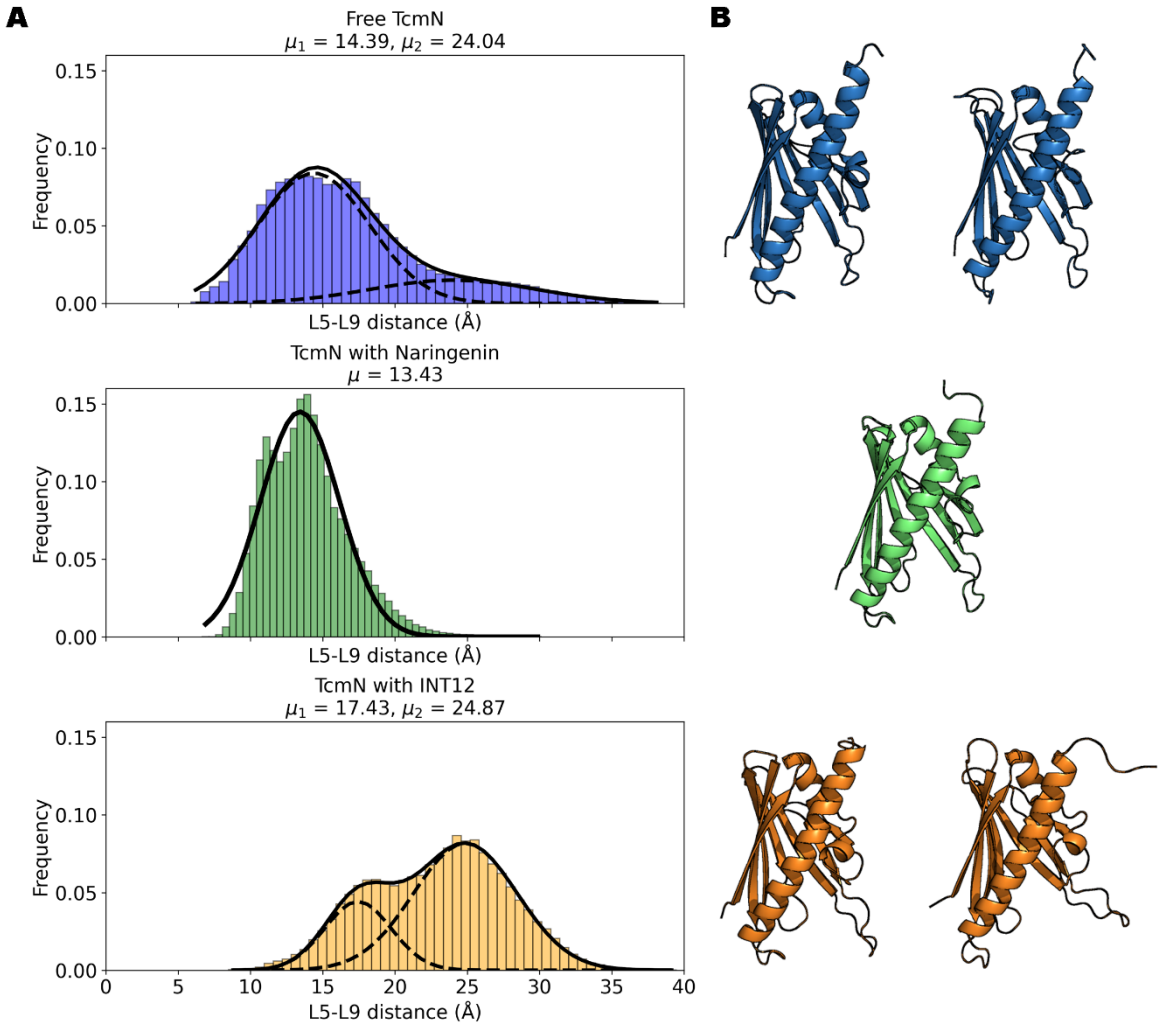

Figure 27 - L5-L9 distance variance on MD simulations

A) Histogram representing the distribution of distances between R61(L5) - P119(L9) from MD simulations of Free TcmN, TcmN with Naringenin and of TcmN with INT12. B) Ribbon representation of TcmN poses of the average distances between L5 and L9 presented in the histograms in A.

5 CONCLUSIONS

This work characterized TcmN's conformational diversity. The naringenin-bound TcmN melting temperature, Tm, measured from DSC experiments, was higher than the free enzyme one. The naringenin binding site on TcmN, mapped by NMR experiments, was located in the core TcmN cavity. Conformational dynamics on the µs-ms timescale were detected for residues in the substrate-binding cavity for the naringenin-bound TcmN. MD simulation results suggest that INT12 shifts the TcmN conformation to an opened loop population. TcmN's ability to adopt

various conformations is essential for its enzymatic activity, as the enzyme must undergo significant conformational changes at different stages of the chemical reaction. In its free state, TcmN adopts a more closed conformation. The protein opens up to accommodate the ligand when interacting with the linear polyketide and INT12.

6 REFERENCES

ABRAHAM, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, v. 1–2, p. 19–25, 2015.

AMES, BRIAN D. *et al.* Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 14, p. 5349–5354, 8 abr. 2008. Acesso em: 10 dez. 2019.

AMES, BRIAN D. *et al.* Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. *Biochemistry*, v. 50, n. 39, p. 8392–8406, 2011.

AMES, BRIAN DOUGLAS *et al.* Crystal structure and functional analysis of tetracenomycin ARO / CYC : Implications for cyclization specificity of aromatic polyketides. v. 105, n. 51, p. 20147–20148, 2008.

ANGULO, J.; NIETO, P. M. STD-NMR: application to transient interactions between biomolecules—a quantitative approach. *European Biophysics Journal*, v. 40, n. 12, p. 1357–1369, 2011.

ASAM, C. *et al.* Bet v 1 – a Trojan horse for small ligands boosting allergic sensitization? *Clinical & Experimental Allergy*, v. 44, n. 8, p. 1083–1093, 1 ago. 2014. Disponível em: <https://doi.org/10.1111/cea.12361>.

BISANG, C. *et al.* A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, v. 401, n. 6752, p. 502–505, 1999. Disponível em: <https://doi.org/10.1038/46829>.

CALDARA-FESTIN, G. *et al.* Structural and functional analysis of two di-domain aromatase/cyclases from type II polyketide synthases. *Proceedings of the National Academy of Sciences of the United States of America*, v. 112, n. 50, p. E6844–E6851, 2015.

CARR, H. Y.; PURCELL, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review*, v. 94, n. 3, p. 630–638, 1 maio 1954.

CHEN, A.; RE, R. N.; BURKART, M. D. Type II fatty acid and polyketide synthases: deciphering protein-protein and protein-substrate interactions. *Natural Product Reports*, v. 35, n. 10, p. 1029–1045, 2018.

CINO, E. A.; CHOY, W.-Y.; KARTTUNEN, M. Characterization of the Free State Ensemble of the CoRNR Box Motif by Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*, v. 120, n. 6, p. 1060–1068, 18 fev. 2016.

CINO, E. A.; CHOY, W.-Y.; KARTTUNEN, M. Conformational Biases of Linear Motifs. *The Journal of Physical Chemistry B*, v. 117, n. 50, p. 15943–15957, 19 dez. 2013.

DEPRISTO, M. A.; WEINREICH, D. M.; HARTL, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Reviews Genetics*, v. 6, n. 9, p. 678–687, 2005.

DISEASES, K.; BAY, C. W.; KONG, H. NMRPipe : A multidimensional spectral processing system based on UNIX pipes. v. 6, p. 277–293, 1995.

ESTER, M. *et al.* A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD'96, 1996, [S.l.]: AAAI Press, 1996. p. 226–231.

FERROLINO, M. C. *et al.* Delicate Balance between Functionally Required Flexibility and Aggregation Risk in a β-Rich Protein. *Biochemistry*, v. 52, n. 49, p. 8843–8854, 10 dez. 2013.

FITZGERALD, J. T. *et al.* Analysis and refactoring of the A-74528 biosynthetic pathway. *Journal of the American Chemical Society*, v. 135, n. 10, p. 3752–3755, 13 mar. 2013.

GERSHENSON, A. *et al.* Energy landscapes of functional proteins are inherently risky. *Nature Chemical Biology*, v. 10, n. 11, p. 884–891, 2014.

GOMES, E. S.; SCHUCH, V.; LEMOS, E. G. DE M. Biotechnology of polyketides: New breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Brazilian Journal of Microbiology*, v. 44, n. 4, p. 1007–1034, 2013.

HEINZELMANN, G.; HENRIKSEN, N. M.; GILSON, M. K. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *Journal of Chemical Theory and Computation*, v. 13, n. 7, p. 3260–3275, 2017.

HERTWECK, C. *et al.* Type II polyketide synthases: Gaining a deeper insight into enzymatic teamwork. *Natural Product Reports*, v. 24, n. 1, p. 162–190, 2007.

HUANG, J.; MACKERELL, A. D. J. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *Journal of computational chemistry*, v. 34, n. 25, p. 2135–2145, set. 2013.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, v. 14, n. 1, p. 33–38, 1996.

HYBERTS, S. G. *et al.* Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *Journal of Biomolecular NMR*, v. 52, n. 4, p. 315–327, 2012.

HYBERTS, S. G.; TAKEUCHI, K.; WAGNER, G. Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data. *Journal of the American Chemical Society*, v. 132, n. 7, p. 2145–2147, 24 fev. 2010.

JAVIDPOUR, P. *et al.* The determinants of activity and specificity in actinorhodin type II polyketide ketoreductase. *Chemistry and Biology*, v. 20, n. 10, p. 1225–1234, 2013.

KARTTUNEN, M.; CHOY, W.-Y.; CINO, E. A. Prediction of Binding Energy of Keap1 Interaction Motifs in the Nrf2 Antioxidant Pathway and Design of Potential High-Affinity Peptides. *The Journal of Physical Chemistry B*, v. 122, n. 22, p. 5851–5859, 7 jun. 2018.

KATSUYAMA, Y.; OHNISHI, Y. *Type III polyketide synthases in microorganisms*. 1. ed. [S.l.]: Elsevier Inc., 2012. v. 515.

KEATINGE-CLAY, A. T. *et al.* An antibiotic factory caught in action. *Nature Structural & Molecular Biology*, v. 11, n. 9, p. 888–893, 2004. Disponível em: <https://doi.org/10.1038/nsmb808>.

KOFLER, S. *et al.* Crystallographically Mapped Ligand Binding Differs in High and Low IgE Binding Isoforms of Birch Pollen Allergen Bet v 1. *Journal of Molecular Biology*, v. 422, n. 1, p. 109–123, 2012.

LEE, M. Y.; AMES, B. D.; TSAI, S. C. Insight into the molecular basis of aromatic polyketide cyclization: Crystal structure and in vitro characterization of WhiE-ORFVI. *Biochemistry*, v. 51, n. 14, p. 3079–3091, 2012.

LEŠNIK, U. *et al.* Construction of a New Class of Tetracycline Lead Structures with Potent Antibacterial Activity through Biosynthetic Engineering. *Angewandte Chemie International Edition*, v. 54, n. 13, p. 3937–3940, 23 mar. 2015.

LI, J.; ZHANG, L.; LIU, W. Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. *Synthetic and Systems Biotechnology*, v. 3, n. 2, p. 83–89, 2018.

LIU, X. *et al.* Heterologous Biosynthesis of Type II Polyketide Products Using E. coli. *ACS Chemical Biology*, 11 dez. 2019.

LORIA, J. P.; RANCE, M.; PALMER, A. G. A Relaxation-Compensated Carr−Purcell−Meiboom−Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *Journal of the American Chemical Society*, v. 121, n. 10, p. 2331–2332, 1 mar. 1999.

MATSUOKA, D. *et al.* Molecular Dynamics Simulations of Heart-type Fatty Acid Binding Protein in Apo and Holo Forms, and Hydration Structure Analyses in the Binding Cavity. *The Journal of Physical Chemistry B*, v. 119, n. 1, p. 114–127, 8 jan. 2015.

MAYER, M.; MEYER, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *Journal of the American Chemical Society*, v. 123, n. 25, p. 6108–6117, 1 jun. 2001.

MAZURENKO, S. *et al.* CalFitter: a web server for analysis of protein thermal denaturation data. *Nucleic Acids Research*, v. 46, n. W1, p. W344–W349, 14 maio 2018.

MCDANIEL, R., HUTCHINSON, C. R., KHOSLA, C. OF THE AMERICAN CHEMICAL Engineered Biosynthesis of Novel Polyketides : Analysis of. v. 117, n. 26, p. 1546–1550, 1995.

MEIBOOM, S.; GILL, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments*, v. 29, p. 688–691, 1 ago. 1958.

MOGENSEN, J. E. *et al.* The major birch allergen, Bet v 1, shows affinity for a broad spectrum of physiological ligands. *The Journal of biological chemistry*, v. 277, n. 26, p. 23684–92, jun. 2002.

MORAES, A. H. *et al.* 1H, 13C and 15N resonance assignments and second structure information of Fag s 1: Fagales allergen from Fagus sylvatica. *Biomolecular NMR assignments*, v. 10, n. 1, p. 45–48, abr. 2016.

MORAES, A. H. *et al.* Structural basis for cross-reactivity and conformation fluctuation of the major beech pollen allergen Fag s 1. *Scientific Reports*, v. 8, n. 1, p. 10512, 2018. Disponível em: <https://doi.org/10.1038/s41598-018-28358-1>.

MOSS, G. P.; SMITH, P. A. S.; TAVERNIER, D. *Glossary of class names of organic compounds and reactivity intermediates based on structure (IUPAC Recommendations 1995)*. *Pure and Applied Chemistry*. [S.l: s.n.]. , 1995

OLSSON, M. H. M. *et al.* PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation*, v. 7, n. 2, p. 525–537, 8 fev. 2011.

OROGUCHI, T.; NAKASAKO, M. Changes in hydration structure are necessary for collective motions of a multi-domain protein. *Scientific Reports*, v. 6, n. 1, p. 26302, maio 2016a.

OROGUCHI, T.; NAKASAKO, M. Changes in hydration structure are necessary for collective motions of a multi-domain protein. *Scientific Reports*, v. 6, n. 1, p. 26302, maio 2016b.

PALMER, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chemical Reviews*, v. 104, n. 8, p. 3623–3640, 1 ago. 2004.

PETTERSEN, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, v. 25, n. 13, p. 1605–1612, out. 2004.

PFEIFER, B. A.; KHOSLA, C. Biosynthesis of Polyketides in Heterologous Hosts. *Microbiology and Molecular Biology Reviews*, v. 65, n. 1, p. 106–118, 2001.

PIANA, S.; LINDORFF-LARSEN, K.; SHAW, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophysical Journal*, v. 100, n. 9, p. L47–L49, 4 maio 2011.

RADAUER, C.; LACKNER, P.; BREITENEDER, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC evolutionary biology*, v. 8, p. 1–19, jan. 2008.

RAWLINGS, B. J. Biosynthesis of polyketides (other than actinomycete macrolides). *Natural Product Reports*, v. 16, n. 4, p. 425–484, 1999.

RIVERS, M. A. J.; LOWELL, A. N. Expanding the Biosynthetic Toolbox: The Potential and Challenges of In Vitro Type II Polyketide Synthase Research. *SynBio*, v. 2, n. 1, p. 85–111, 2024. Disponível em: <https://www.mdpi.com/2674-0583/2/1/6>.

ROHR, J.; HERTWECK, C. Type II PKS. *Comprehensive Natural Products II: Chemistry and Biology*, v. 1, p. 227–303, 2010.

SUMMERS, R. G. *et al.* Nucleotide sequence of the tcmII-tcmIV region of the tetracenomycin C biosynthetic gene cluster of Streptomyces glaucescens and evidence that the tcmN gene encodes a multifunctional cyclase-dehydratase-O-methyl transferase. *Journal of Bacteriology*, v. 174, n. 6, p. 1810–1820, 1 mar. 1992. Disponível em: <https://doi.org/10.1128/jb.174.6.1810-1820.1992>.

TANG, Y. *et al.* Ketosynthases in the Initiation and Elongation Modules of Aromatic Polyketide Synthases Have Orthogonal Acyl Carrier Protein Specificity. *Biochemistry*, v. 42, n. 21, p. 6588–6595, 2003. Disponível em: <https://doi.org/10.1021/bi0341962>.

TSUJISHITA, Y.; HURLEY, J. H. Structure and lipid transport mechanism of a StAR-related domain. *Nature structural biology*, v. 7, n. 5, p. 408–414, 2000.

UEBERSCHAAR, N. *et al.* Synthetic Remodeling of the Chartreusin Pathway to Tune Antiproliferative and Antibacterial Activities. *Journal of the American Chemical Society*, v. 135, n. 46, p. 17408–17416, 20 nov. 2013.

VALADARES, V. S. *et al.* Conformational dynamics of Tetracenomycin aromatase/cyclase regulate polyketide binding and enzyme aggregation propensity. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 1865, n. 9, p. 129949, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0304416521001070>.

VAN DER MAATEN, L.; HINTON, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, v. 9, n. 86, p. 2579–2605, 2008.

VIRTANEN, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, v. 17, n. 3, p. 261–272, mar. 2020.

VRANKEN, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, v. 59, n. 4, p. 687–696, 1 jun. 2005.

WAUDBY, C. A. *et al.* Two-Dimensional NMR Lineshape Analysis. *Scientific Reports*, v. 6, n. 1, p. 24826, 2016. Disponível em: <https://doi.org/10.1038/srep24826>.

WAUDBY, C. A. *et al.* Two-dimensional NMR lineshape analysis of single, multiple, zero and double quantum correlation experiments. *Journal of Biomolecular NMR*, v. 74, n. 1, p. 95–109, 2020. Disponível em: <https://doi.org/10.1007/s10858-019-00297-7>.

WEISSMAN, K. J. Chapter 1 Introduction to Polyketide Biosynthesis. *Methods in Enzymology*, v. 459, n. B, p. 3–16, 2009.

WEISSMAN, K. J.; LEADLAY, P. F. Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*, v. 3, n. 12, p. 925–936, 2005.

ZHANG, Z.; PAN, H.-X.; TANG, G.-L. New insights into bacterial type II polyketide biosynthesis. *F1000Research*, v. 6, p. 172, 21 fev. 2017.