

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOQUÍMICA E IMUNOLOGIA**

VERONICA SILVA VALADARES

**BIOPHYSICAL CHARACTERIZATION OF THE N-TERMINAL DOMAIN OF
TETRACENOMYCIN ARO/CYC**

**BELO HORIZONTE
2020**

VERONICA SILVA VALADARES

**BIOPHYSICAL CHARACTERIZATION OF THE N-TERMINAL DOMAIN OF
TETRACENOMYCIN ARO/CYC**

**Orientador: Elio Anthony Cino, PhD
Professor Adjunto do Departamento de
Bioquímica e Imunologia da Universidade
Federal de Minas Gerais**

**Coorientador: Adolfo Henrique de Moraes
Silva, PhD
Professor Adjunto do Departamento de
Química da Universidade Federal de Minas
Gerais**

**Dissertação submetida ao Departamento de
Bioquímica e Imunologia do Instituto de
Ciências Biológicas da Universidade Federal
de Minas Gerais, como requisito parcial para
a obtenção do grau de Mestre em Bioquímica
e Imunologia.**

BELO HORIZONTE

2020

- 043 Valadares, Veronica Silva.
Biophysical characterization of the N-terminal domain of tetracenomycin ARO/CYC [manuscrito] / Veronica Silva Valadares. – 2020.
64 f. : il. ; 29,5 cm.
- Orientador: Elio Anthony Cino. Coorientador: Adolfo Henrique de Moraes Silva.
- Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Departamento de Bioquímica e Imunologia.
1. Bioquímica. 2. Aromatase. 3. Ciclização. 4. Policetídeos. 5. Tetraciclina. I. Cino, Elio Anthony. II. Moraes, Adolfo Henrique de. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 577.1



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

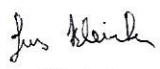


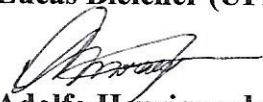
ATA DA DEFESA DA DISSERTAÇÃO DE MESTRADO DE VERONICA SILVA VALADARES. Aos vinte e oito dias do mês de fevereiro de 2020 às 09:00 horas, reuniu-se no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, a Comissão Examinadora da dissertação de Mestrado, indicada *ad referendum* do Colegiado do Curso, para julgar, em exame final, o trabalho intitulado "Biophysical Characterization of the N-Terminal Domain of Tetracenomycin ARO/CYC", requisito final para a obtenção do grau de Mestre em Bioquímica e Imunologia, área de concentração: Bioquímica. Abrindo a sessão, o Presidente da Comissão, Prof. Elio Anthony Cino, 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. Liza Figueiredo Felicori Vilela (Universidade Federal de Minas Gerais), aprovada; Dr. Lucas Bleicher (Universidade Federal de Minas Gerais), aprovada; Dr. Adolfo Henrique de Moraes Silva - Coorientador (Universidade Federal de Minas Gerais), aprovada; Dr. Elio Anthony Cino - Orientador (Universidade Federal de Minas Gerais), aprovada. Pelas indicações a candidata foi considerada:

- APROVADA
 REPROVADA


O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 28 de fevereiro de 2020.


 Dr. Liza Figueiredo Felicori Vilela (UFMG)


 Dr. Lucas Bleicher (UFMG)


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


 Dr. Elio Anthony Cino - Orientador (UFMG)


 Profª Maria de Fátima Martins Hort
 Sub Coordenadora do Curso de Pós-Graduação
 em Bioquímica e Imunologia
 ICB-UFMG

THIS STUDY WAS FINANCED IN PART BY THE FUNDAÇÃO DE AMPARO À PESQUISA DE MINAS GERAIS (FAPEMIG) AND BY THE BRAZILIAN INNOVATION AND RESEARCH FUNDING AGENCY (FINEP – GRANT MCTI/FINEP/ FNDCT 01/2016/01.16.0050.00).

ACKNOWLEDGEMENT

Agradeço aos meus pais, Vagner e Romilda, ao meu irmão Filipe e à minha avó Ruth, que são os principais apoiadores de todas as minhas conquistas. Vocês são o meu exemplo de amor, perseverança, força e fé. Obrigada pelo amor incondicional, por acreditarem em mim, por me proporcionarem oportunidades para enfrentar o mundo, encarar os meus sonhos e minhas dificuldades e por me darem forças para lutar e alcançar meus objetivos. Amo vocês!

Agradeço aos meus orientadores, prof. Adolfo e prof. Elio pela orientação e imenso apoio ao longo desses dois anos. Sou muito grata pela paciência, sabedoria, conselhos e incentivos que recebi de vocês.

Ao Luan, pela colaboração e ajuda com os dados de MD.

Aos amigos do Laboratório Macromol, passados e presentes, pelo companheirismo, conversas durante os cafés e os momentos de descontração: Philipe, Mozart, Raquel, Aline, Samara, Carlos, Diego, prof Tiago, profa Amanda, Yan e Brunno.

Aos amigos que sempre estiveram ao meu lado, em especial a Bruna, Marina, Luísa, Pedro e Fábio por fazerem parte da minha história, por me entenderem, me aguentarem e terem tornado a caminhada mais leve e feliz no tempo do mestrado.

Ao grupo Pé Descalço e às amigas que o forró de BH me trouxe, em especial o André, Elvys e Álvaro, por me acolherem e pelos diversos momentos de alegria. A dança me ajudou a lidar melhor com as minhas emoções, ansiedade, autoconfiança e autoestima. Benefícios que foram importantes para enfrentar os desafios do mestrado e que se estendem a vários campos da minha vida.

À Universidade Federal de Minas Gerais por tudo o que me foi concedido. À FAPEMIG pelo financiamento e pelas oportunidades empreendedoras que o DELA me trouxe.

Agradeço aos membros da banca por aceitarem prontamente o convite para participarem da avaliação dessa dissertação.

Agradeço a todos que fizeram parte da minha vida nesses dois anos e que, de alguma maneira, tornaram a realização deste sonho possível.

ABSTRACT

Tetracenomycin aromatase/cyclase (Tcm) is an enzyme derived from *Streptomyces glaucescens* involved in polyketide cyclization, aromatization, and folding. Polyketides are a diverse class of secondary metabolites produced by certain groups of bacteria, fungi, and plants with pharmaceutical applications. Examples include antibiotics, such as tetracycline, and anticancer drugs, such as doxorubicin. The N-terminal domain of Tcm (TcmN) participates in the specific cyclization of polyketides, and is classified as belonging to the Bet v 1-like superfamily. TcmN catalyzes the formation of the first (C9-C142) and second (C7-C16) rings of polyketides *in vivo* and *in vitro*. This work characterized the stability and conformational diversity TcmN, which is important for its possible use in industrial applications, such as engineering polyketide biosynthesis. TcmN was expressed in *Escherichia coli* BL21(DE3) and purified. Secondary structure and stability were assessed in different buffer conditions by circular dichroism (CD) and fluorescence spectroscopy. Nuclear Magnetic Resonance (NMR) spectra were acquired at different field strengths (900, 800, and 600 MHz). 6 x 1 μ s Molecular Dynamics (MD) simulations of TcmN were performed in explicit solvent. Thermal denaturation of TcmN is irreversible, and the denaturation temperature is reduced at higher concentrations. TcmN is most stable around pH 8, and in the presence of NaI and NaCl. NMR and MD findings suggest that substrate binding and product release can be modulated by flexibility of specific loop segments. Together, the results indicate that TcmN exists in equilibrium between an open conformation that favors ligand binding in the main hydrophobic cavity, and a closed state that protects it against solvent exposure and aggregation.

Keywords: Aromatase, Aromatization, ARO/CYC, Cyclase, Cyclization, NMR, Polyketides, Tetracenomycin, Tcm, TcmN

RESUMO

A tetracenomicina aromatase/ciclase (Tcm) é uma enzima presente em *Streptomyces glaucescens* envolvida na ciclização, aromatização e enovelamento de policetídeos. Os policetídeos são uma classe diversificada de metabólitos secundários produzidos por certos grupos de bactérias, fungos e plantas, com aplicações farmacêuticas. Exemplos incluem antibióticos, como a tetraciclina, e drogas anticâncer, como a doxorrubicina. O domínio N-terminal da Tcm (TcmN) participa de ciclizações específicas de policetídeos e é classificada como pertencente à superfamília Bet v 1. A TcmN catalisa a formação do primeiro (C9-C14) e segundo (C7-C16) anéis de policetídeos *in vivo* e *in vitro*. Este trabalho caracterizou a estabilidade e a diversidade conformacional da TcmN, informações importantes para o possível uso da TcmN em aplicações industriais, como na engenharia da biossíntese de policetídeos. A TcmN foi expressa em *Escherichia coli* BL21 (DE3) e purificada. Sua estrutura secundária e estabilidade foram avaliadas em diferentes condições de tampão por dicróismo circular e espectroscopia de fluorescência. Os espectros de Ressonância Magnética Nuclear foram adquiridos em espectrômetros de diferentes campos magnéticos (900, 800 e 600 MHz). 6 x 1 μ s de simulações de Dinâmica Molecular da TcmN foram realizadas em solvente explícito. A desnaturação térmica da TcmN é irreversível e a temperatura de desnaturação é reduzida em concentrações mais altas da proteína. A TcmN é mais estável em torno de pH 8 e na presença de NaI e NaCl. Os resultados de ressonância magnética nuclear e de dinâmica molecular sugerem que a ligação do substrato e a liberação do produto podem ser moduladas pela flexibilidade de algumas alças. Juntos, os resultados apresentados neste trabalho sugerem que a TcmN existe em equilíbrio entre a conformação aberta, que favorece a ligação do ligante na cavidade hidrofóbica, e um principal estado fechado, que o protege a cavidade contra a exposição de solventes e à agregação.

Palavras-chave: Aromatase, Aromatização, ARO/CYC, Ciclase, Ciclização, RMN, Policetídeos, Tetracenomicina, Tcm, TcmN

LIST OF FIGURES

Figure 1 – The state of polyketones during aromatic PKS iterative elongation.....	11
Figure 2 – Schematic diagram of ARO/CYC activity and cyclization specificity in representative type II PKSs.	13
Figure 3 – TcmN structure.	15
Figure 4 – Comparison of TcmN, WhiE and ZhuI.	17
Figure 5 – Proposed biosynthetic pathway for Tcm C (1), RM80 (4), RM80b (5), SEK15 (6), and SEK15b (7).....	18
Figure 6 – Energy landscape of protein folding and misfolding.	19
Figure 7 – CD effect on the circularly polarized components A_L and A_R	20
Figure 8 – Typical fav-UV spectra for α -helices, β -sheets, and random coil secondary structures.....	21
Figure 9 – Perri-Jablonski diagram including vibrational levels for absorbance, non-radiative transition, and fluorescence.	22
Figure 10 – Protein motion and NMR timescales.....	23
Figure 11 – TcmN expression test.	31
Figure 12 – 12% SDS-PAGE of TcmN Ni ²⁺ affinity purification.	32
Figure 13 – Overlapping of the CBCACONH spectra (pink) with the HNCACB (blue and green).....	33
Figure 14 – TcmN NMR assignment.	34
Figure 15 – Stability of TcmN against temperature and pH change.....	36
Figure 16 – Thermal denaturation of TcmN under different ions type.....	37
Figure 17 – Thermal denaturation of TcmN under different concentrations.	38
Figure 18 – Aggregation assay of TcmN and of the aggregation-prone protein p53.	39
Figure 19 – TcmN transient self-interaction monitored by NMR.	40
Figure 20 – Backbone molecular dynamics of TcmN at ps-ns timescale monitored by NMR and MD analysis.	41
Figure 21 – Backbone molecular dynamics of TcmN monitored by NMR.	42
Figure 22 – Side chains of residues showing R_{ex} above 3 Hz colored in orange on the TcmN crystal structure.....	43
Figure 23 – Residue-specific fluctuations of TcmN from the MD simulations.	44
Figure 24 – TcmN conformational diversity.	45
Figure 25 – TcmN aggregation propensity.....	47

Figure 26 – Proposed interplay between conformational flexibility and aggregation of TcmN.	48
Figure 28 – TcmN and ANS interaction.	49
Figure 29 – 2D diagram of docking results of ANS.	50
Figure 30 – TcmN and Naringerin interaction.	51
Figure 31 – 2D diagram of docking results of Naringerin.	52

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 ₂ O	Deuterium Oxide
Da	Dalton
<i>E. coli</i>	<i>Escherichia coli</i>
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
MRE	Molar residue ellipticity
N	Nitrogen
NaCl	Sodium chloride
NaI	Sodium Iodide
NaPO ₄	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
S ²	Order parameter
TcmN	N-terminal of Tcm ARO/CYC protein
ThT	Thioflavin T
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.....	10
1.1 Polyketide synthases.....	10
1.2 N-terminal domain of tetracenomycin ARO/CYC (TcmN).....	13
1.3 Protein aggregation	18
1.4 Techniques for characterizing protein structure and dynamics	20
CD spectroscopy.....	20
Fluorescence spectroscopy.....	21
NMR spectroscopy	22
MD simulations	23
2 OBJECTIVES	25
3 MATERIALS AND METHODS	26
3.1 Expression and purification of TcmN	26
3.2 Fluorescence spectroscopy	26
3.3 CD spectroscopy	27
3.4 NMR spectroscopy.....	27
3.5 MD simulations.....	29
4 RESULTS AND DISCUSSION	31
4.1 Expression and purification	31
4.2 TcmN backbone resonance assignment.....	32
4.3 Characterization of TcmN stability	35
4.4 TcmN flexibility and dynamics.....	40
4.5 TcmN interactions.....	48
5 CONCLUSION	52
6 REFERENCES.....	53

1 INTRODUCTION

1.1 Polyketide synthases

Polyketides are secondary metabolites produced by bacteria, fungi, and plants, comprising a large class of natural products with several biological activities and pharmaceutical properties. They are synthesized by polyketide synthases (PKSs) from repeated condensation of acetyl coenzyme A, via malonyl-CoA, giving rise to a molecular structure of alternating carbonyl and methylene groups with multiple β -hydroxyketone or β -hydroxyaldehyde functional groups (MOSS; SMITH; TAVERNIER, 1995) (Fig. 1 and 2).

Three types of PKSs have been described to date. Type I PKSs, also called modular PKSs, are large multi-domain enzymes with a series of active sites for catalyzing the various stages of polyketide synthesis. Type II PKSs produce aromatic polyketides and are composed of separate mono and bi-functional enzymes that interact during synthesis (GOMES; SCHUCH; LEMOS, 2013). Type III PKSs are homodimers of ketosynthases which 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), a chain length factor (CLF), an acyl carrier protein (ACP), and a malonyl transferase (MAT) (CHEN, A.; RE; BURKART, 2018). An initial precursor, malonyl-CoA, forms a thioester bond with ACP via MAT. Next, the KS-CLF elongates the acyl-ACP to a β -ketoacyl-ACP by a Claisen condensation reaction, and proceeds through an iterative pathway, elongating the polyketide chain by two carbons in each cycle (CHEN, A.; RE; BURKART, 2018) (Fig. 1). Additional enzymes, such as cyclases, aromatases, and ketoreductases, act on this chain producing aromatic and cyclic groups. After the release of the polyketide from the ACP, oxidases, transferases, and hydrolases may act on the chain to introduce further modifications (WEISSMAN, 2009). The versatility of type II PKSs, along with a broad array of post-PKS tailoring enzymes, makes the group of type II PKS-derived secondary metabolites one of the most structurally diverse groups found in nature (ROHR; HERTWECK, 2010). Several compounds produced by PKSs are used clinically, with some well-known examples including tetracyclines as antibiotics, and doxorubicin and mithramycin as anticancer drugs (WEISSMAN, 2009).

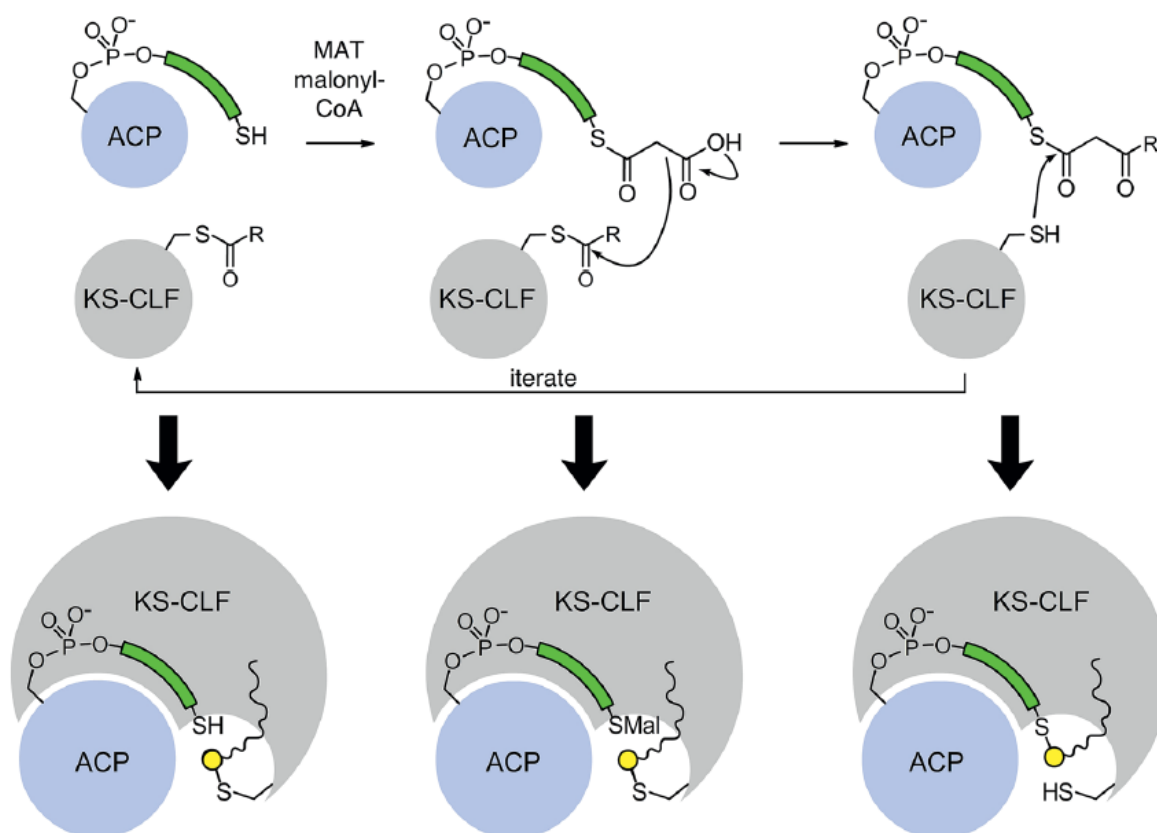


Figure 1 – The state of polyketones during aromatic PKS iterative elongation.

The ACP and KS-CLF adopt three states during aromatic polyketide biosynthesis: (1) holo-ACP and acylated KS-CLF; (2) malonyl-ACP and acylated KS-CLF; (3) b-ketoacyl-ACP and holo-KS-CLF. Each of these states involves the elongated polyketone intermediate bound within the KS-CLF, indicating that the KS-CLF must stabilize these highly reactive elongated species via yet unknown mechanisms. Adapted from Chen, A. et al, 2018. Usage of this figure has been granted by the Royal Society of Chemistry (license number 1013768-1).

The development of combinatorial biosynthesis by “mixing and matching” of different PKS components, including the combinatorial manipulation of ketoreductases (KRs), aromatases (AROs), and cyclases (CYCs), in heterologous expression systems has resulted in a large number of 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, their posttranslational modification requirements be met, 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 demonstrated to be an attractive means for the construction of new pharmaceuticals and bioactive molecules (LI; ZHANG; LIU, 2018). However, the use of the biosynthetic machinery *in vitro* to make new natural products through control of 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 to control cyclization and prevent aberrant cyclization events. Previous studies have established the biosynthetic routes for aromatic polyketides in “reducing” (KR present) and “non-reducing” (KR absent) PKS systems (AMES *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). In a non-reducing system, the growing poly- β -ketone intermediate is transported directly from the ketosynthase to the ARO/CYC. There are two types of ARO/CYCs in non-reducing systems (Fig. 2): 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 *et al.*, 2008; CALDARA-FESTIN *et al.*, 2015; MCDANIEL, R., HUTCHINSON, C. R., KHOSLA, 1995).

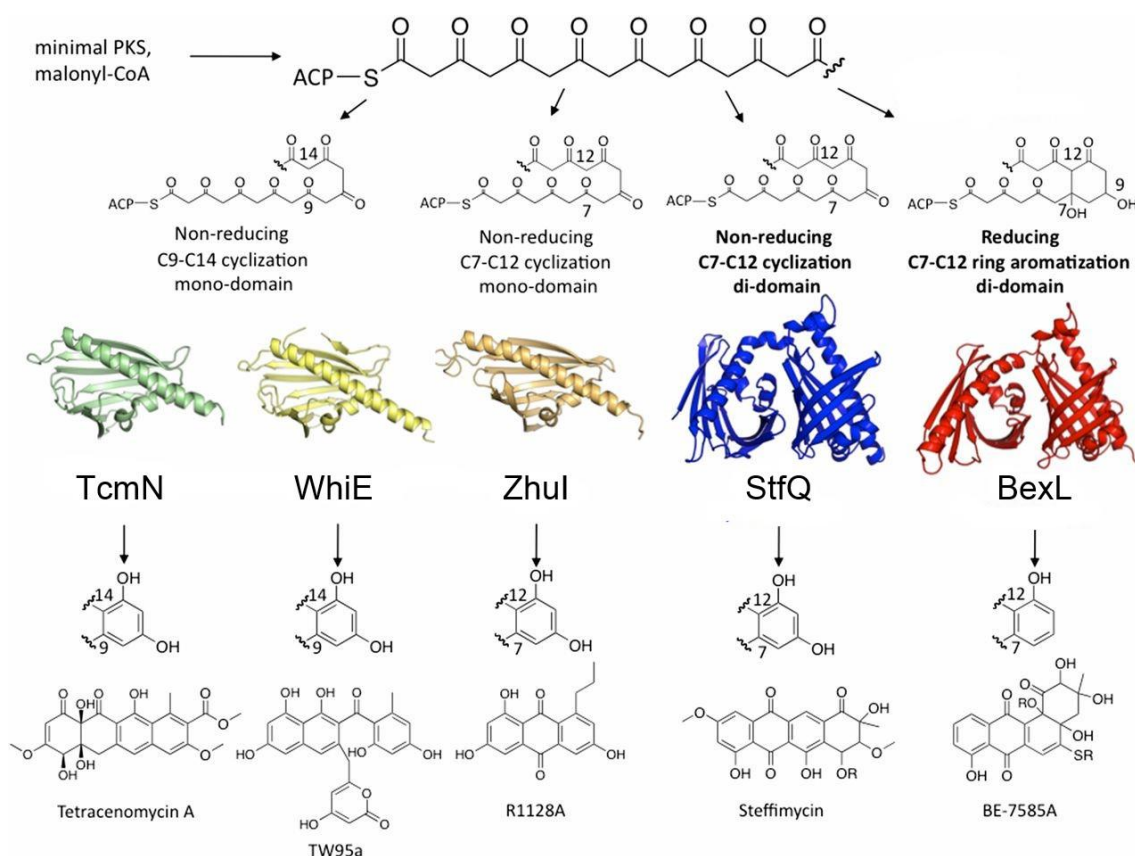


Figure 2 – Schematic diagram of ARO/CYC activity and cyclization specificity in representative type II PKSs.

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. 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)

TcmN is a bifunctional enzyme from *Streptomyces glaucescens* that promotes C9–C14 first ring, C7–C16 second-ring cyclization, and aromatization of unreduced linear polyketides (AMES *et al.*, 2011; MCDANIEL, R., HUTCHINSON, C. R., KHOSLA, 1995). In 2008, Ames and colleagues published the 1.9 Å crystal structure of TcmN (pdb accession codes 2RER, 2RES, and 2REZ) (AMES *et al.*, 2008). 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 beta-sandwich, and two small helices between β_1 and β_2 forming a helix–loop–helix motif that seals one end of the β -sandwich (Fig. 3a). Nascent polyketide intermediates bind to residues

within the interior pocket of TcmN where regiospecific cyclization and aromatization occurs. The deep pocket can accommodate an ACP-bound decaaketide (20 carbons). Pocket residues W28, F32, W65, S67, R69, M91, and W95 are highly conserved among monodomain, and the N-terminal half of didomain ARO/CYCs, while residues T54, L93, H128, T132, T133, and N136 are unique to TcmN. R69, T35, W08, and Q110 orient the polyketide chain inside the cavity, while S67 and R69 anchor the polyketide, promoting chain folding and C9-C14 first-ring cyclization. The pocket has a narrow neck region defined by residues W65, F88, L129, Y35, T132, T133, E34, and R82, which participate in C7-C16 second-ring cyclization (AMES *et al.*, 2008). The TcmN pocket can accommodate up to three linearly fused rings, and it is thought that the third ring cyclization likely happens spontaneously, after which 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 Tetracenomyacin (AMES *et al.*, 2008).

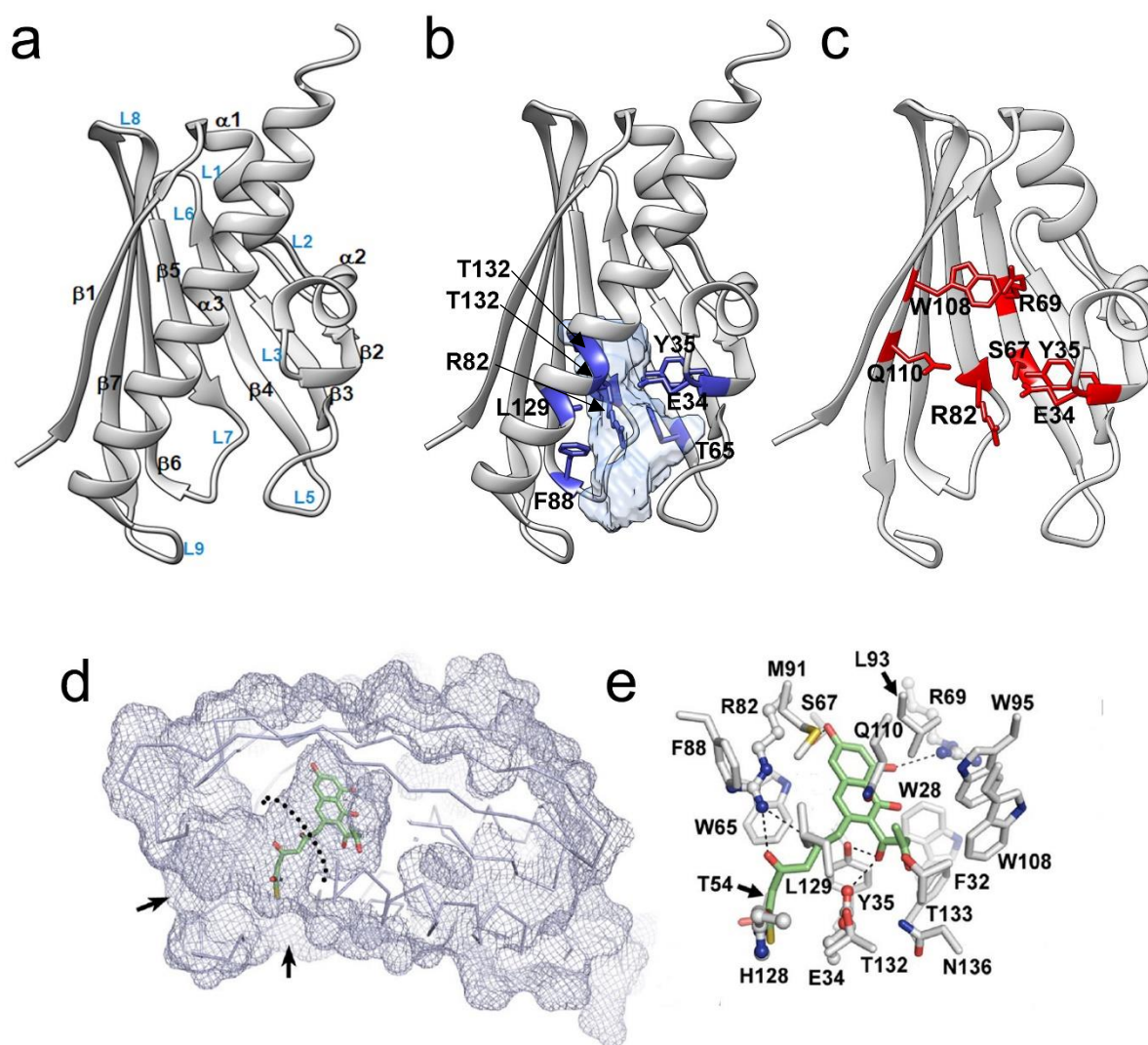


Figure 3 – TcmN structure.

(a) Ribbon representation of the TcmN crystal structure (pdb accession code 2RER) characterized by a helix-grip fold. (b) Narrow neck residues side chains represented by blue sticks. (c) Important residues side chains that orient, anchor and participate in cyclization and aromatization of the polyketide chains shown in red sticks. (d) Surface cutaway view of TcmN with docked intermediate Tcm F2 (green sticks) with two rings. The two entrances to the pocket are indicated by arrows and the narrow neck region is marked by dotted line. (e) Zoomed stereo view of (d), showing all pocket residues discussed in the section “N-Terminal of Tetracenomycin ARO/CYC”. 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.

Crystal structures of two other mono-domain ARO/CYCs from nonreducing type II PKS systems have also been obtained. WhiE (pdb accession code 3TVR) catalyzes the regiospecific C9–C14 and C7–C16 cyclization and aromatization of a 24-carbon polyketide chain (LEE; AMES; TSAI, 2012), and ZhuI (pdb accession code 3TFZ) acts on unreduced polyketide

intermediates to generate C7–C12 cyclized and aromatized products (AMES *et al.*, 2011). Sequence comparison reveals that WhiE shares an overall identity and similarity of 54 and 65% with TcmN, with 80 and 93%, respectively, for interior pocket residues. ZhuI shares 17 and 36% of overall identity and similarity with TcmN, and 27% identity of pocket residues. The alignments show that there is a significant switch in pocket residue composition between the C9–C14 and the C7–C12 ARO/CYCs (Fig. 4a). Subtle conformational and residue changes in the pocket of WhiE, effectively increase its pocket size compared to TcmN in order to accommodate the binding of a 24-carbon polyketide substrate cyclized at C9–C14 first ring, and C6–C16 second ring.

Structural superposition between TcmN and WhiE reveals high similarity (Fig. 4b). The main difference is the C-terminal helix, which extends for an additional 1.5 helical turns at the C-terminal end in WhiE, while the corresponding region in TcmN is random coil (LEE; AMES; TSAI, 2012). The structural superposition between TcmN and ZhuI shows several significant differences (Fig. 4c). For example, the L9 loop at the entrance to the interior pocket is extended by 12 residues in ZhuI (AMES *et al.*, 2011).

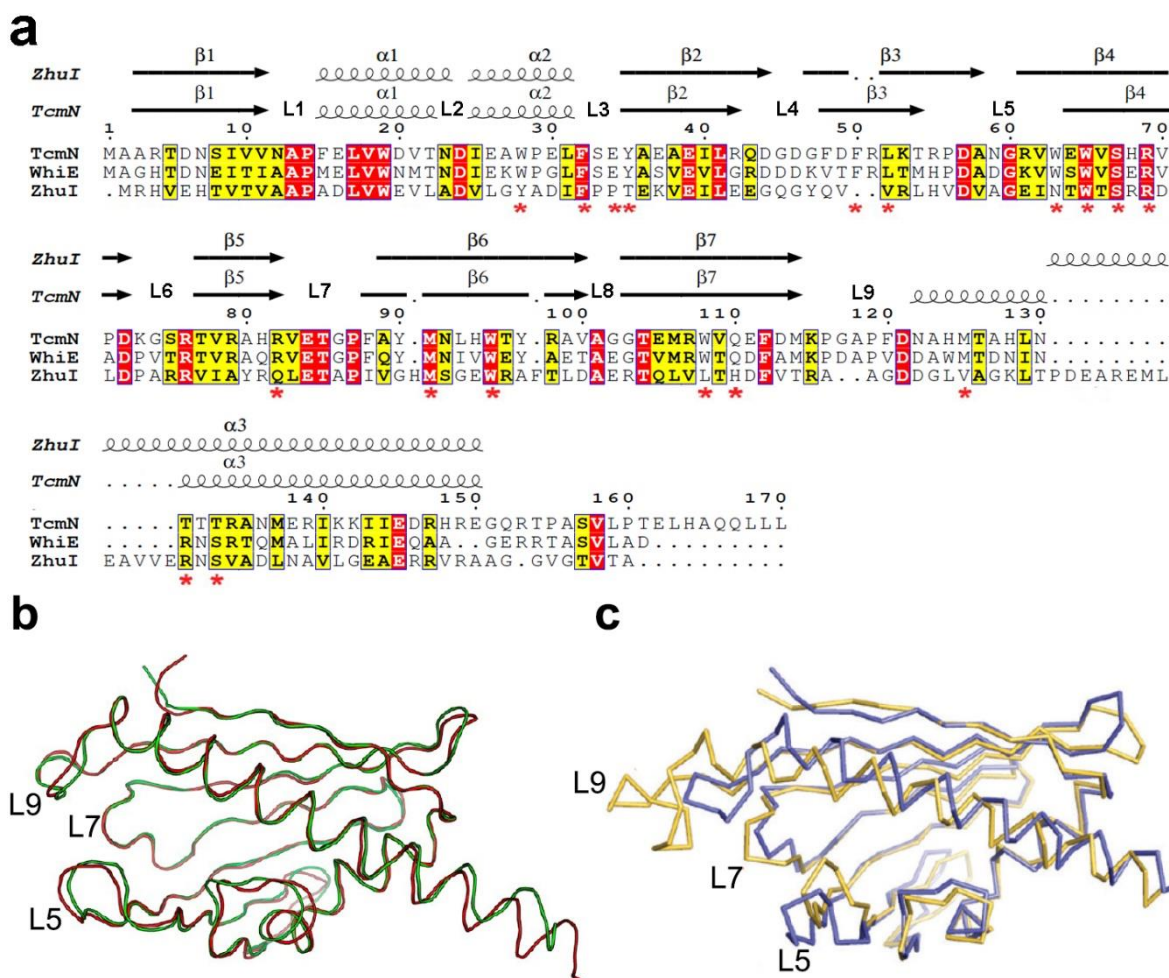


Figure 4 – Comparison of TcmN, WhiE and ZhuI.

(a) Sequence alignment of TcmN, WhiE and ZhuI rendered using ESPrpt (Risler similarity matrix with global score set to 0.7). The top line illustrates secondary structure elements according to the TcmN and ZhuI structure. The asterisks below the alignment indicate the residues that define the TcmN interior pocket. (b) Backbone overlay of TcmN (green) and WhiE (red). (c) Backbone overlay of TcmN (blue) and ZhuI (yellow). Loops segments L5, L7 and L9 of TcmN are identified in (b) and (c). Adapted from Ames *et al*, *Biochemistry* 2011, 50, 39, 8392-8406 © 2011 American Chemical Society.

Heterologous expression of TcmN-lacking minimal PKS results in the production of two differently cyclized products, SEK15 and SEK15b (MCDANIEL, R., HUTCHINSON, C. R., KHOSLA, 1995). The inclusion of TcmN, in the absence of downstream enzymes, produces RM80/RM80b, which have the natural cyclization and aromatization pattern for the first two rings (ZAWADA; KHOSLA, 1999) (Fig. 5). Using different combinations of PKS subunits, it has been shown that WhiE can functionally replace TcmN, because expression of the Tcm minimal PKS with WhiE or TcmN yields the same 20-carbon, C9–C14 first-ring (and C7–C16 second-ring) cyclized polyketides, RM80 and RM80b (LEE; AMES; TSAI, 2012).

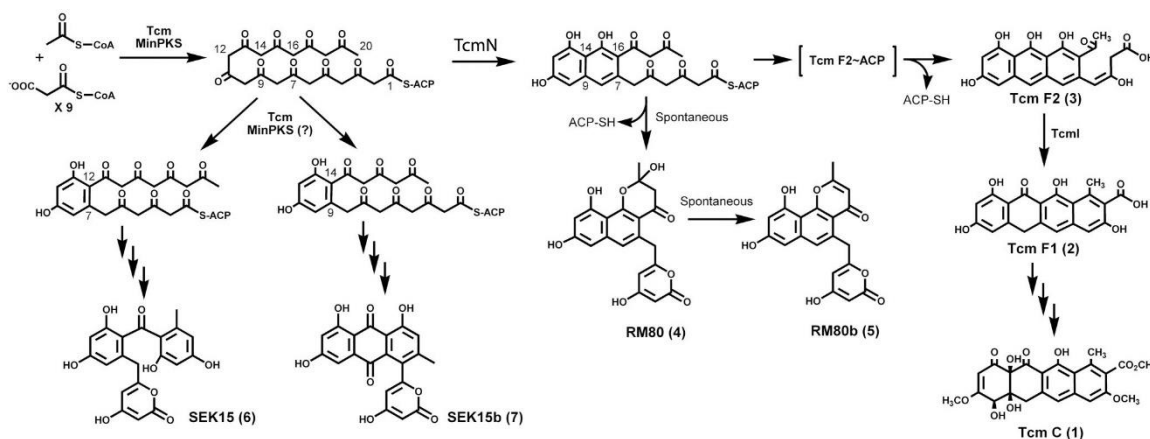


Figure 5 – Proposed biosynthetic pathway for Tcm C (1), RM80 (4), RM80b (5), SEK15 (6), and SEK15b (7).

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.

The tertiary structure of TcmN is similar to members of the Bet v1-like superfamily. The ubiquitous distribution of Bet v1-related proteins among all super kingdoms suggests that a Bet v1-like protein was already present in the last universal common ancestor. During evolution, the ancestral protein diversified into numerous families with varying degrees of sequence similarity, while maintaining the ancestral fold. The internal cavity has been proposed as a versatile scaffold for binding small molecules, such as phytosteroid hormones, lipids, polyketides, and cholesterol (IYER; KOONIN; ARAVIND, 2001). Bet v1-like proteins must undergo conformational changes for substrate binding, catalysis, and product release. Involvement of different conformational states during these processes has been suggested, but they have not been thoroughly characterized (MORAES *et al.*, 2016, 2018). An important goal of polyketide biosynthetic engineering is to generate large collections of candidate drug molecules. Understanding the dynamics and mechanism that ARO/CYCs use to catalyze cyclization and aromatization is critical for developing protein engineering strategies.

1.3 Protein aggregation

Using natural enzymes in cell-free conditions can impose some challenges. For instance, protective mechanisms present in the natural organism, such as chaperones, often need to be accounted for (MUNTAU *et al.*, 2014). For biotechnological applications, enzyme stability is of critical importance. The equilibrium between the unfolded state, folding intermediates, and the folded conformation must be evaluated to insure optimal enzyme functionality (Fig. 6). It

is also critical to avoid conditions that promote protein aggregation. Proteins typically present a complex free-energy landscape (ONUICHIC; LUTHEY-SCHULTEN; WOLYNES, 1997; PLOTKIN; ONUICHIC, 2002a, b). The roughness of the landscape affects the ease at which a polypeptide chain folds, and determines the behavior of the folded structure, thus influencing its functionality and stability (TAVERNELLI; COTESTA; DI IORIO, 2003). Alternative conformations may enable functional interactions by exposing particular surfaces; however, there is also a chance that the exposed surfaces are aggregation prone, thus creating a risk of misfolding, aggregation and/or polymerization (GERSHENSON *et al.*, 2014).

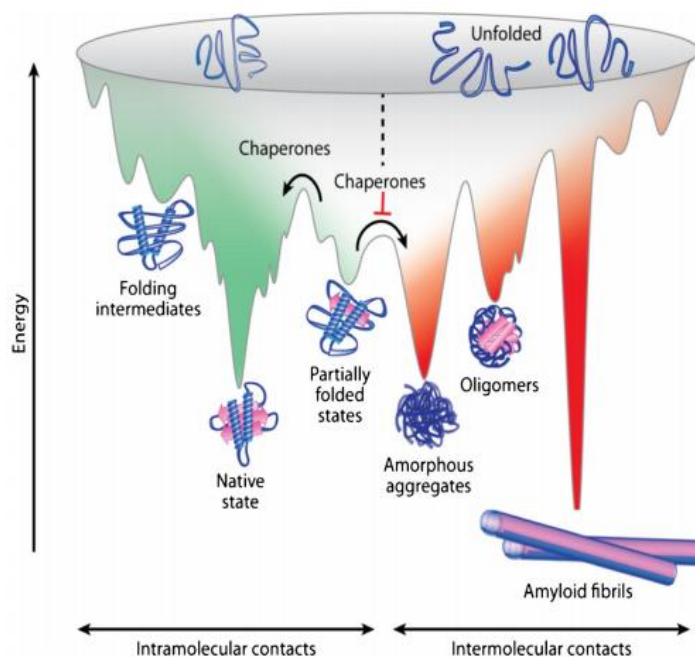


Figure 6 – Energy landscape of protein folding and misfolding.

From the unfolded state toward the natively folded state, protein folding, and protein aggregation are competing reactions. Some intramolecular interactions (green) are associated with an increase in conformational stability in the course of downhill protein folding toward the native state. During this process, proteins may adopt energetically favorable but nonnative conformations that lead to populations of kinetically trapped states. These are located in low-energy wells (partially folded or misfolded states). Moreover, the kinetically trapped states are prone to establish intermolecular interactions (red), which result in protein aggregation (amorphous aggregates, β -sheet-rich oligomers, and amyloid fibrils). Usage of this figure has been granted by the John Wiley and Sons (license number 4750401393307)

Because aggregation is often mediated by association of exposed hydrophobic sequences, local fluctuations of the native state may give rise to near-native, aggregation-prone states, which in turn may lead to partial solvent exposure of hydrophobic stretches (GERSHENSON *et al.*, 2014; KNOWLES; VENDRUSCOLO; DOBSON, 2014). In this work,

the energy landscape of TcmN, and its impact on enzyme function and stability was evaluated using experimental and theoretical techniques, as detailed in the following sections.

1.4 Techniques for characterizing protein structure and dynamics

CD spectroscopy

Circular Dichroism (CD) is a physical phenomenon characterized by the differential absorption of left-handed (A_L) and right-handed (A_R) circularly polarized light by a specific sample. When the light passes through an optically active sample, there will be an absorption of the light if electrons are transiently promoted from a ground to a higher energy state. CD instruments measure the absorbance difference of the circular components of linearly polarized light ($\Delta A = A_L - A_R$) as a function of wavelength (λ). The difference is usually described in terms of ellipticity (θ), which represents the tangent obtained from the ratio of the smallest to the largest axis of the ellipse (KELLY, S. M.; JESS; PRICE, 2005) (Fig. 7)

$$\tan \theta = \frac{A_R - A_L}{A_R + A_L} \quad [1]$$

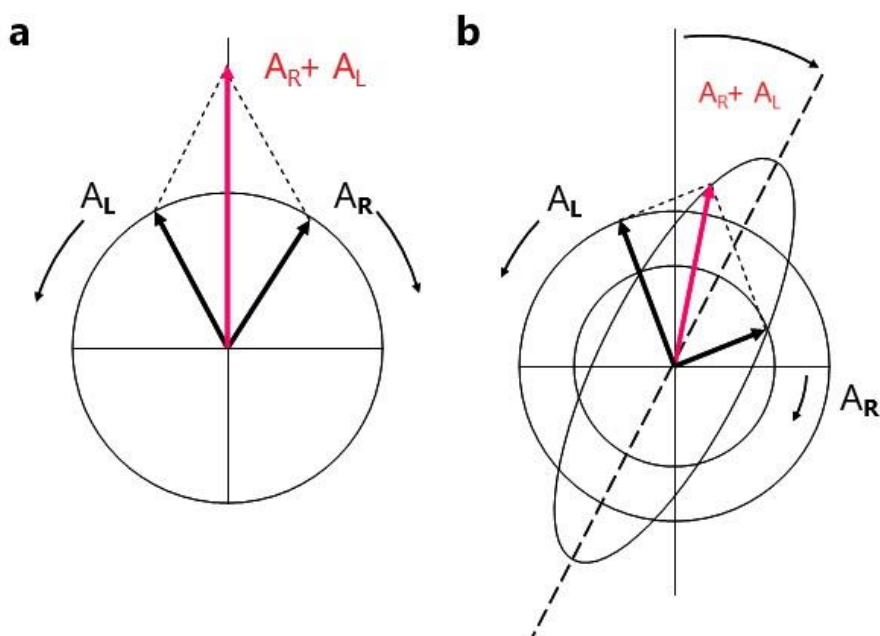


Figure 7 – CD effect on the circularly polarized components A_L and A_R .

(a) Left- and right-hand polarized components of equal amplitude and phase which, when combined, generate polarized plane radiation (red). (b) Different absorption of the left- and right-hand polarized component leads to ellipticity. Adapted from Kelly, S. M. *et al*, 2005.

The ellipticity may take positive or negative values depending on which light component is preferably absorbed. Far-UV CD corresponds to the spectral region over the wavelength range 160 to 240 nm. The peptide bond has two transitions in this part of the spectrum: an $n \rightarrow \pi^*$ transition at 220 nm, and a $\pi \rightarrow \pi^*$ transition at 190 nm. These transitions result in distinct spectra for different secondary structural elements. The most apparent spectral signatures for α -helices are negative bands at 222 nm and 208 nm. β -sheets give rise to a positive peak around 195 nm, and a negative peak around 215 nm. A protein with no dominant secondary structure has a negative peak around 200 nm. Fig. 8 displays illustrative CD spectra for α -helical, a β -sheet, and disordered proteins. CD can also be used to study protein stability.

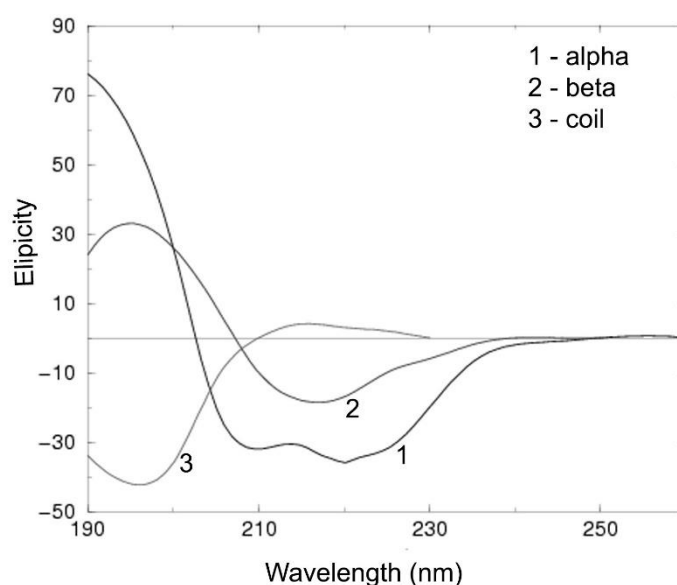


Figure 8 – Typical far-UV spectra for α -helices, β -sheets, and random coil secondary structures.

Adapted from Kelly, S. M. *et al*, 2005

Fluorescence spectroscopy

Fluorescence spectroscopy involves using a beam of light to excite the electrons of certain compounds, causing them to emit light when they return to a lower energy level. A molecule is first excited from its ground electronic state, S_0 , to an excited electronic state, S_1 (Fig. 9). The molecule then drops back to the ground state, emitting a photon in the process. As molecules may drop down into any of several vibrational levels in the ground state, the emitted photons will have different energies, and thus frequencies. By analyzing the different frequencies of light emitted in fluorescent spectroscopy, along with their relative intensities, the structure of the different vibrational levels can be determined. The technique can be used to

verify the interaction of a protein with the ligand, or to verify the loss or gain in structure from the change in fluorescence emission intensity.

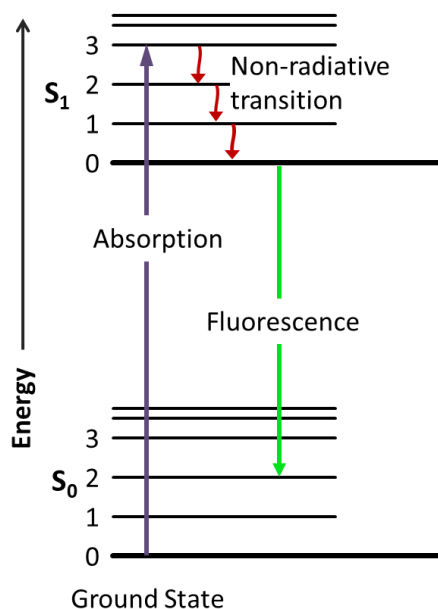


Figure 9 – Perri-Jablonski diagram including vibrational levels for absorbance, non-radiative transition, and fluorescence.

Adapted from Lichtman and Conchello, 2005.

NMR spectroscopy

NMR is a powerful tool to determine three-dimensional protein structure, study protein dynamics, investigate, and map protein interactions. NMR active nuclei atoms have odd number of protons, neutrons, or both, with spins values of $\frac{1}{2}$, such as ^1H , ^{15}N , ^{13}C , can assume two different orientations when placed in a uniform magnetic field (B_0): spin-aligned or spin-opposed orientations parallel to the field. Nuclei in the spin aligned state have a slightly lower energy and are slightly more populated than the spin opposed nuclei. The energy difference (ΔE) between the two orientations is described by Equation 2, and the population difference follows a Boltzmann distribution (WÜTHRICH, 1986).

$$\Delta E = \frac{\gamma h B_0}{2\pi} \quad [2]$$

where h is the Planck constant, and γ is the gyromagnetic ratio for a particular nucleus.

By applying electromagnetic radiation, in the form of pulses of radio waves, NMR active nuclei can be temporarily excited to the higher energy, unaligned state. The resonance signal is detected as the spin returns to the lower energy state. This signal is processed by Fourier transform, resulting in a frequency domain spectrum. Various types and patterns of electromagnetic pulses are combined to generate specific NMR experiments to obtain the desired information (e.g. chemical environment, dynamics, etc) about the nuclei.

To compare NMR data collected at different field strengths, the signal frequencies are converted to unitless values (ppm) called chemical shift (δ):

$$\delta \text{ (ppm)} = \frac{\nu - \nu_{ref}}{\nu_{ref}} * 10^{-6} \quad [3]$$

where ν is the signal position to be measured and ν_{ref} is the signal position relative to a reference compound, typically DSS or TMS.

Proteins have varying degrees of flexibility. Dynamics can play an important role in different processes, such as enzymatic catalysis, which often involves protein conformational changes. NMR is suitable for the study of these dynamics processes, as motions occurring at various time scales (i.e. ps to s) can be monitored (Fig. 10).

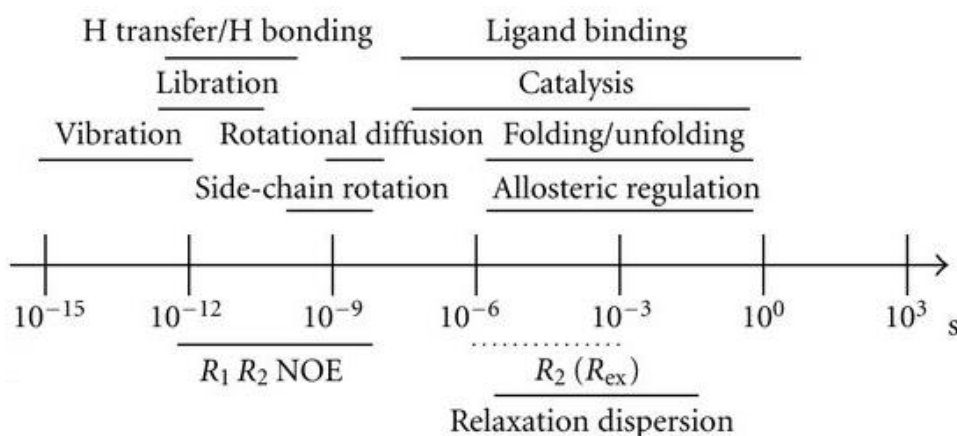


Figure 10 – Protein motion and NMR timescales.

Adapted from Henzler-Wildman, K. *et al*, 2007

MD simulations

MD is a computational approach of simulating the movements and interactions of atoms as a function of time, providing a unique glimpse into protein behavior. The technique was

developed by Adler and Wainwright in the late 1950s, and is known as a classical method because it is based on Newton's second law for a system of interacting particles:

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d\vec{v}_i}{dt} = m_i \frac{d^2\vec{r}_i}{dt^2} \quad [4]$$

where F_i is the force exerted on particle i , m_i is the mass of particle i , a_i is the acceleration of particle i , v_i is the velocity of particle i and r_i is the position of particle i at time t .

The initial position may come from a crystal or NMR protein structure, and the velocities are often taken from a Maxwellian distribution at the desired temperature. A trajectory is calculated by allowing the particles to move for a short period of time, and after obtaining the resulting forces, a new set of coordinates for the particles will be produced at a time interval $t+1$, and this process is repeated as long as necessary. Force calculation is the most computationally expensive step in MD, and is determined by the force-field, which contains specific energy terms for bonded (i.e. bond lengths, angles, and dihedrals) and nonbonded (i.e. Van der Waals and electrostatic) interactions (PATODIA; BAGARIA; CHOPRA, 2014).

2 OBJECTIVES

This work aimed to characterize fundamental biophysical properties of TcmN, such as stability, conformational dynamics, and aggregation propensity under different conditions. The results are expected to be useful for assessing its suitability in industrial applications for engineering polyketide biosynthesis. Specific objectives of the project included:

1. Characterization of TcmN stability under temperature, pH, and salt concentration variation
2. Backbone resonance assignment of TcmN
3. Characterization of TcmN dynamics by NMR
4. Identification and characterization of different TcmN conformational states by MD

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 the company GenScript (Piscataway, USA) in a pET28a plasmid. *E. coli* BL21 (DE3) cells were transformed with the plasmid and were grown at 37°C with 200 rpm shaking in 50 mL Luria Bertani (LB) medium with 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin overnight. Cells were harvested by centrifugation at 3000 rpm and used to inoculate 500 mL of LB, or ^{15}N and ^{13}C enriched M9 medium to a starting optical density at 600 nm (OD_{600}) of 0.2. Cells were grown up to an OD_{600} of 0.8, and protein overexpression was induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 16 h at 22°C. Cells were harvested by centrifugation for 40 min at 4°C and at 4500 rpm in a Thermo Scientific TX-750 Swinging Bucket Rotor. Harvested cells were resuspended in 50 mM sodium phosphate pH 8 containing 300 mM NaCl and 10 mM imidazole. The cell suspension was frozen at -20°C, thawed at room temperature with 3 $\text{mg}\cdot\text{mL}^{-1}$ of lysozyme, and passed 10x through a 0.55 mm gauge needle for cell disruption. The lysate was centrifuged for 40 min at 4°C and at 11000 rpm, and the protein was purified from the soluble fraction using HisTrap HP Ni-NTA protein purification column (GE Healthcare, USA) in 50 mM sodium phosphate pH 8 containing 300 mM NaCl and 10 mM imidazole, and eluted with imidazole gradient up to 500 mM. The purified protein was dialyzed against 20 mM sodium phosphate pH 7 containing 100 mM NaCl. Purity was determined by SDS-PAGE, and its concentration estimated by UV-vis spectroscopy at 280 nm using an extinction coefficient of 37470 $\text{M}^{-1}\cdot\text{cm}^{-1}$, obtained from the His6-tagged TcmN amino acid primary sequence using the server protparam (<https://web.expasy.org/protparam/>).

3.2 Fluorescence spectroscopy

The thermal stability of TcmN was characterized by fluorescence spectroscopy. Experiments were carried out on Prometheus NT.48 (Nanotemper, USA), and in Cary Eclipse (Varian, USA) spectrophotometers. Fluorescence emission spectra were acquired over wavelengths ranging from 300 to 450 nm using a 1 cm path length cuvette. The temperature was varied from 25 to 85°C. For each temperature, the sample was equilibrated for three minutes before spectrum acquisition. Fluorescence data were normalized using the following equation:

$$P_F = \frac{I - I_U}{I_F - I_U} \quad [5]$$

where P_F is the folded population of protein, I , I_U , and I_F are the ratio of fluorescence intensities measured at 350 and 330 nm, fluorescence intensity of the unfolded population, and fluorescence intensity of folded population, respectively. T_m^* (observable melting temperature) analysis was performed with web-based software CalFitter (MAZURENKO *et al.*, 2018). The influence of ionic strength and salt type on TcmN thermal stability was evaluated using 6 μ M TcmN in 20 mM sodium phosphate pH 7 and NaCl (50 and 100 mM), and NaI (50 and 100 mM). The effect of TcmN concentration on its stability was tested over a 1-80 μ M range. Amyloidogenic aggregation of TcmN was assessed using 20 μ M of fluorescent dye thioflavin T (ThT). Measurements were performed at 40°C under agitation on a Cytation™ 5 spectrofluorometer (Biotek, USA). ThT binding to TcmN and p53 DBD was monitored by exciting samples at 440 nm and recording the emission spectra at 485 nm.

3.3 CD spectroscopy

Far-UV CD measurements were performed on a Jasco spectropolarimeter (J-815) equipped with Peltier temperature controller. 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, taking the average of three scans. Thermal denaturation assays were performed by measuring the CD intensity at 208 nm from 25 to 85°C, at 0.5°C increments, with measurements at each temperature following a 5 seconds equilibration period. The population of the folded state was estimated using the following equation:

$$P_F = \frac{I_{CD} - I_{CD,U}}{I_{CD,F} + I_{CD,U}} \quad [6]$$

where I_{CD} , $I_{CD,U}$, and $I_{CD,F}$ correspond to the molar residue ellipticity (MRE) values at 208 nm at the measurement temperature, 25°C (folded state), and 90°C (unfolded state), respectively. T_m^* analysis was performed with web-based software CalFitter (MAZURENKO *et al.*, 2018).

3.4 NMR spectroscopy

For backbone assignment experiments, 200 μ M of ¹⁵N and ¹³C labelled TcmN was prepared in 20 mM sodium phosphate pH 7, 100 mM NaCl, and 10% D₂O. NMR experiments were performed at 308 K using Bruker 900 MHz Avance Neo and 800 MHz Avance III

spectrometers equipped with TXI 5 mm triple-resonance probes, both located at the National Center of Nuclear Magnetic Resonance – CNRMN of the Federal University of Rio de Janeiro.

The following NMR triple-resonance experiments were acquired and analyzed: ^1H - ^{15}N HSQC, ^1H - ^{13}C HSQC, HNC0, HN(CA)CO, HNCA, HNCACB and CBCA(CO)NH. 2D experiments were acquired using traditional full acquisition schedule, and 3D experiments were collected using a non-uniform Poisson Gap Sampling schedule (HYBERTS *et al.*, 2012; HYBERTS; TAKEUCHI; WAGNER, 2010). The spectra were processed using Topspin 3.7 (Bruker Corporation, USA) and NMRpipe (DELAGLIO *et al.*, 1995), and analyzed using CCPN (VRANKEN *et al.*, 2005).

The backbone dynamics of TcmN were characterized by monitoring relaxation rates R_1 , R_2 ^1H - ^{15}N HetNOE obtained from ^{15}N relaxation experiments acquired 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. ^1H - ^{15}N HetNOE, R_1 and R_2 parameters were measured using pulse sequences provided by the manufacturer: HSQCNOEF3GPSI (NOE) (parameters: number of scans (ns) 128; recovery time (d1): 4 s; transient points in the indirect dimension (td f1): 120), HSQCT1ETF3GPSI3D (R_1) (ns: 24; d1: 2.2 s; td f1: 100), and HSQCT2ETF3GPSI3D (R_2) (ns: 32; d1: 1.2 s; td f1: 128). The experiments were carried out at 25°C with 180 μM TcmN in 20 mM sodium phosphate pH 7, and 100 mM NaCl. R_1 and R_2 rates were obtained from fitting the intensity exponential decay of resonance intensities as a function of the relaxation delays: 0.05, 0.1, 0.2, 0.3, 0.5, 0.6, 0.8, 0.9 and 1.0 s for R_1 and 16.96, 33.92, 50.88, 67.84, 84.8, 101.76, 118.72, 135.68, 203.52 and 237.44 ms for R_2 , respectively (FARROW *et al.*, 1994; KAY; TORCHIA; BAX, 1989). ^1H - ^{15}N HetNOE rates were acquired from ^1H - ^{15}N correlation experiments with and without saturation of ^{15}N magnetization, using recovery time of 5.0 s (KAY; TORCHIA; BAX, 1989). The relaxation rates mentioned above were adjusted to the Lipari-Szabo theoretical model (LIPARI; SZABO, 1982) using the software TENSOR2 (DOSSET *et al.*, 2000). The overall protein correlation time, τ_c , was calculated from the R_2/R_1 ratios of the residues in which the values of the relaxation rates did not differ more or less than one standard deviation from the mean. The residues 4, 5, 6, 7, 8, 10, 12, 13, 15, 16, 17, 18, 19, 20, 21, 24, 25, 26, 27, 32, 35, 37, 39, 40, 42, 43, 45, 46, 50, 72, 73, 74, 75, 76, 77, 78, 79, 86, 90, 95, 96, 97, 98, 99, 102, 103, 104, 105, 106, 109, 110, 111, 113, 117, 118, 120, 123, 126, 128, 129, 134, 136, 138, 139, 144, 148, 149, 154 were used to obtain the parameters calculated by

TENSOR2 program. The internal dynamics parameters: generalized order parameter (S^2) and the chemical exchange relaxation rate between μs -ms (R_{ex} (s^{-1})) were calculated using the axial symmetry diffusion model.

Residue-specific dynamics of TcmN at micro-millisecond time scale was monitored by relaxation-compensated ^{15}N single-quantum Carr-Purcell-Meiboom-Gill (CPMG), RC- ^{15}N CPMG relaxation dispersion experiments (LORIA; RANCE; PALMER, 1999). RC- ^{15}N CPMG experiments were acquired at 308 K and using Bruker 800 MHz Avance III spectrometers equipped with TXI 5 mm triple-resonance probe and Bruker Avance Neo 600 MHz spectrometer equipped with PABBO 5 mm Smart Probe using the pulse sequence HSQCREXETF3GPSITC3D provided by the manufacturer. Experimental R_2 effective rates, $R_{2,\text{eff}}$ (Hz), were obtained from ^1H - ^{15}N correlation spectra intensities using the CCPN (VRANKEN *et al.*, 2005) and applying the following equation:

$$R_{2,\text{eff}} = \frac{1}{T_{\text{relax}}} \ln \left(\frac{I_0}{I_{\text{CPMG}}} \right) \quad [7]$$

where T_{relax} is the relaxation delay of 30 ms, I_{CPMG} is the intensity measured in ^1H - ^{15}N spectra acquired with CPMG frequency, ν_{CPMG} , ranging from 66.7 to 1000 Hz. I_0 is the peak intensity obtained with T_{relax} of 0 ms. The relaxation dispersion CPMG experiments were acquired with the following ν_{CPMG} frequencies: 66.7, 133.3, 200, 266.7, 333.3, 400, 466.7, 533.3, 600, 666.7, 866.7 and 1000 Hz. Experimental errors were calculated from the standard deviation of noise in relation to the resonance intensities.

3.5 MD simulations

Six replicates of MD simulations were performed using GROMACS 5, utilizing the CHARMM22* force field and standard TIP3P water model.

The TcmN crystal structure (pdb accession code 2RER) (AMES *et al.*, 2008) was used with protonation states of titratable residues predicted by PropKa (OLSSON *et al.*, 2011) at pH 7. The CHARMM22* force field (PIANA; LINDORFF-LARSEN; SHAW, 2011) and standard TIP3P water model were used. Na^+ and Cl^- were added to simulate a physiological concentration of 0.15 M. MD was performed using GROMACS 5 (ABRAHAM *et al.*, 2015). After steepest-descents minimization, six 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, in house scripts, and other tools were employed to analyze and visualize 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 C α atoms of the initial structure (CINO; CHOY; KARTTUNEN, 2013). Bidimensional histograms were prepared by binning the trajectory projections on the first two principal components and expressed in energy units using Boltzmann's factor. Cavity analysis was performed with CASTp 3.0 (TIAN *et al.*, 2018). ^1H - ^{15}N NMR spin relaxation values were calculated for the six replicas individually using the experimentally determined global tumbling time of 9.3 ns, and standard methods (CINO; KARTTUNEN; CHOY, 2012), as implemented in SpinRelax (CHEN, P. *et al.*, 2018).

4 RESULTS AND DISCUSSION

4.1 Expression and purification

Several expression tests were performed with different *E. coli* strains: BL21(DE3), BL21(DE3) pLysS, and BL21(DE3) RIPL. In addition, the expressions tests included variation of IPTG concentration, induction growth temperature, and induction time. The best expression condition found was in BL21(DE3) induced with 0.5 mM IPTG for 16 h at 22°C (Fig. 11) in both LB medium or ¹⁵N and ¹³C enriched M9 medium.

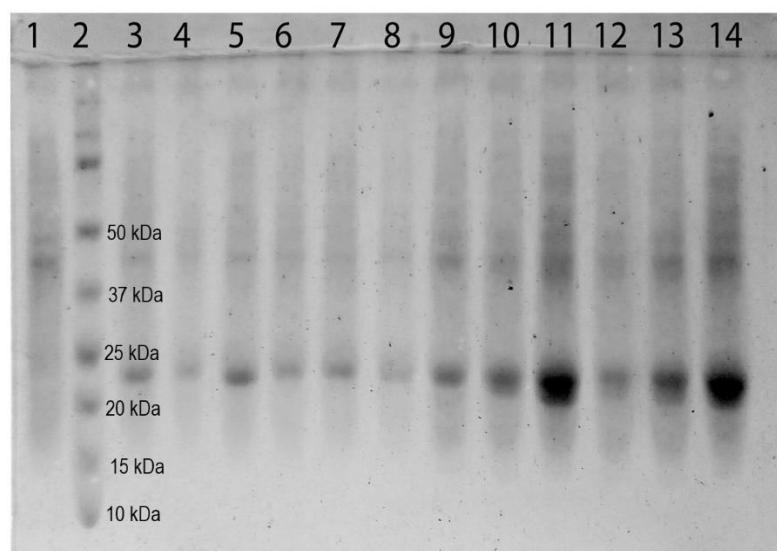


Figure 11 – TcmN expression test.

12% SDS-PAGE of TcmN expression tests in LB medium. Lane 1: total protein extract before induction. Lane 2: molecular weight standard. Lanes 3, 4, and 5: total protein extract induced with 0.5 mM IPTG at 37°C for 4, 6, and 12 h, respectively. Lanes 6, 7, and 8: total protein extract induced with 1 mM IPTG at 37°C for 4, 6, and 12 h, respectively. Lanes 9, 10, and 11: total protein extract induced with 0.5 mM IPTG at 22°C for 4, 6, and 12 h, respectively. Lanes 12, 13, and 14: total protein extract induced with 1 mM IPTG at 22°C for 4, 6, and 12 h, respectively. His-tag fused TcmN has 182 amino acids, and molecular weight of 21.06 kDa.

TcmN was purified from the soluble fraction by Ni²⁺ affinity chromatography (Fig. 12), as described in the methods section. Typical final yield per 500 mL culture was 13 mg.

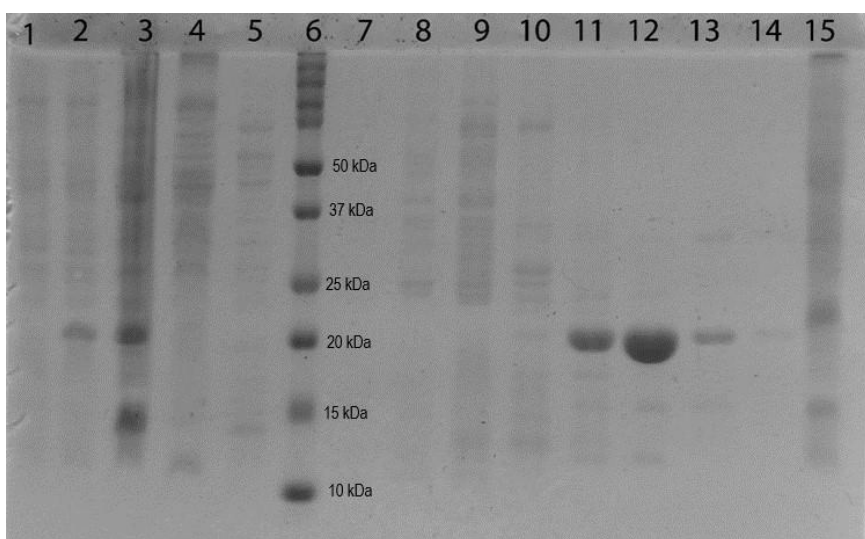


Figure 12 – 12% SDS-PAGE of TcmN Ni²⁺ affinity purification.

Lane 1: total pre-induction protein extract. Lane 2: total protein extract after induction with 0.5 mM IPTG at 22°C for 12 h. Lane 3: lysate soluble fraction. Lane 4: flow through. Lane 5: wash unbound fraction. Lane 6: molecular weight standard. Lanes 7 to 14: elution of TcmN by imidazole gradient. Lane 15: lysate insoluble fraction.

4.2 TcmN backbone resonance assignment

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

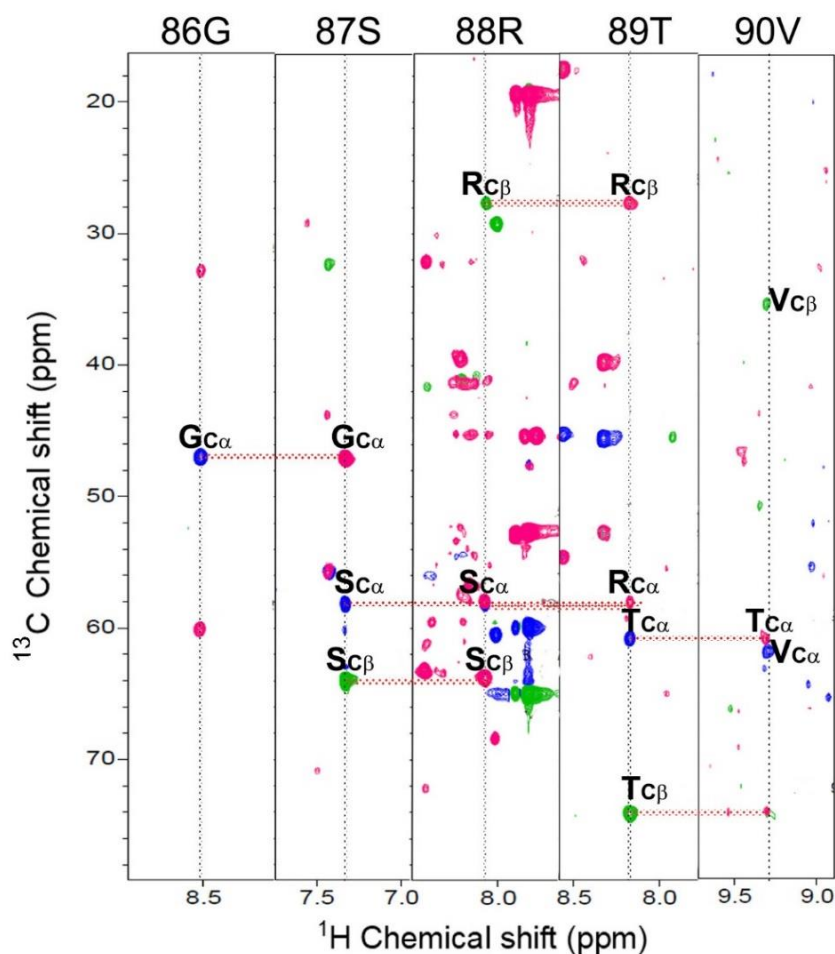


Figure 13 – Overlapping of the CBCACONH spectra (pink) with the HNCACB (blue and green).

At HNCACB, the blue (positive) peaks are $C\alpha$ and green (negative) peaks are $C\beta$. The chemical shifts of 1H and ^{13}C plane 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 the its peaks on the CBCACONH.

Fig. 14a shows the assigned 1H - ^{15}N -HSQC TcmN spectrum. The spectrum has a good chemical shift dispersion, showing that the protein is well-folded. In this spectrum 238 peaks were expected, including the 66 HN side chain peaks. However, due to the overlapping peaks, only 205 peaks were observed. 80% of the backbone atoms has been assigned and the region missing assignment are parts of β -sheet 4 (residues V66, S67, H68, R69, V70), β -sheet 5 (residues H81, R82, V83, E84 and T85), β -sheet 6 (residues A89, M91, N92, L93, H94) near loop L7, and part of the L9 (residues D113, M124, K115 and P116) (Fig. 4c). 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 μ s-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 ^1H - ^{15}N -HSQC spectrum, they could have rather low intensity in the HNCA and other triple-resonance spectra used to assign the backbone.

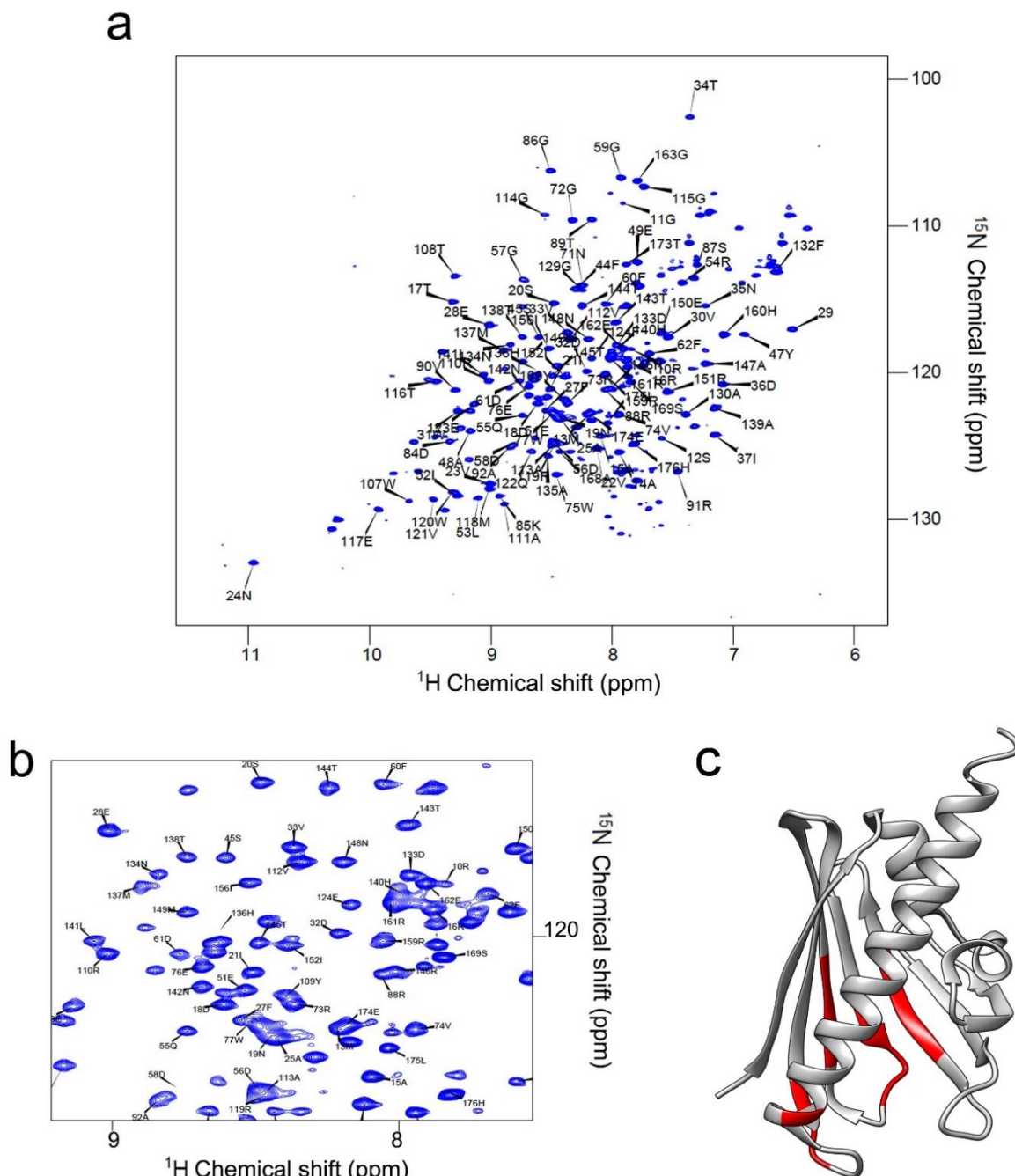


Figure 14 – TcmN NMR assignment.

(a) ^1H - ^{15}N HSQC spectrum showing the assignments and (b) zoom region. In this experiment, each peak contains information on the frequency of ^1HN and ^{15}N of a specific residue, except for prolines. (c) Ribbon representation of the TcmN with unassigned residues shown in red.

4.3 Characterization of TcmN stability

To evaluate TcmN folding, and to compare protein stability under different conditions, CD spectra were collected in the far-UV wavelength range. The CD spectra at 20 and 40°C indicate a folded conformation, which was irreversibly lost between 40 and 55°C (Fig. 15a). TcmN forms insoluble aggregates that can be seen after thermal denaturation (Fig. 15d). TcmN structural stability was also pH dependent. TcmN retained its secondary structure at pH 8.1, 6.8, and 5.6, but exhibited detectable loss of signal intensity at pH 4.0, indicative of partial structure loss (Fig. 15b). TcmN thermal plot shows that TcmN is less stable at pH 5.6 with a T_m^* (observable melting temperature) of 47°C. T_m^* values at pH 6.8 and pH 8.1 were 56°C and 60°C, respectively (Fig. 15c). Thermal unfolding is generally used to determine thermodynamic protein stability, which is the Gibbs free energy difference between the folded and the unfolded states, $\Delta G = G_{\text{Folded}} - G_{\text{Unfolded}}$ (BENJWAL *et al.*, 2006). It is important to note that when the unfolding process is irreversible and also is dependent on protein concentration, the melting point cannot be recognized by equilibrium thermodynamics, since the kinetics of self-interaction also affects the unfolding process (BENJWAL *et al.*, 2006; DUY; FITTER, 2005; GOYAL; CHAUDHURI; KUWAJIMA, 2015). The T_m^* indicates the midpoint of the unfolding/aggregation curves.

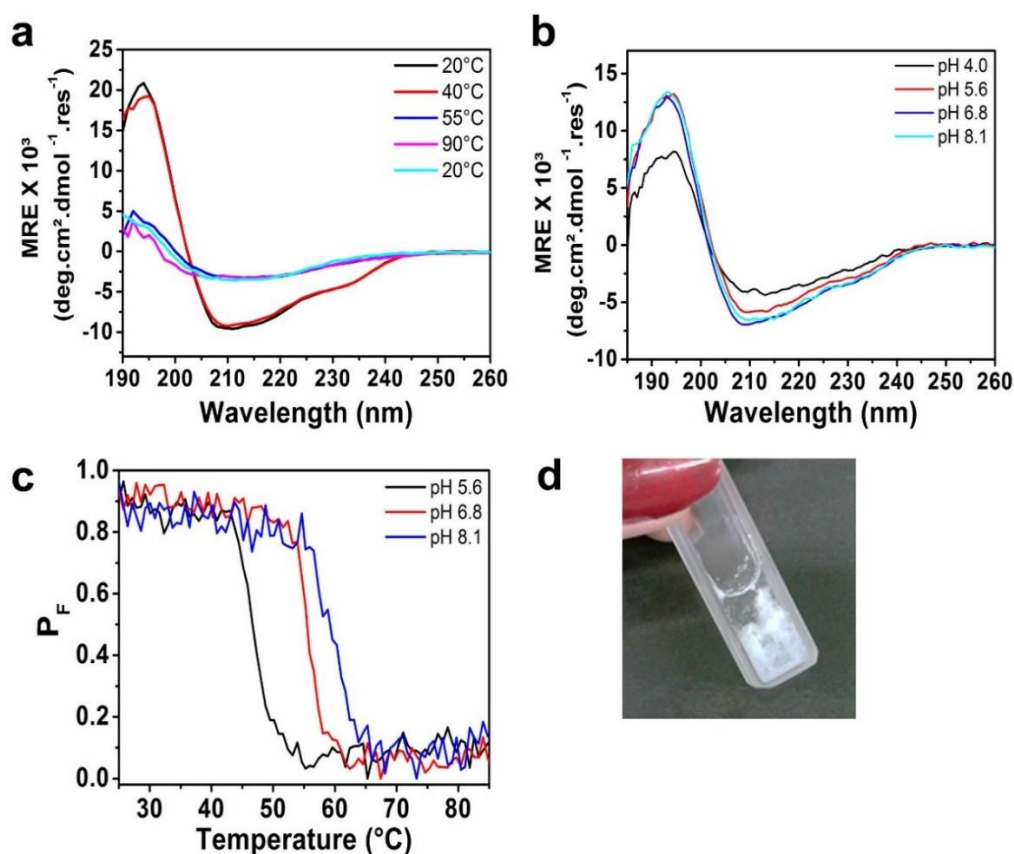


Figure 15 – Stability of TcmN against temperature and pH change.

(a) Overlay of far-UV CD spectra of 95 μM TcmN in 20 mM sodium phosphate pH 7, 100 mM NaCl at different temperatures. (b) Overlay of far-UV CD spectra of 50 μM and 66 μM TcmN in 20 mM sodium acetate pH 4.0 and 5.5, respectively, containing 100 mM NaCl, and of 88 μM and 57 μM TcmN in 20 mM sodium phosphate pH 6.8 and 8.0, respectively, containing 100 mM NaCl. (c) Thermal plot of folded population of TcmN at different pH values. (d) Photo of cuvette with TcmN aggregate after thermal denaturation.

Intrinsic tryptophan fluorescence assays were also used to evaluate TcmN stability. TcmN was most stable in the presence of 100 mM NaI, but was also stabilized, to a lesser extent, in the presence of 100 mM NaCl, compared to plain phosphate buffer (Fig. 16). Ames and colleagues suggested that TcmN is monomeric in 10 mM HEPES pH 7, however, addition of 100 mM NaI resulted in the formation of an iodide-bridged homodimer (AMES *et al.*, 2008). Experiments performed in 20 mM sodium phosphate pH 7 with 100 mM NaCl exhibit a tumbling time consistent with that of a ~21 kDa protein (i.e. TcmN in its monomeric form) (Fig. 21a). The finding that NaI stabilizes TcmN more than NaCl is consistent with the Hofmeister rankings (OKUR *et al.*, 2017), and similar effects of dissolved salts on protein unfolding rates have been reported (BROERING; BOMMARIUS, 2005; SCHWIERZ; HORINEK; NETZ, 2010). The sensitivity of NMR probes can be greatly reduced by high salt

concentrations (KELLY, A. E. *et al.*, 2002). Furthermore, dimerization can also reduce spectral quality. To avoid these issues, experiments were conducted in 20 mM sodium phosphate pH 7 with 100 mM NaCl.

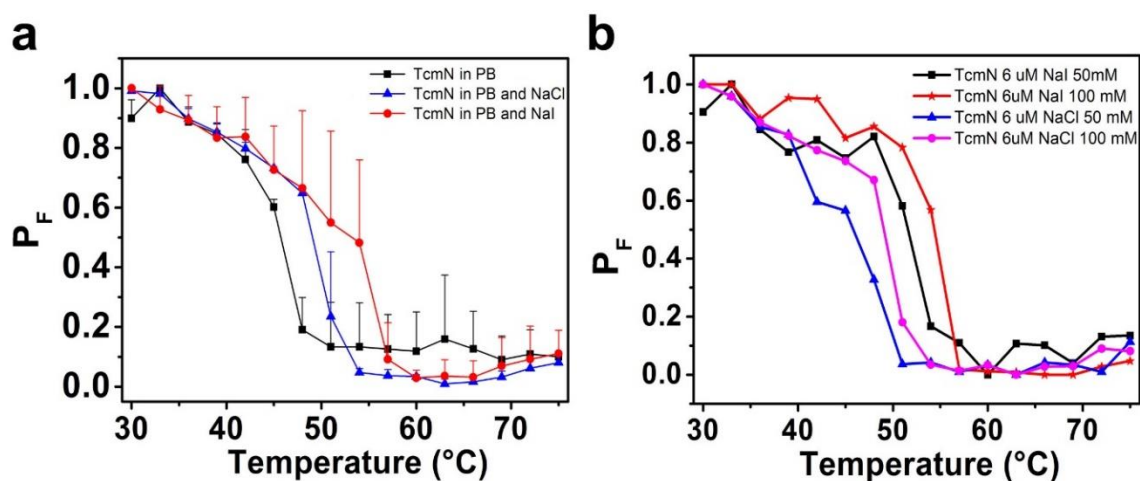


Figure 16 – Thermal denaturation of TcmN under different ions type.

(a) Thermal plot of folded population of 6 μM TcmN in 20 mM sodium phosphate pH 7 and in the presence of 100 mM NaCl and NaI by fluorescence. The error bars are the result of triplicate measurements. (b) Thermal plot of folded population of 6 μM TcmN in 20 mM sodium phosphate pH 7 in the presence of different concentrations of NaCl and NaI by fluorescence.

Protein concentration also affects TcmN thermal stability. Thermo stability of TcmN decays as protein concentration increases (Fig. 17a). This dependence may be explained by transient self-interactions. Higher concentrations increase the probability of interaction, and could facilitate unfolding and aggregation. The T_m^* values of the thermal assays evaluating the thermal denaturation of TcmN under different concentrations are shown in the table below:

CD Thermal assays – T_m^*			
pH 7.0	Near UV (291 nm)	60 μM	54.33°C
		125 μM	52.03°C
pH 7.0	Far UV (208 nm)	60 μM	50.48°C
		125 μM	48.93°C
Fluorescence spectroscopy assays – T_m^*			
pH 7.0	5 μM	51.76°C	
	10 μM	55.62°C	
	16 μM	53.85°C	
	20 μM	52.97°C	
	40 μM	51.63°C	
	60 μM	49.99°C	

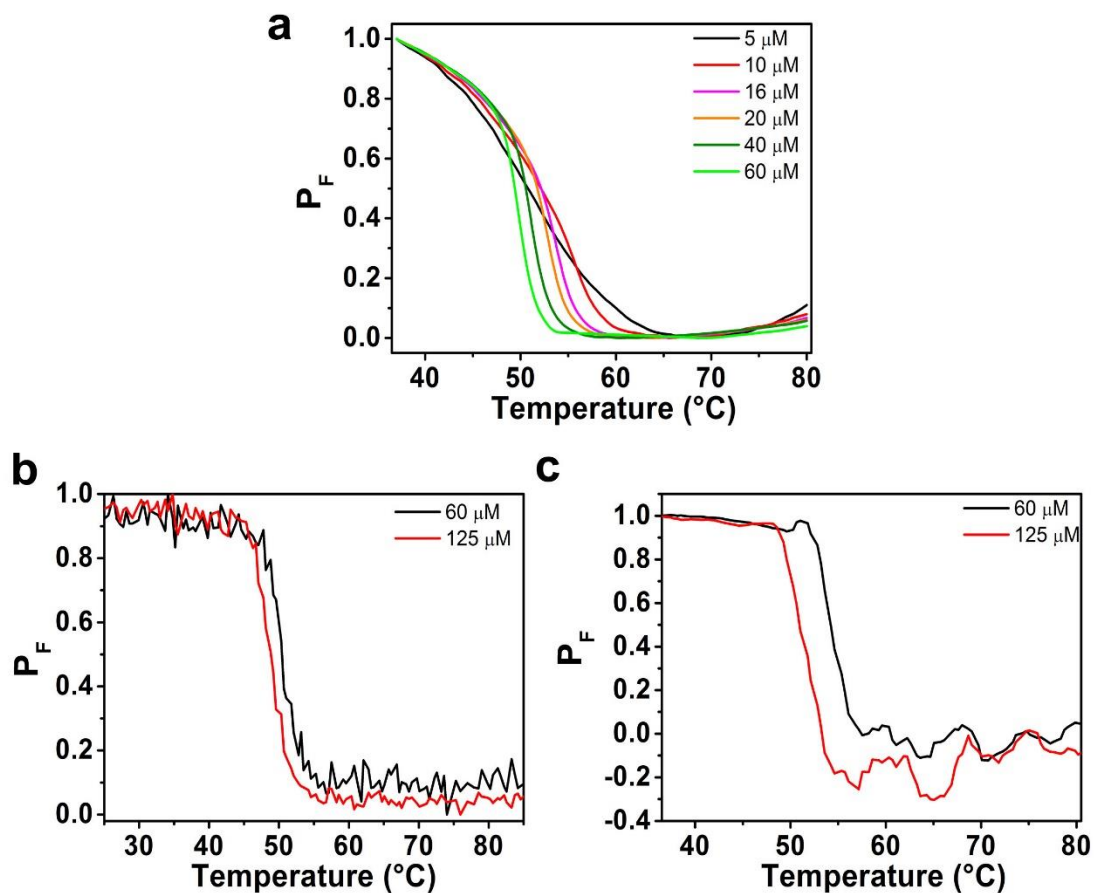


Figure 17 – Thermal denaturation of TcmN under different concentrations.

(a) Thermal plot of folded population of different concentrations of TcmN in 20 mM sodium phosphate pH 7, 100 mM NaCl by fluorescence. (b) Thermal plot of folded population of 60 μM and 125 μM TcmN in 20 mM sodium phosphate pH 7, 100 mM NaCl estimated from the CD intensities (b) at 208 nm and (c) 291 nm.

CD is useful for following protein unfolding transitions because the far-UV signals fade as the secondary structure is disrupted (CLARKE, 2012; PRICE, 2000; SANCHO, 2013). Far and near UV thermal plots show that TcmN T_m^* changes in a concentration dependent manner, and is less stable at 125 μM than at 60 μM (Fig. 17b and c). The T_m^* was lower in the far UV range compared to near UV at the same TcmN concentration, which may indicate that the secondary structure begins to be disturbed first, followed by loss of tertiary structure and aggregation once it opens and exposes tryptophans in the hydrophobic cavity. Similar phenomena of loss of tertiary structure after loss of secondary structure were observed in thermal denaturation studies of other proteins, like Asparaginase-2 and α -Amylases (BENJWAL *et al.*, 2006; DUY; FITTER, 2005).

ThT binding assays were performed to assess if TcmN aggregates possess amyloid structure. For comparison, the p53 DBD was included as a positive control (CINO *et al.*, 2016).

TcmN samples did not yield detectable ThT fluorescence, indicating that it does not form ordered aggregates (Fig. 18).

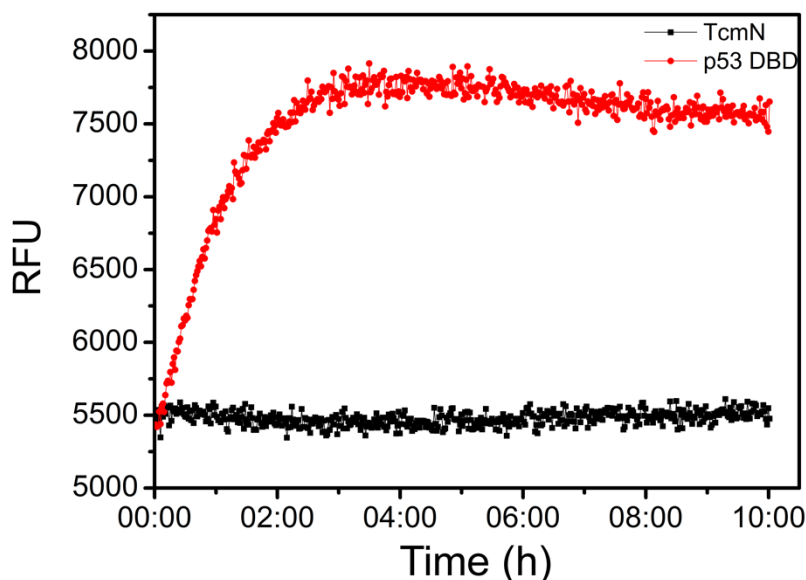


Figure 18 – Aggregation assay of TcmN and of the aggregation-prone protein p53.

The aggregation of 5 μ M TcmN and 5 μ M p53 DBD was monitored by following fluorescent intensity of 20 μ M of fluorescent dye ThT specific for detecting amyloid aggregates.

To further characterize TcmN's concentration dependent aggregation propensity, the possibility of self-interaction was evaluated. ^1H - ^{15}N HSQC spectra of ^{15}N -labelled TcmN titrated with unlabeled TcmN were acquired, and structural changes were identified by mapping the chemical-shift perturbations (CSP). Fig. 19a shows the overlay of free TcmN with that of TcmN in complex with 140 μ M of unlabeled TcmN. Small, but consistent chemical shift variations were seen (Fig. 19b), suggestive of a transient interaction (LIAN, 2013; QIN; GRONENBORN, 2014). The CSP data revealed a broad distribution of residues on TcmN surface; however, residues showing higher CSP values (Fig. 19c) were more concentrated on one face of TcmN (Fig. 19d). One can conjecture that in a condition of high protein concentration, this non-specific self-interaction of TcmN could play an important role in the enzyme aggregation, mainly when these weak interactions occurs between TcmN in an opened conformation, where hydrophobic residues are more exposed.

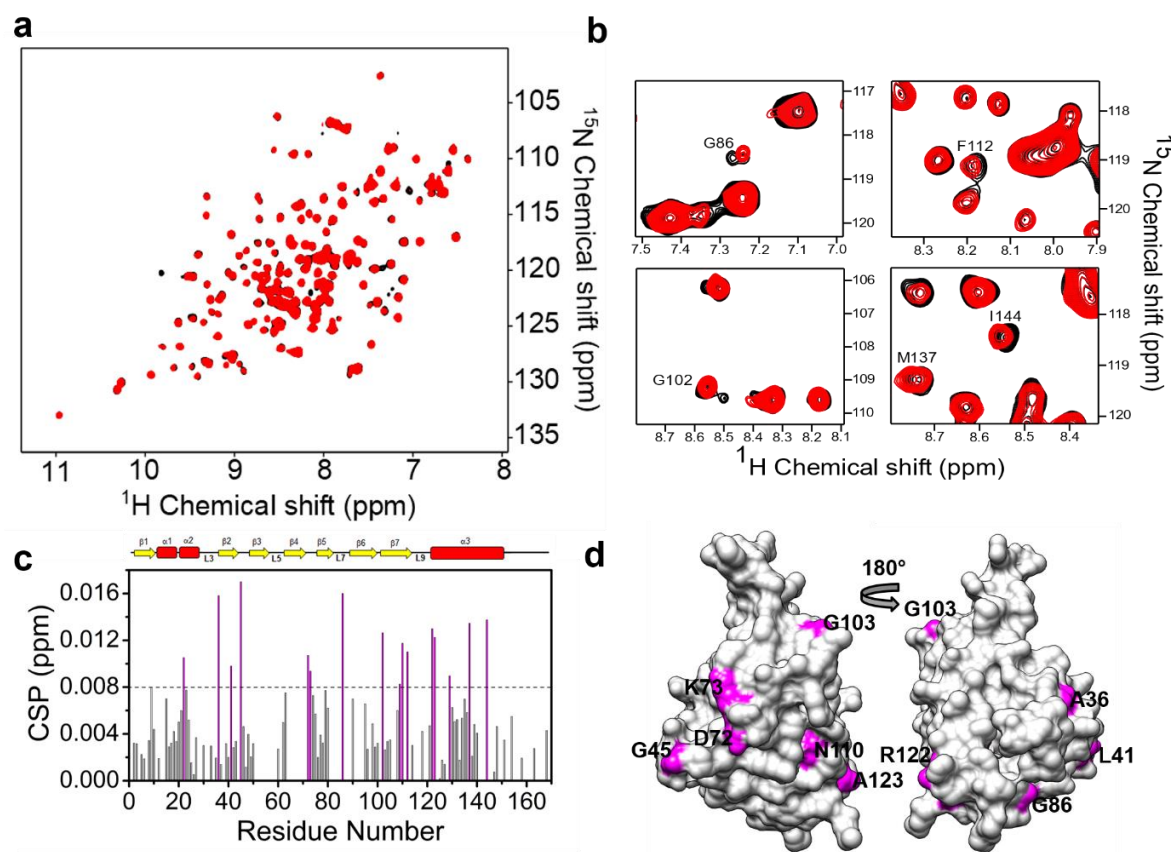


Figure 19 – TcmN transient self-interaction monitored by NMR.

(a) ^1H - ^{15}N HSQC spectra of 70 μM of ^{15}N -labeled TcmN colored in black and 70 μM of ^{15}N -labeled with 140 μM of unlabeled TcmN in red. (b) Selected regions of ^1H - ^{15}N HSQC showing examples of chemical shift variation on ^{15}N -labeled TcmN titrated with unlabeled enzyme. (c) CSP as function of TcmN residue number. (d) Residues showing CSP values more than one standard deviation from the mean.

4.4 TcmN flexibility and dynamics

TcmN dynamics were investigated using NMR relaxation measurements and MD simulations. To analyze the backbone dynamics of TcmN, a complete set of backbone amide ^{15}N relaxation rates (R_1 , R_2 and ^1H - ^{15}N NOE) was measured at 298 K. Fig. 20 shows the results obtained from the NMR ^{15}N relaxation rates, and the theoretical data extracted from the average of the six replicas of MD simulations as a function of residue number. There is a good correlation between the experimental and theoretical data. Overall, the protein is well structured, and most of its residues exhibit similar relaxation parameter values (Fig. 20a and c). Theoretical data shows that exceptions occur in residues from the regions of loops L4, L5 and L9, whose R_2 values are smaller than average and therefore are undergoing fast thermal regime motion (i.e. dynamics on the ps-ns timescale) (Fig. 20b). In contrast, experimental data shows

that residues V11, W19, T22, N23, A36, I40, F48, D49, V62, W65, R76, T77, A80, Y90, H94, W108, Q110, N130, and most of residues from α -helix 3 (120 to 160) had increased R_2 values, indicating conformational exchange processes (Fig. 20b). It is interesting to note that this pattern of dynamics on the ps-ns timescale was different compared to Bet v1 and Fag s1, another protein from the Bet v1-like superfamily, which were less flexible, even for residues located in unstructured regions, such as loops (ASAM *et al.*, 2014; MORAES *et al.*, 2018).

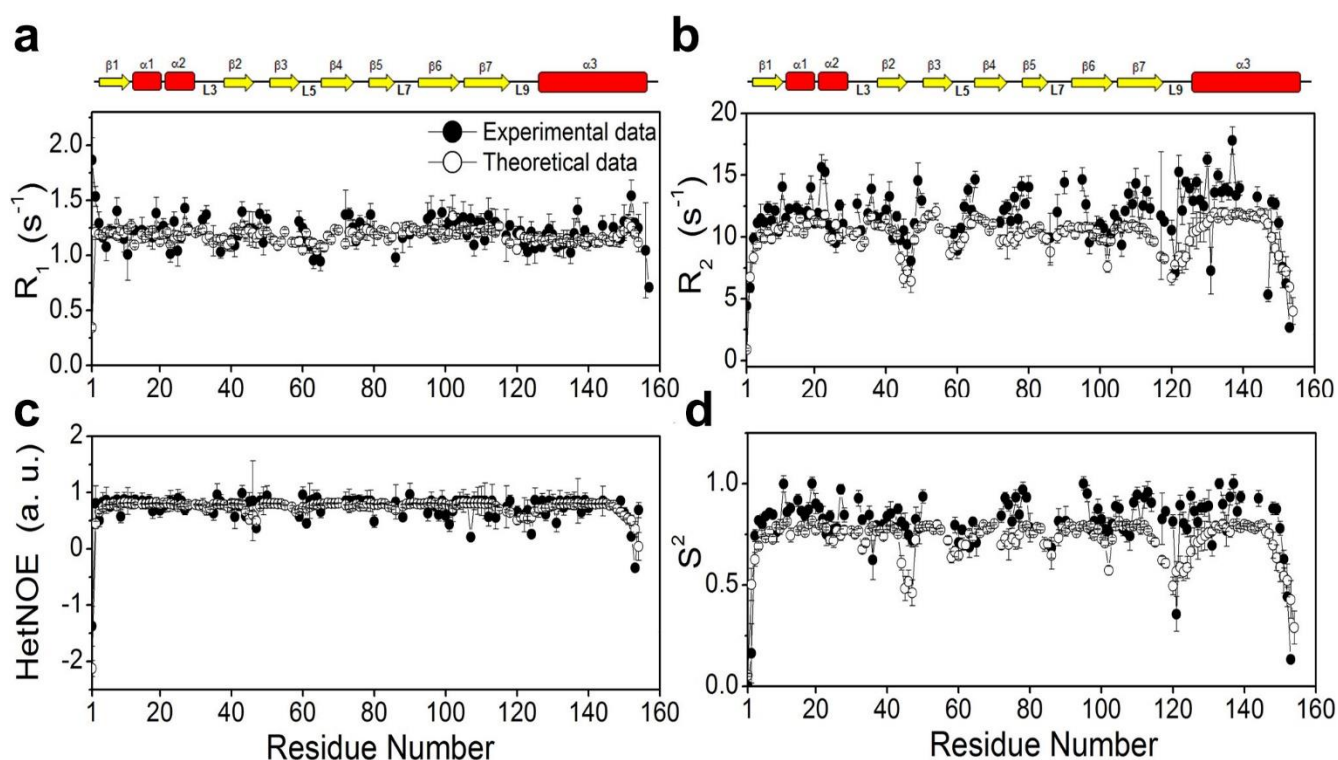


Figure 20 – Backbone molecular dynamics of TcmN at ps-ns timescale monitored by NMR and MD analysis.

NMR experimental data are shown in black and theoretical data in white. (a) R_1 rate, (b) R_2 rate, (c) Het-NOE and (d) S^2 order parameter as function of TcmN residue number.

The results of the extended model-free formalism of Lipari–Szabo using the program Tensor2 are presented in Fig. 20d and Fig. 21b. The value of 9.3 ± 0.7 ns for the correlation time for TcmN agrees with the expected value for a monomeric protein of 21 kDa (Fig. 21a). The dynamic parameters (S_2 and R_{ex}) were calculated using a diffusion model with axial symmetry. The analysis of the R_2/R_1 and R_{ex} ratio indicated segments in conformational exchange (Fig. 21a and b). Residues T22, N23, A36, D49, W63, W65, Y90, W108 and N130 presented higher ratios than average indicating conformation exchange (Fig. 22). Residues in conformational exchange, such as W63 and W65 define the narrow neck region, while W108 and Q110 orient the polyketide chain inside the cavity, suggesting a coupling between TcmN conformation

exchange and fluctuation of its main cavity volume and shape. Similar results were seen for Bet v1 and Fag s1 (ASAM *et al.*, 2014; MORAES *et al.*, 2018).

Carr-Purcell- Meiboom-Gill (CPMG) relaxation dispersion NMR spectroscopy (CARR; PURCELL, 1954; MEIBOOM; GILL, 1958) was measured in order to quantify the conformational exchange identified in some residues of TcmN. $R_{2,eff}$ (s^{-1}) values obtained from ^{15}N CPMG experiments at 298 K (Fig. 21c) showed no significant difference between the $R_{2,eff}$ measured with CPMG frequency of 66.7 and 1000 Hz, meaning that the residues are in exchange regime above 1000 Hz (defined by k_{ex}) and this experiment could not attenuate the conformational exchange contribution of R_2 .

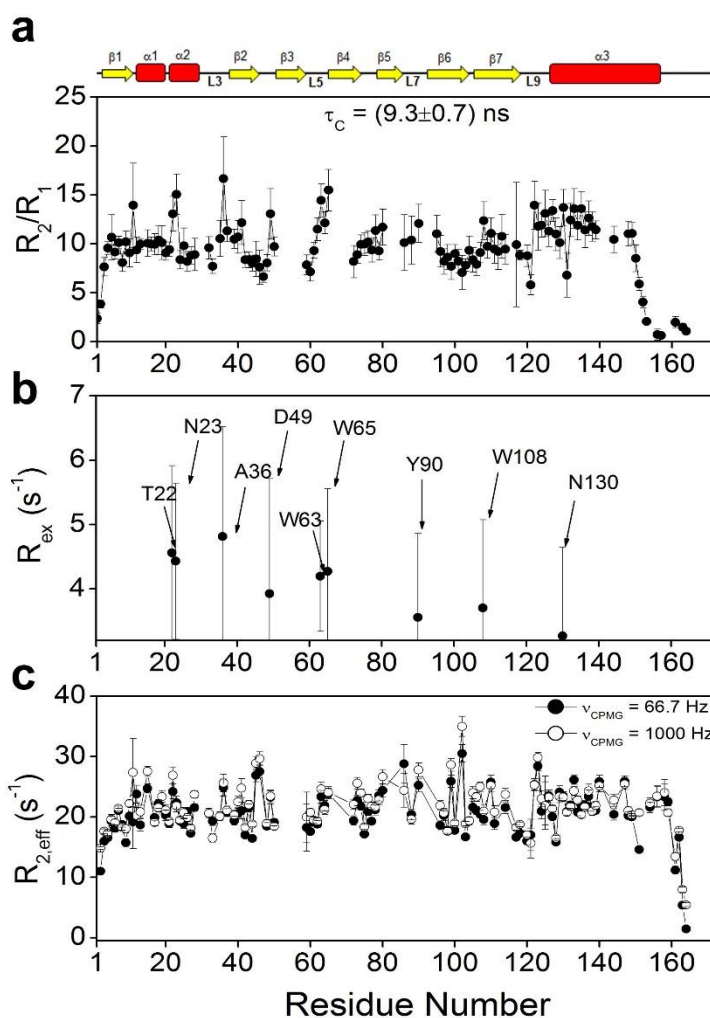


Figure 21 – Backbone molecular dynamics of TcmN monitored by NMR.

(a) ^{15}N R_2/R_1 ratio as function of TcmN residue number. (b) R_{ex} rate, and (c) $R_{2,eff}$ rate as a function of TcmN residue number.

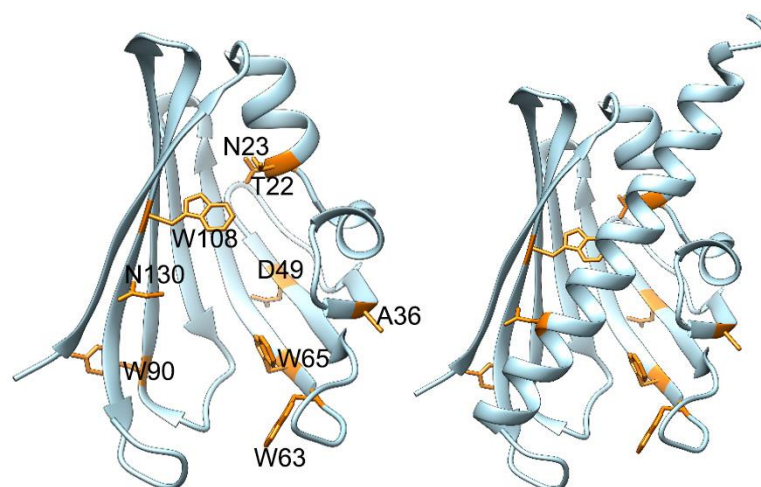


Figure 22 – Side chains of residues showing R_{ex} above 3 Hz colored in orange on the TcmN crystal structure.

$C\alpha$ RMSF values were calculated from the 6 x 1 μ s MD simulations to evaluate atomic-level dynamics of TcmN. The result showed elevated flexibility at L5 and L9, which surround the main cavity (Fig. 23), and was consistent with NMR relaxation measurements. Principal Component Analysis (PCA) was employed to better examine the main contributions to TcmN conformational variance. The first two components, PC1 and PC2, represented 28 and 18% of the $C\alpha$ conformational variance, respectively. These components correspond primarily to movements of the L3, L5, L7, and L9 segments, located at the neck of the central cavity (Fig. 24a). The PC1-PC2 probability-based free energy landscape shows that the two most populated regions correspond to closed and open conformations, respectively (Fig. 24b). Cavity opening/closing motions are well-separated along PC1, with an estimated energy difference of ~ 3 kJ/mol between the closed and open states, with a ~ 6 kJ/mol barrier (Fig. 24c). Cavity volumes of these two states were $1211 \pm 196 \text{ \AA}^3$, and $1713 \pm 132 \text{ \AA}^3$ (t-test $p < 0.0001$), respectively. It can be inferred that movements of L5 and L9 could be necessary to generate a channel through which the polyketide chain can enter the cavity.

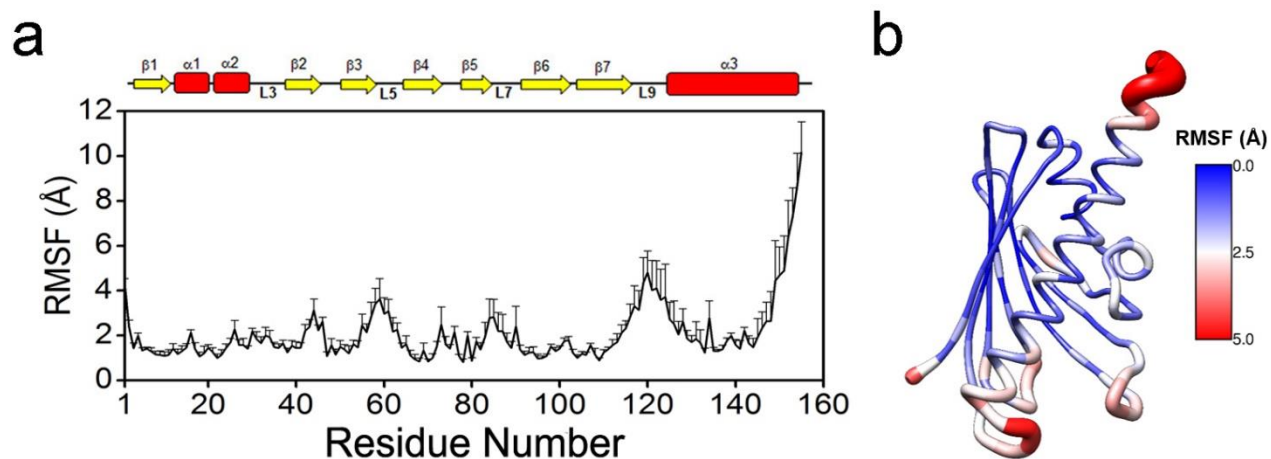


Figure 23 – Residue-specific fluctuations of TcmN from the MD simulations.

(a) Average RMSF \pm standard deviation profile from the 6 x 1 μ s MD simulations. (b) RMSF values of the first replica of TcmN MD simulations represented on TcmN crystal structure by a color gradient and by the ribbon representation width.

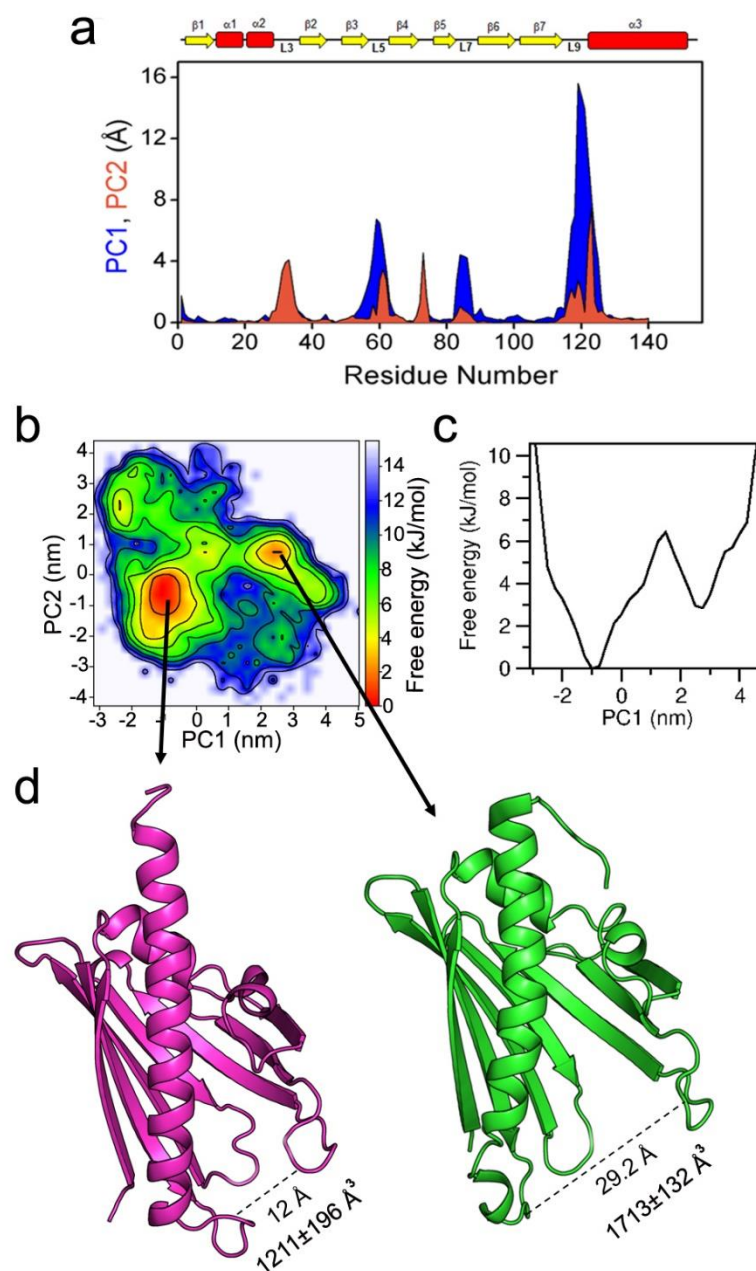


Figure 24 – TcmN conformational diversity.

(a) C α RMSF as a function of TcmN residue number obtained from the PC1 and PC2 projections. 15 highly flexible residues were truncated from the C-terminal of the protein for clarity. (b) PCA correlation plot of the projections of C α variance on the two main components, PC1 and PC2. (c) Free energy variation along PC1. (d) The two minima identified in (b) correspond to the TcmN “closed” (pink) and “opened” (green) ensembles with distance between R61 in L5 and P119 in L9 and cavity volume identified.

To investigate the influence of TcmN dynamics on its aggregation properties, the Aggrescan3D (A3D) Server was used to calculate the aggregation propensity of the two structures corresponding to the closed and open states. In the A3D plot, it is seen that the opened state has more residues with positive values, which are predicted as aggregation-prone (Fig.

25a). The models for each input structure were rendered by their A3D scores (Fig. 25b). The A3D aggregation analysis utilizes an experimentally derived intrinsic aggregation propensity scale for natural amino acids, and projects this scale onto the protein structure, considering the area exposed to solvent (ZAMBRANO *et al.*, 2015). The residues which are aggregation-prone are located inside the TcmN cavity, and are solvent exposed in the open state (Fig. 25b). A3D prediction shows that in the closed conformation, L5 and L9 protect the hydrophobic residues in the cavity (Fig. 25b). However, during substrate binding, the motions of L5 and L9 required to allow ligand entry transiently expose the aggregation-prone regions to the solvent, which may promote aggregation (Fig. 25c). Although the open state of TcmN appears to be at risk of aggregation, MD indicates that it is the less dominant conformation. Evolutionary pressure creates competition between increased binding efficiency and reduced protein stability, balancing the ability to populate multiple states without creating an unacceptable propensity for aggregation (FERROLINO *et al.*, 2013; GERSHENSON *et al.*, 2014). The findings of this work provides an example of a fine tuned balance between TcmN function, conformational variability, and aggregation vulnerability (Fig. 26).

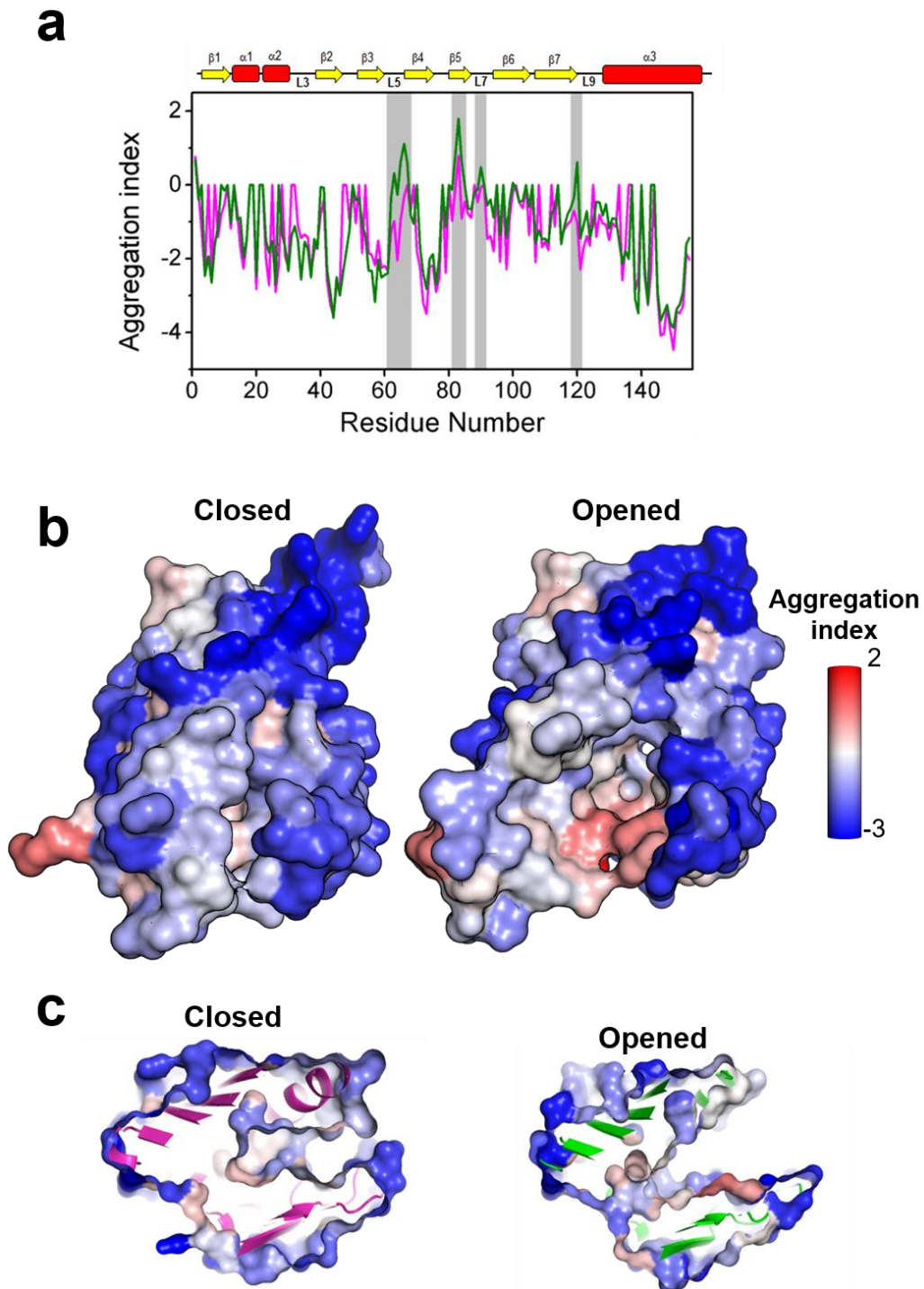


Figure 25 – TcmN aggregation propensity.

(a) Aggregation propensity of TcmN closed (pink) and open (green) states as a function of TcmN residue numbers analyzed by A3D. (b) The protein surface is colored according to A3D score in gradient from red (high-predicted aggregation propensity) to white (negligible impact on protein aggregation) to blue (high-predicted solubility). (c) zoomed view of (b) showing the narrow neck region (left) and the open cavity entrance (right).

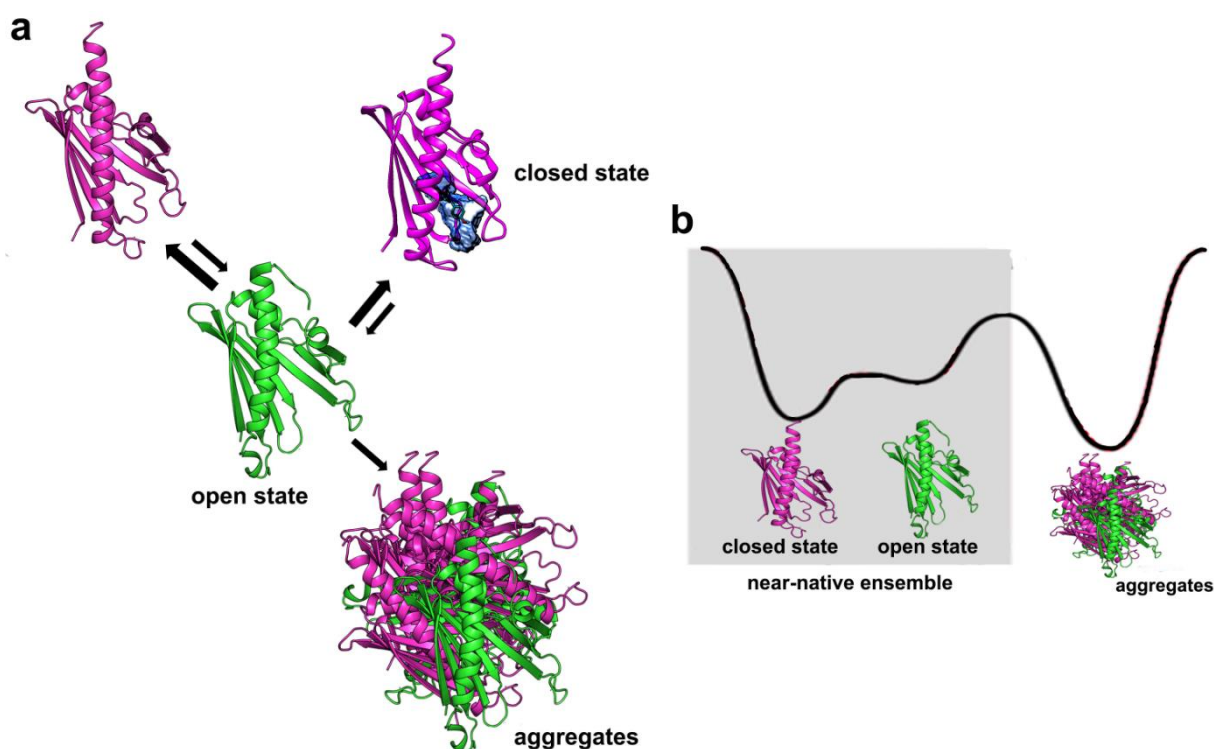


Figure 26 – Proposed interplay between conformational flexibility and aggregation of TcmN.

(a) Schematic representation of TcmN's conformational transitions showing that the opened state (aggregation-prone conformation) is required for ligand binding but also results in aggregation. (b) Energy landscape of TcmN under native conditions comprises a delicate balance between protein conformational flexibility and aggregation.

4.5 TcmN interactions

The TcmN product analogs 8-Anilino-1-naphthalenesulfonic acid (ANS) and Naringerin were used to explore substrate binding in the TcmN enzymatic cavity. Titration of ANS led to considerable changes in the ^1H - ^{15}N HSQC spectra of TcmN, with peaks of several residues disappearing completely during titration (Fig. 28a), which is characteristic for intermediate exchange regime (i.e. dynamics on the μs -ms timescale) of the residues that are interacting with ligand.

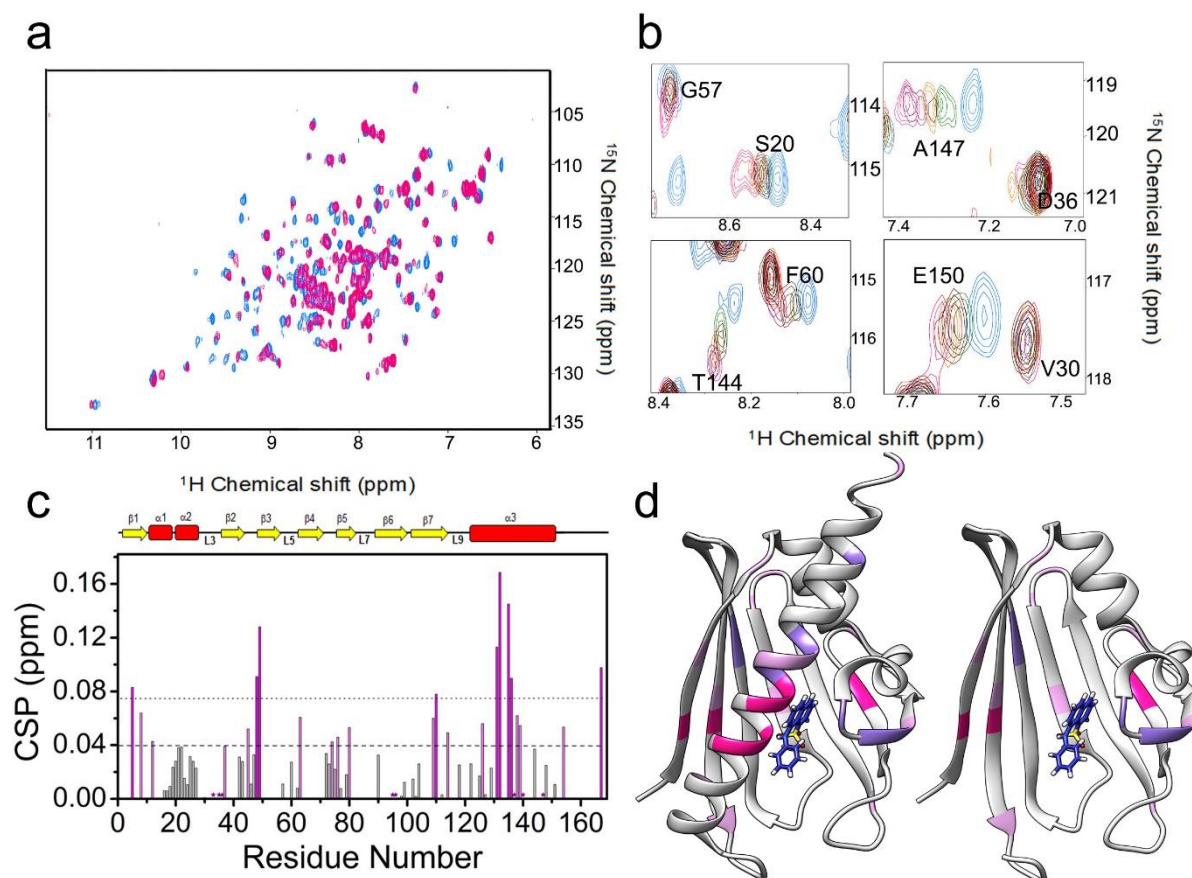


Figure 27 – TcmN and ANS interaction.

(a) ^1H - ^{15}N HSQC spectra of 80 μM of ^{15}N -labeled TcmN colored in blue and 80 μM of ^{15}N -labeled with 560 μM ANS in pink. (b) Selected regions of ^1H - ^{15}N HSQC showing examples of chemical shift variation on ^{15}N -labeled TcmN under titration of ANS. (c) CSP as a function of TcmN residue number. (d) Residues showing CSP values higher than the mean are colored in light pink; values higher than the mean plus the standard deviation are colored in pink and residues that disappeared during titration are shown in purple.

Similar, but more subtle, spectral changes were observed in the titration with Naringerin (Fig. 30a). Fig. 28b and 30b shows examples of chemical shift perturbation indicating residues fast exchange rate, which indicate a dissociation constant (K_d) varying from μM to mM. The calculated CSP for both ANS and Naringerin showed that perturbation occurred in similar regions (residues that shape the cavity and residues of α -helix 3) and residues G45, W63, R76, A80, V109, Q110, T126, T131, T132, A135, N136, and E138 appear to interact with both ANS and Naringerin. Rigid body docking simulations using AutoDock Vina predicted the positions of both ANS and Naringerin inside the TcmN cavity (Fig. 28d and Fig. 30d). ANS was predicted to interact with Y35, W65, S67, A80, R82, L93, M91 and L129 (Fig. 29), and Naringerin with R82, M91, Q110 and M125 (Fig. 31). Among these residues, S67, R82, M91 and L93 were not assigned in the ^1H - ^{15}N -HSQC spectrum, and therefore their CSP could not

be calculated. Residues W35 and A80 showed high CSP values in the ANS titration, and Q110 and M125 showed increased CSP values in the Naringerin titration.

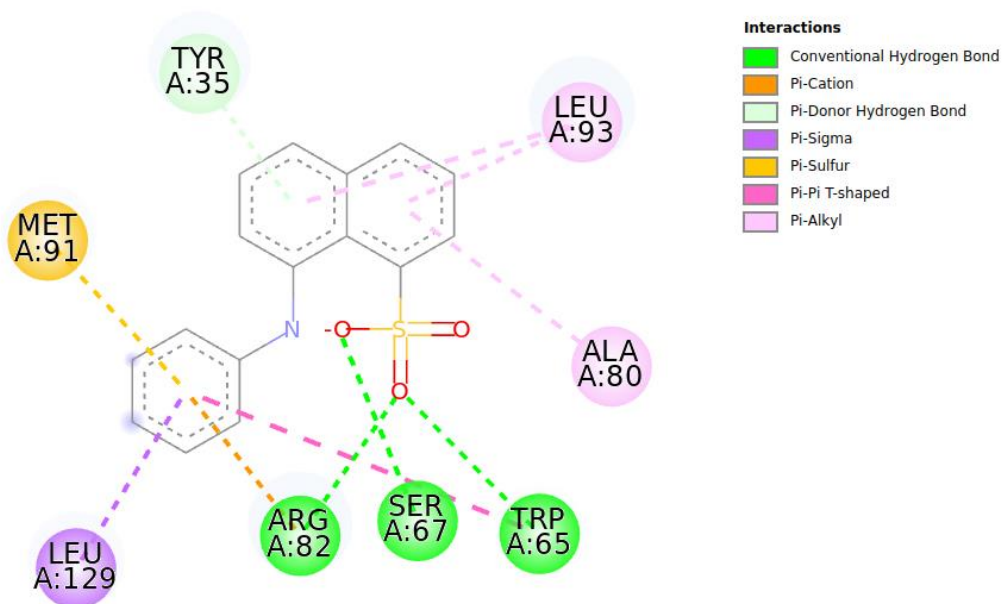


Figure 28 – 2D diagram of docking results of ANS.

2D diagram of the output of AutoDock Vina rendered by Discovery Studio software showing the predicted binding site residues and types of contacts formed between TcmN and ANS.

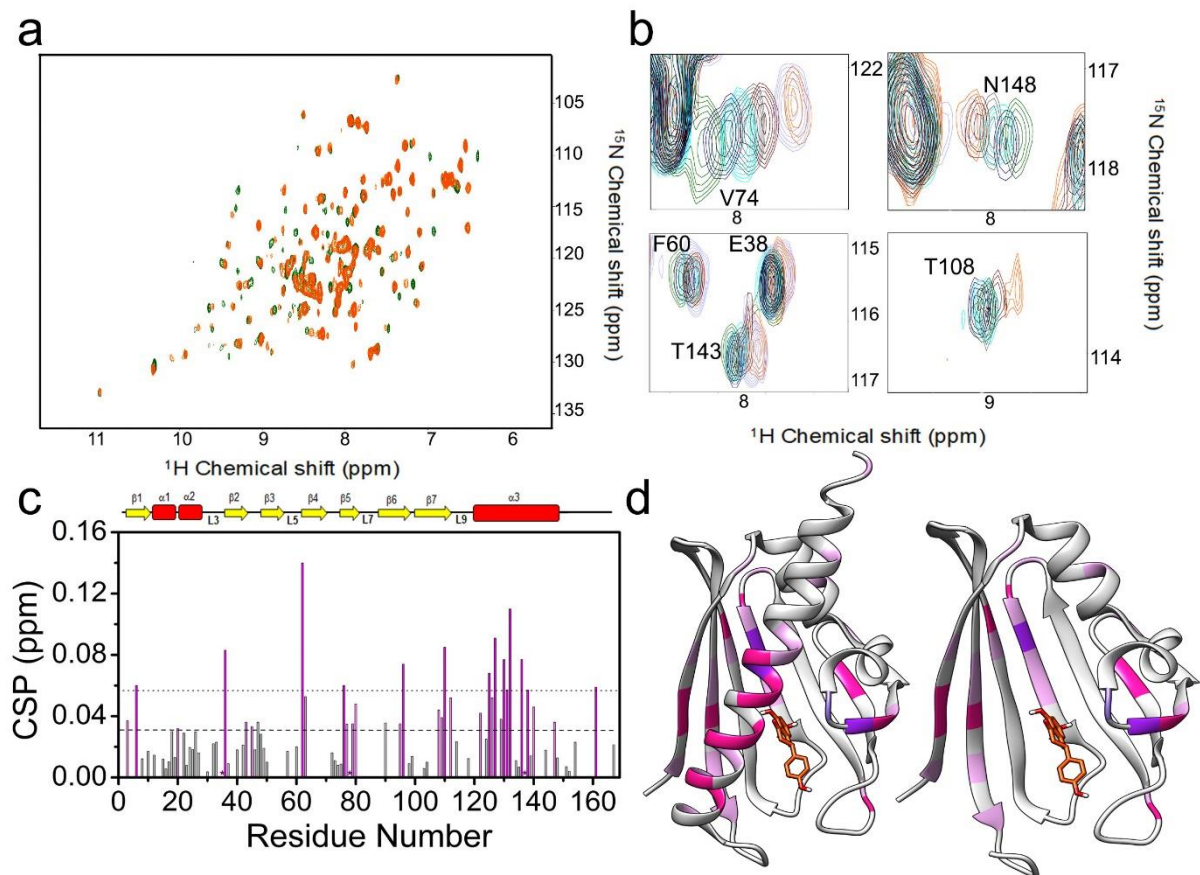


Figure 29 – TcmN and Naringerin interaction.

(a) ^1H - ^{15}N HSQC spectra of 80 μM of ^{15}N -labeled TcmN colored in green and 80 μM of ^{15}N -labeled TcmN with 560 μM Naringerin in orange. (b) Selected regions of ^1H - ^{15}N HSQC showing examples of chemical shift variation on ^{15}N -labeled TcmN under titration of Naringerin. (c) CSP as a function of TcmN residue number. (d) Residues showing CSP values higher than the mean are colored in light pink; values higher than the mean plus the standard deviation are colored in pink and residues that disappeared during titration are shown in purple.

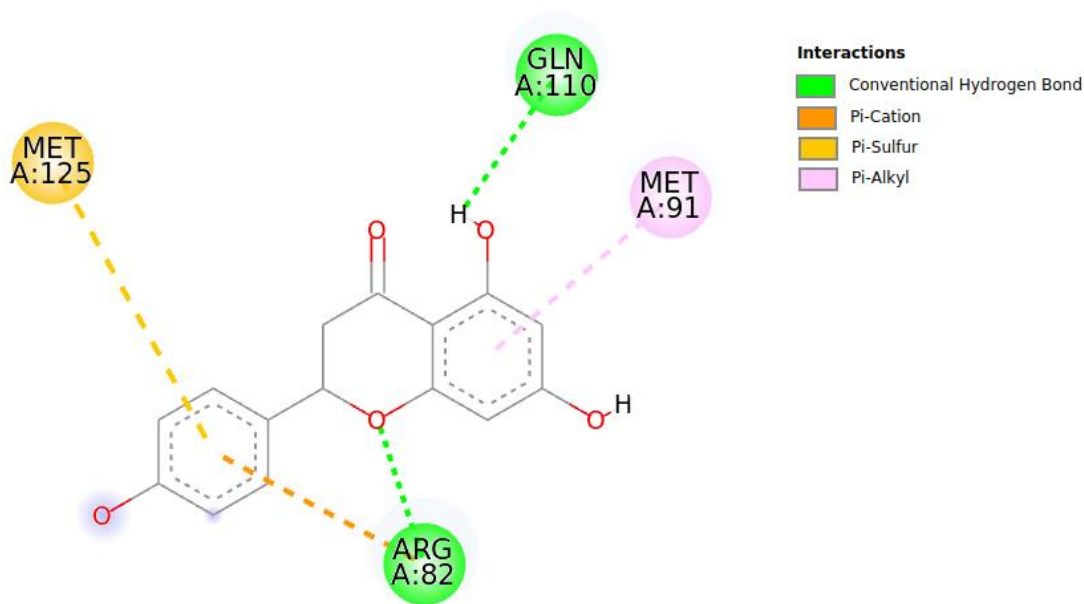


Figure 30 – 2D diagram of docking results of Naringerin.

2D diagram of the output of AutoDock Vina rendered by Discovery Studio software showing the predicted binding site residues and types of contacts formed between TcmN and Naringerin.

5 CONCLUSION

This work characterized TcmN's stability and conformational diversity. Thermal denaturation of TcmN is irreversible, and the denaturation temperature is reduced at higher concentrations. TcmN is more stable at pH 8, and in the presence of NaI and NaCl. Conformational dynamics suggests that substrate binding, and release of the product can be modulated by the conformation of L5 and L9, suggesting a mechanism where the major equilibrium states regulate ligand binding in the main hydrophobic cavity, and protection against solvent exposure and self-aggregation. Understanding the principles underlying TcmN stability, dynamics, and interactions with substrates is expected to be useful for developing it as a platform for *in vitro* polyketide biosynthesis.

6 REFERENCES

- ABRAHAM, M. J.; MURTOLA, T.; SCHULZ, R.; PÁLL, S.; SMITH, J. C.; HESS, B.; LINDAHL, E. **GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.** *SoftwareX*, v. 1–2, p. 19–25, 2015. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>.
- AMES, B. D.; KORMAN, T. P.; ZHANG, W.; SMITH, P.; VU, T.; TANG, Y.; TSAI, S. C. **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. DOI: [10.1073/pnas.0709223105](https://doi.org/10.1073/pnas.0709223105).
- AMES, B. D.; LEE, M. Y.; MOODY, C.; ZHANG, W.; TANG, Y.; TSAI, S. C. **Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway.** *Biochemistry*, v. 50, n. 39, p. 8392–8406, 2011. DOI: [10.1021/bi200593m](https://doi.org/10.1021/bi200593m).
- ASAM, C.; BATISTA, A. L.; MORAES, A. H.; DE PAULA, V. S.; ALMEIDA, F. C. L.; AGLAS, L.; KITZMÜLLER, 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. DOI: [10.1111/cea.12361](https://doi.org/10.1111/cea.12361).
- BENJWAL, S.; VERMA, S.; RÖHM, K.-H.; GURSKY, O. **Monitoring protein aggregation during thermal unfolding in circular dichroism experiments.** *Protein Science*, v. 15, n. 3, p. 635–639, 1 mar. 2006. DOI: [10.1110/ps.051917406](https://doi.org/10.1110/ps.051917406).
- BROERING, J. M.; BOMMARIUS, A. S. **Evaluation of Hofmeister Effects on the Kinetic Stability of Proteins.** *The Journal of Physical Chemistry B*, v. 109, n. 43, p. 20612–20619, 1 nov. 2005. DOI: [10.1021/jp053618+](https://doi.org/10.1021/jp053618+).
- CALDARA-FESTIN, G.; JACKSON, D. R.; BARAJAS, J. F.; VALENTIC, T. R.; PATEL, A. B.; AGUILAR, S.; NGUYEN, M. C.; *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. DOI: [10.1073/pnas.1512976112](https://doi.org/10.1073/pnas.1512976112).
- 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. DOI: [10.1103/PhysRev.94.630](https://doi.org/10.1103/PhysRev.94.630).
- 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. DOI: 10.1039/c8np00040a.
- CHEN, P.; HOLOGNE, M.; WALKER, O.; HENNIG, J. **Ab Initio Prediction of NMR Spin Relaxation Parameters from Molecular Dynamics Simulations.** *Journal of Chemical Theory and Computation*, v. 14, n. 2, p. 1009–1019, 13 fev. 2018. DOI: 10.1021/acs.jctc.7b00750.
- 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. DOI: 10.1021/jp407536p.
- CINO, E. A.; KARTTUNEN, M.; CHOY, W.-Y. **Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins.** *PLOS ONE*, v. 7, n. 11, p. e49876, 26 nov. 2012.
- CINO, E. A.; SOARES, I. N.; PEDROTE, M. M.; DE OLIVEIRA, G. A. P.; SILVA, J. L. **Aggregation tendencies in the p53 family are modulated by backbone hydrogen bonds.** *Scientific Reports*, v. 6, n. 1, p. 32535, 2016. DOI: 10.1038/srep32535.
- CLARKE, D. T. **Circular Dichroism in Protein Folding Studies.** *Current Protocols in Protein Science*, v. 70, n. 1, p. 28.3.1–28.3.17, 1 nov. 2012. DOI: 10.1002/0471140864.ps2803s70.
- DELAGLIO, F.; GRZESIEK, S.; VUISTER, G. W.; ZHU, G.; PFEIFER, J.; BAX, A. **NMRPipe: A multidimensional spectral processing system based on UNIX pipes.** *Journal of Biomolecular NMR*, v. 6, n. 3, p. 277–293, 1995. DOI: 10.1007/BF00197809.
- DOSSET, P.; HUS, J.-C.; BLACKLEDGE, M.; MARION, D. **Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data.** *Journal of Biomolecular NMR*, v. 16, n. 1, p. 23–28, 2000. DOI: 10.1023/A:1008305808620.
- DUY, C.; FITTER, J. **Thermostability of irreversible unfolding alpha-amylases analyzed by unfolding kinetics.** *The Journal of biological chemistry*, v. 280, n. 45, p. 37360–5, 11 nov. 2005. DOI: 10.1074/jbc.M507530200.
- FARROW, N. A.; MUHANDIRAM, R.; SINGER, A. U.; PASCAL, S. M.; KAY, C. M.; GISH, G.; SHOELSON, S. E.; *et al.* **Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by ¹⁵N NMR Relaxation.** *Biochemistry*, v. 33, n. 19, p. 5984–6003, 17 maio 1994. DOI: 10.1021/bi00185a040.
- FERROLINO, M. C.; ZHURAVLEVA, A.; BUDYAK, I. L.; KRISHNAN, B.; GIERASCH,

L. M. **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. DOI: 10.1021/bi4013462.

FITZGERALD, J. T.; CHARKOUDIAN, L. K.; WATTS, K. R.; KHOSLA, C. **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. DOI: 10.1021/ja311579s.

GERSHENSON, A.; GIERASCH, L. M.; PASTORE, A.; RADFORD, S. E. **Energy landscapes of functional proteins are inherently risky.** *Nature Chemical Biology*, v. 10, n. 11, p. 884–891, 2014. DOI: 10.1038/nchembio.1670.

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. DOI: 10.1590/S1517-83822013000400002.

GOYAL, M.; CHAUDHURI, T. K.; KUWAJIMA, K. **Irreversible Denaturation of Maltodextrin Glucosidase Studied by Differential Scanning Calorimetry, Circular Dichroism, and Turbidity Measurements.** *PLOS ONE*, v. 9, n. 12, p. e115877, 30 dez. 2015.

HERTWECK, C.; LUZHETSKYY, A.; REBETS, Y.; BECHTHOLD, A. **Type II polyketide synthases: Gaining a deeper insight into enzymatic teamwork.** *Natural Product Reports*, v. 24, n. 1, p. 162–190, 2007. DOI: 10.1039/b507395m.

HYBERTS, S. G.; MILBRADT, A. G.; WAGNER, A. B.; ARTHANARI, H.; WAGNER, G. **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. DOI: 10.1007/s10858-012-9611-z.

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. DOI: 10.1021/ja908004w.

IYER, L. M.; KOONIN, E. V.; ARAVIND, L. **Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily.** *Proteins: Structure, Function, and Bioinformatics*, v. 43, n. 2, p. 134–144, 1 maio 2001. DOI: 10.1002/1097-0134(20010501)43:2<134::AID-PROT1025>3.0.CO;2-I.

JAVIDPOUR, P.; BRUEGGER, J.; SRITHAHAN, S.; KORMAN, T. P.; CRUMP, M. P.;

- CROSBY, J.; BURKART, M. D.; *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. DOI: 10.1016/j.chembiol.2013.07.016.
- 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. DOI: 10.1021/acs.jpcc.8b03295.
- KATSUYAMA, Y.; OHNISHI, Y. *Type III polyketide synthases in microorganisms*. 1. ed. [S.l.]: Elsevier Inc., 2012. v. 515.
- KAY, L. E.; TORCHIA, D. A.; BAX, A. **Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease.** *Biochemistry*, v. 28, n. 23, p. 8972–8979, 14 nov. 1989. DOI: 10.1021/bi00449a003.
- KELLY, A. E.; OU, H. D.; WITHERS, R.; DÖTSCH, V. **Low-Conductivity Buffers for High-Sensitivity NMR Measurements.** *Journal of the American Chemical Society*, v. 124, n. 40, p. 12013–12019, 1 out. 2002. DOI: 10.1021/ja026121b.
- KELLY, S. M.; JESS, T. J.; PRICE, N. C. **How to study proteins by circular dichroism.** *Biochimica et Biophysica Acta - Proteins and Proteomics*, v. 1751, n. 2, p. 119–139, 10 ago. 2005. DOI: 10.1016/j.bbapap.2005.06.005.
- KNOWLES, T. P. J.; VENDRUSCOLO, M.; DOBSON, C. M. **The amyloid state and its association with protein misfolding diseases.** *Nature Reviews Molecular Cell Biology*, v. 15, n. 6, p. 384–396, 2014. DOI: 10.1038/nrm3810.
- 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. DOI: 10.1021/bi201705q.
- LEŠNIK, U.; LUKEŽIČ, T.; PODGORŠEK, A.; HORVAT, J.; POLAK, T.; ŠALA, M.; JENKO, B.; *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. DOI: 10.1002/anie.201411028.
- 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. DOI: 10.1016/j.synbio.2018.02.002.

LIAN, L.-Y. **NMR studies of weak protein–protein interactions.** *Progress in Nuclear Magnetic Resonance Spectroscopy*, v. 71, p. 59–72, 2013. DOI: <https://doi.org/10.1016/j.pnmrs.2012.11.002>.

LIPARI, G.; SZABO, A. **Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results.** *Journal of the American Chemical Society*, v. 104, n. 17, p. 4559–4570, 1 ago. 1982. DOI: 10.1021/ja00381a010.

LIU, X.; HUA, K.; LIU, D.; WU, Z.-L.; WANG, Y.; ZHANG, H.; DENG, Z.; *et al.* **Heterologous Biosynthesis of Type II Polyketide Products Using E. coli.** *ACS Chemical Biology*, 11 dez. 2019. DOI: 10.1021/acscchembio.9b00827.

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. DOI: 10.1021/ja983961a.

MAZURENKO, S.; STOURAC, J.; KUNKA, A.; NEDELJKOVIĆ, S.; BEDNAR, D.; PROKOP, Z.; DAMBORSKY, J. **CalFitter: a web server for analysis of protein thermal denaturation data.** *Nucleic Acids Research*, v. 46, n. W1, p. W344–W349, 14 maio 2018. DOI: 10.1093/nar/gky358.

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. DOI: 10.1063/1.1716296.

MORAES, A. H.; ASAM, C.; ALMEIDA, F. C. L.; WALLNER, M.; FERREIRA, F.; VALENTE, A. P. **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. DOI: 10.1038/s41598-018-28358-1.

MORAES, A. H.; ASAM, C.; BATISTA, A.; ALMEIDA, F. C. L.; WALLNER, M.; FERREIRA, F.; VALENTE, A. P. **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. DOI: 10.1007/s12104-015-9634-y.

- 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
- MUNTAU, A. C.; LEANDRO, J.; STAUDIGL, M.; MAYER, F.; GERSTING, S. W. **Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators**. *Journal of Inherited Metabolic Disease*, v. 37, n. 4, p. 505–523, 1 jul. 2014. DOI: 10.1007/s10545-014-9701-z.
- OKUR, H. I.; HLADÍLKOVÁ, J.; REMBERT, K. B.; CHO, Y.; HEYDA, J.; DZUBIELLA, J.; CREMER, P. S.; *et al.* **Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions**. *The Journal of Physical Chemistry B*, v. 121, n. 9, p. 1997–2014, 9 mar. 2017. DOI: 10.1021/acs.jpcc.6b10797.
- OLSSON, M. H. M.; SØNDERGAARD, C. R.; ROSTKOWSKI, M.; JENSEN, J. H. **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. DOI: 10.1021/ct100578z.
- ONUCHIC, J. N.; LUTHEY-SCHULTEN, Z.; WOLYNES, P. G. **THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective**. *Annual Review of Physical Chemistry*, v. 48, n. 1, p. 545–600, 1 out. 1997. DOI: 10.1146/annurev.physchem.48.1.545.
- PALMER, A. G. **NMR Characterization of the Dynamics of Biomacromolecules**. *Chemical Reviews*, v. 104, n. 8, p. 3623–3640, 1 ago. 2004. DOI: 10.1021/cr030413t.
- PATODIA, S.; BAGARIA, A.; CHOPRA, D. **Molecular Dynamics Simulation of Proteins: A Brief Overview**. *Journal of Physical Chemistry & Biophysics*, v. 4, 1 jan. 2014. DOI: 10.4172/2161-0398.1000166.
- PFEIFER, B. A.; KHOSLA, C. **Biosynthesis of Polyketides in Heterologous Hosts**. *Microbiology and Molecular Biology Reviews*, v. 65, n. 1, p. 106–118, 2001. DOI: 10.1128/mmbr.65.1.106-118.2001.
- 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. DOI: 10.1016/j.bpj.2011.03.051.
- PLOTKIN, S. S.; ONUCHIC, J. N. **Understanding protein folding with energy landscape theory Part I: Basic concepts**. *Quarterly Reviews of Biophysics*, v. 35, n. 2, p. 111–167, 2002a. DOI: DOI: 10.1017/S0033583502003761.

PLOTKIN, S. S.; ONUCHIC, J. N. **Understanding protein folding with energy landscape theory Part II: Quantitative aspects.** *Quarterly Reviews of Biophysics*, v. 35, n. 3, p. 205–286, 2002b. DOI: DOI: 10.1017/S0033583502003785.

PRICE, S. M. K. AND N. C. *The Use of Circular Dichroism in the Investigation of Protein Structure and Function.* *Current Protein & Peptide Science.* [S.l: s.n.]. , 2000

QIN, J.; GRONENBORN, A. M. **Weak protein complexes: challenging to study but essential for life.** *The FEBS journal*, v. 281, n. 8, p. 1948–1949, abr. 2014. DOI: 10.1111/febs.12744.

RAWLINGS, B. J. **Biosynthesis of polyketides (other than actinomycete macrolides).** *Natural Product Reports*, v. 16, n. 4, p. 425–484, 1999. DOI: 10.1039/a900566h.

ROHR, J.; HERTWECK, C. **Type II PKS.** *Comprehensive Natural Products II: Chemistry and Biology*, v. 1, p. 227–303, 2010. DOI: 10.1016/b978-008045382-8.00703-6.

SANCHO, J. **The stability of 2-state, 3-state and more-state proteins from simple spectroscopic techniques... plus the structure of the equilibrium intermediates at the same time.** *Archives of Biochemistry and Biophysics*, v. 531, n. 1, p. 4–13, 2013. DOI: <https://doi.org/10.1016/j.abb.2012.10.014>.

SCHWIERZ, N.; HORINEK, D.; NETZ, R. R. **Reversed Anionic Hofmeister Series: The Interplay of Surface Charge and Surface Polarity.** *Langmuir*, v. 26, n. 10, p. 7370–7379, 18 maio 2010. DOI: 10.1021/la904397v.

TAVERNELLI, I.; COTESTA, S.; DI IORIO, E. E. **Protein Dynamics, Thermal Stability, and Free-Energy Landscapes: A Molecular Dynamics Investigation.** *Biophysical Journal*, v. 85, n. 4, p. 2641–2649, 1 out. 2003. DOI: 10.1016/S0006-3495(03)74687-6.

TIAN, W.; CHEN, C.; LEI, X.; ZHAO, J.; LIANG, J. **CASTp 3.0: computed atlas of surface topography of proteins.** *Nucleic Acids Research*, v. 46, n. W1, p. W363–W367, 1 jun. 2018. DOI: 10.1093/nar/gky473.

UEBERSCHAAR, N.; XU, Z.; SCHERLACH, K.; METSÄ-KETELÄ, M.; BRETSCHNEIDER, T.; DAHSE, H.-M.; GÖRLS, H.; *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. DOI: 10.1021/ja4080024.

VRANKEN, W. F.; BOUCHER, W.; STEVENS, T. J.; FOGH, R. H.; PAJON, A.; LLINAS, M.; ULRICH, E. L.; *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. DOI: 10.1002/prot.20449.

WEISSMAN, K. J. **Chapter 1 Introduction to Polyketide Biosynthesis.** *Methods in Enzymology*, v. 459, n. B, p. 3–16, 2009. DOI: 10.1016/S0076-6879(09)04601-1.

WEISSMAN, K. J.; LEADLAY, P. F. **Combinatorial biosynthesis of reduced polyketides.** *Nature Reviews Microbiology*, v. 3, n. 12, p. 925–936, 2005. DOI: 10.1038/nrmicro1287.

WÜTHRICH, K. **NMR of Proteins and Nucleic Acids.** *Wiley, John & Sons*, v. 1st, p. 13 – 203, 1986.

ZAMBRANO, R.; JAMROZ, M.; SZCZASIUK, A.; PUJOLS, J.; KMIĘCIK, S.; VENTURA, S. **AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures.** *Nucleic Acids Research*, v. 43, n. W1, p. W306–W313, 16 abr. 2015. DOI: 10.1093/nar/gkv359.

ZAWADA, R. J. X.; KHOSLA, C. **Heterologous expression, purification, reconstitution and kinetic analysis of an extended type II polyketide synthase.** *Chemistry and Biology*, v. 6, n. 9, p. 607–615, 1999. DOI: 10.1016/S1074-5521(99)80112-4.

ZHANG, Z.; PAN, H.-X.; TANG, G.-L. **New insights into bacterial type II polyketide biosynthesis.** *F1000Research*, v. 6, p. 172, 21 fev. 2017. DOI: 10.12688/f1000research.10466.1.