A kinetic and structure-guided characterization of Scabin, a novel mono-ADP-ribosyltransferase produced by Streptomyces scabies

by

Bronwyn J. Lyons

A Thesis
presented to
The University of Guelph

In partial fulfillment of requirements
for the degree of
Master of Science
in
Molecular and Cellular Biology

Guelph, Ontario, Canada
© Bronwyn J. Lyons, July, 2017
ABSTRACT

A KINETIC AND STRUCTURE-GUIDED CHARACTERIZATION OF SCABIN, A NOVEL MONO-ADP-RIBOSYLTRANSFERASE PRODUCED BY STREPTOMYCES SCABIES

Bronwyn J. Lyons
University of Guelph, 2017

Advisor: Dr. Rod Merrill

Mono-ADP-ribosyltransferase toxins are produced by pathogenic bacteria as virulence factors that target important macromolecules in host cells. In a few cases, the cellular target may be DNA. This family of enzymes transfers an ADP-ribose moiety from NAD$^+$ to the target macromolecule, leading to an altered function of the target and ultimately host-cell death. A bioinformatics strategy was used to identify Scabin, a mono-ADP-ribosyltransferase from the plant pathogen Streptomyces scabies. A detailed kinetic analysis was performed on Scabin, revealing the target as genomic DNA. The crystal structure of Scabin with NADH as a substrate analog was determined, which provided important insights into the active site structure of the enzyme. Residues involved in activity and binding of DNA were identified. Hydrogen-deuterium exchange coupled with mass spectrometry was used to characterize the Scabin-DNA interface, revealing key interacting regions. Understanding the mechanism of Scabin will allow for a more targeted approach in the development of inhibitors against the potentially toxic activity of this enzyme.
Acknowledgements

I would first like to thank my advisor, Dr. Rod Merrill, for his constant support during my Master’s and my Undergraduate projects. I truly would not be half the researcher that I am today without the opportunities I have had in your lab! I would also like to thank my committee members, Dr.’s David Josephy and Richard Manderville, for their guidance and valuable discussions.

I appreciate all the collaborators involved on this work, including Dr. Derek Wilson (Chemistry Department, York University), Dr. Jaideep Mathur (Molecular and Cellular Biology Department, University of Guelph) and Dr. Praveen Saxena (Gosling Research Institute for Plant Preservation, University of Guelph). Thank you to Cristina Lento (Chemistry Department, York University), Kathleen Delfosse (Molecular and Cellular Biology Department, University of Guelph) and Dr. Sherif Sherif (Gosling Research Institute for Plant Preservation, University of Guelph). I am thankful for the assistance of Dr. Dyanne Brewer with mass spectrometric analysis, and the beamline scientists at the Canadian Macromolecular Crystallography Facility at the Canadian Light Source (University of Saskatchewan) for their collection of data. This research was sponsored financially by the Canadian government through a Natural Sciences and Engineering Research Council (NSERC) Canadian Graduate Scholarship (CGS-M).

I have had the privilege of mentoring several students who have all contributed to my project: Stephanie Carlin, Taylor Lidster, Thomas McAusland, Juanita Francis, Flor Ali Salehie-Barbosa, Jesse Rentz, Taylor Elliot, Haya Ahmed and Maritza Vatta. Past and present members of the Merrill lab have provided endless help with my project. Thank you to Dr.’s Debajyoti Dutta, Ravi Ravulapalli and Miguel Lugo for all of your guidance. Thank you to Kayla Heney, Lachae Hood, Rebecca Jagroop, Dan Krska, Amanda Poole, Simone Renwick, Madison Turner, Olivier Tremblay and Chantal Wood, you have all made long days in the lab that much more enjoyable.
Thank you to everyone on the second floor, especially Olivia, Connor, Evan and Kanu. I would also like to thank my fantastic house/office mates, Megan Brasher and Patrick Ryan. I am so grateful for the hilarious adventures that we have had these past two years. Thank you to my mom, dad and brother, for always being there for moral support. Last but definitely not least, thank you to my dear friends Meagan, Nicole and Reaghan.
Declaration of work performed

In some cases, it has been necessary to include experiments performed by other researchers in order for a complete story to be written. All the work contained in this thesis is my own, with the following exceptions: T. Keeling performed all cloning and mutagenesis for the Q158AxE160A variant. T. McAusland performed mutagenesis for the W155A and W199A variants. J. Lanoue crystallized apo-Scabin and D. Dutta solved the structure; R. Ravulapalli solved the PJ34 and P6-E co-crystal structures. J. Lanoue characterized the glycohydrolase activity of wild-type Scabin and identified and characterized inhibition by PJ34 and P6-F compounds. Inhibition assays for P6-C, P6-D and P6-E were performed by S. Carlin. D. Brewer performed and analyzed mass spectrometry data for native Scabin (2.2) and reaction products (2.10). S. Carlin performed all kinetic assays (both glycohydrolase and transferase) for Scabin W128Y, W155A and W199A. T. Lidster developed and performed all experiments for the monobromobimane assay (A1.2). T. Lidster crystallized the Scabin W128Y, W155A, N110A and S117A variants. K. Heney crystallized Scabin W128Y with NADH. Hydrogen-deuterium exchange experiments were performed by C. Lento at York University; data were analyzed by myself under close consultation with C. Lento. M. Lugo performed and analyzed all molecular modelling experiments described in: A.1.7, A1.8, and A1.9. Deconvolution of Trp emission spectra, as described under A.11, were performed by M. Lugo. M. Lugo developed the model for analyzing anisotropy data (2.13, A1.3). S. Sherif performed the plant studies described under A1.6 and A1.7. J. Mathur developed the seedling assay described under A1.5; experiments were performed under close consultation with K. Delfosse.
Abbreviations:

ADP – Adenosine diphosphate
ADPRT – Adenosine diphosphate-ribosyltransferase
ARTT – ADP-ribosyl-turn-turn
CD – Circular dichroism
cGMP – Cyclic guanosine monophosphate
CHAPS – 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CT – Cholera toxin
dG – Deoxyguanosine
dI – Deoxyinosine
DMSO – Dimethyl sulfoxide
DNA – Deoxyribonucleic acid
DSF – Differential scanning fluorimetry
DTT - Dithiothreitol
DT – Diphtheria toxin
εAMP – Etheno adenosine monophosphate
EDTA – Ethylenediaminetetraacetic acid
εEF2 – Eukaryotic elongation factor two
εNAD – Etheno nicotinamide adenine dinucleotide
GDP – Guanosine diphosphate
GH – Glycohydrolase
HB-EGF – Heparin binding epidermal growth factor
HDX – Hydrogen-deuterium exchange
HPLC – High performance liquid chromatography
ICE – Integrative conjugating element
IMAC – Immobilized metal affinity chromatography
IPTG – Isopropyl β-D-1-thiogalactopyranoside
LB – Luria Broth
LC – Liquid chromatography
LIST – Ligand-induced slow transition model
LRP-1 – Lipoprotein receptor-related protein 1
MALDI-TOF – Matrix assisted laser desorption/ionization – time of flight
mART – Mono adenosine diphosphate ribosyltranferase
MDR – Multi-drug resistance
MS – Murashige and Skoog
MTX – Mosquitocidal toxin
NAD – Nicotinamide adenine dinucleotide
NER – Nuceotide excision repair
PARP – Poly-ADP-ribosylpolymerase
P6-C – 8-fluoro-1H,2H,3H,4H,5H,6H-benzo[c]1,6-naphthyridin-6-one
P6-D – 8-fluoro-2-[3-(piperidin-1-yl)propyl]-1H,2H,3H,4H,5H,6H-benzo[c]1,6-naphthyridin-6-one
P6-E – 4-[8-fluoro-6-oxo 1H,2H,3H,4H,5H,6H-benzo[c]1,6-naphthyridin-2-yl]butanoic acid
P6-F – 8-fluoro-2-[3 (piperidin-1-yl)propanesulfonyl]-1H,2H,3H,4H, 5H,6H-benzo[c]1,6-naphthyridin-6-one
PJ34 – N-(6-Oxo-5,6-dihydrophenanthridin-2-yl)-(N,N-dimethylamino)acetamide hydrochloride
PMSF – Phenylmethanesulfonyl fluoride
PN – Phosphate-nicotinamide
PT – Pertussis toxin
Q-TOF – Quadrupole-time of flight
SAAB – Selection and amplification binding assay
SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis
$S_{N1}$ – Unimolecular nucleophilic substitution
T3SS – Type III secretion system
Tat – Twin arginine transport
vdW – van der Waals
List of Figures

FIGURE 1.1: Structural comparison of DT- and CT-like toxins......................................................... 7
FIGURE 1.2: Schematic of TatABC translocase complex................................................................. 10
FIGURE 1.3: Schematic for the ADP-ribosylation mechanism of guanine nucleotides.............. 14
FIGURE 1.4: Streptomyces scabies and the common scab ................................................................. 20
FIGURE 1.5: Schematic of the cellobiose ABC transporter, CebEFG ............................................. 22

FIGURE 2.1: TRESI schematic ........................................................................................................... 32

FIGURE 3.1: Multiple-sequence alignment of Scabin with various Pierisin-like mART toxins. 43
FIGURE 3.2: Folded stability and enzymatic activity of Scabin......................................................... 46
FIGURE 3.3: Mass spectrometry of Scabin with oligonucleotide substrates................................. 51
FIGURE 3.4: Proposed general kinetic scheme to describe Scabin’s observed hysteretic activity. 54
FIGURE 3.5: P-series inhibitors effective against Scabin GH activity.............................................. 57
FIGURE 3.6: Scabin-apo crystal structures......................................................................................... 61
FIGURE 3.7: Scabin inhibitor crystal structures................................................................................ 65
FIGURE 3.8: The Scabin·PJ34 complex ............................................................................................ 68

FIGURE 4.1: Scabin binds competitive inhibitor/substrate NADH................................................ 74
FIGURE 4.2: Spectral data of Trp variants. ......................................................................................... 81
FIGURE 4.3: Analysis of WT vs. variant Scabin structures ............................................................... 88
FIGURE 4.4: NADH and DNA binding............................................................................................. 93
FIGURE 4.5: Scabin-NADH-DNA model......................................................................................... 98

FIGURE 5.1: HDX profile of apоСcabin ............................................................................................ 107
FIGURE 5.2: HDX of Scabin with bound NADH substrate analog.................................................. 111
FIGURE 5.3: HDX of Scabin bound with dsDNA substrate............................................................. 116
FIGURE 5.4: Interaction of PARP-1 with a single-strand break (SSB) DNA ................................. 118

FIGURE A.1: Experimental set up for ‘Treatment of G-mito Arabidopsis thaliana seedlings with Scabin toxin’ .................................................................................................................. 143
FIGURE A.2: Scabin induces cellular stress response in Solanum tuberosum .............................. 147
FIGURE A.3: Scabin induces a clustered mitochondria response in G-mito Arabidopsis thaliana seedlings .................................................................................................................................... 149
FIGURE A.4: Proposed map of binding pocket for ternary complex of Scabin ............................. 150
List of Tables

TABLE 2.1: Summary of variants developed for studying Scabin toxin. .................................................. 27

TABLE 3.1: Kinetic parameters and NAD^+ binding affinity of Scabin and Q158A/E160A variant ................................................................. 48

TABLE 3.2: Comparison of $K_D$, IC_{50}, and solubility parameter values for inhibitors of Scabin . 58

TABLE 3.3: Crystallographic data and refinement statistics for apo-Scabin, Scabin-PJ34 and Scabin P6-E ............................................................................................................ 62

TABLE 4.1A: Crystallographic data and refinement statistics for Scabin-NADH complex, Scabin W128Y, and Scabin W128Y-NADH complex. ............................................................... 75

TABLE 4.1B: Crystallographic data and refinement statistics for Scabin W155A, Scabin S117A, and Scabin N110A ............................................................................................................. 76

TABLE 4.2: Percentage of free thiols present in different Scabin variants ............................................. 78

TABLE 4.3: Fluorescence parameters of WT Scabin and tryptophan variants ........................................ 82

TABLE 4.4: Kinetic parameters of Scabin WT and variants for GH activity and NAD^+ binding. ................................................................................................................................. 86

TABLE 4.5: Kinetic parameters of Scabin WT and variants for transferase activity ......................... 90

TABLE 4.6: Binding constants of Scabin WT and variants with blunt-end double-stranded DNA. ................................................................................................................................. 95

TABLE 4.7: Binding constants of Scabin WT for different synthetic double-stranded DNA. .... 96
# TABLE OF CONTENTS

_Abtract_ .............................................................................................................................. i  
_Acknowledgements_ ............................................................................................................. iii  
_Declaration of work performed_ ............................................................................................ v  
_Abbreviations_ ....................................................................................................................... vi  
_List of Figures_ ...................................................................................................................... ix  
_List of Tables_ ....................................................................................................................... x

## CHAPTER 1: INTRODUCTION

1.1: Antibiotic resistance ........................................................................................................... 1  
1.2: Anti-virulence strategy .................................................................................................... 2  
1.3: mART toxins .................................................................................................................. 3  
1.4: Classification .................................................................................................................... 3  
1.4.1: DT-like ....................................................................................................................... 4  
1.4.2: CT-like ....................................................................................................................... 5  
1.5: Mechanisms of secretion ................................................................................................ 6  
1.6: Examples of important mART toxins ............................................................................. 7  
1.6.1: Mosquitocidal toxin from _Bacillus sphaericus_ ......................................................... 8  
1.6.2: HopU1 from _Pseudomonas syringae_ ........................................................................ 9  
1.6.3: A new mART subfamily - Pierisin-1 from _Pieris rapae_ ........................................... 10  
1.7: Mechanism ..................................................................................................................... 11  
1.8: Structural motifs of DNA-binding proteins ..................................................................... 12  
1.9: Methods for structural characterization of macromolecular complexes ...................... 13  
1.9.1: X-ray crystallography ................................................................................................. 14  
1.9.2: Hydrogen-deuterium exchange ............................................................................... 15  
1.9.3: Small-angle X-ray scattering .................................................................................... 16  
1.10: Streptomyces scabies and the common scab disease of potatoes ............................. 17  
1.10.1: _Streptomyces scabies_ ............................................................................................. 18  
1.10.2: Scabin, a putative mono-ADP-ribosyltransferase toxin ....................................... 19  
1.11: Research rationale and objectives .............................................................................. 20

## CHAPTER 2: MATERIALS AND METHODS

2.1: Expression and purification of Scabin .............................................................................. 24  
2.2: Confirmation of Scabin identity by mass spectrometry ................................................. 25  
2.3: Circular dichroism spectra ............................................................................................... 26  
2.4: Protein crystallography ................................................................................................... 27  
2.4.1: Scabin-apo structure ................................................................................................. 28  
2.4.2: Scabin PJ34 and P6-E structures ............................................................................. 29  
2.5: Hydrogen-deuterium exchange of Scabin .................................................................... 30  
2.6: Hydrogen-deuterium exchange data analysis ............................................................... 31  
2.7: β-NAD and inhibitor binding ......................................................................................... 32  
2.8: Glycohydrolase activity ................................................................................................. 33  
2.9: ADP-ribosyltransferase reaction .................................................................................. 34  
2.10: Mass spectrometry of mART reaction products ......................................................... 34
5.3: Conclusion ............................................................................................................... 119

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS ........................................ 120
6.1: Conclusions ......................................................................................................... 121
6.2: Future directions ............................................................................................... 123
   6.2.1: Characterizing the macromolecular complex of Scabin with DNA .......... 123
   6.2.2: Identifying the target gene for Scabin ......................................................... 124
   6.2.3: The role of Scabin in Streptomyces scabies pathogenesis ......................... 125
6.3: Concluding remarks ........................................................................................... 126

REFERENCES ............................................................................................................ 127

APPENDIX .................................................................................................................. 139
A.1: Methods ............................................................................................................. 140
   A.1.1: Microfluidic chip fabrication ................................................................. 140
   A.1.2: Determination of the percentage of free thiols in Scabin variants .......... 140
   A.1.3: Derivation for equation 2 used for representing DNA binding of Scabin 141
   A.1.4: Arabidopsis thaliana growth conditions ................................................. 141
   A.1.5: Treatment of G-mito Arabidopsis thaliana seedlings with Scabin toxin .... 141
   A.1.6: Transient expression of Scabin in potato and tobacco ......................... 142
   A.1.7: Visualization of H₂O₂ accumulation ....................................................... 142
   A.1.8: Structure preparation and molecular mechanics calculations .......... 144
   A.1.9: Modelling wild-type and variant Scabin complexed with NAD⁺ ............ 144
   A.1.10: Modeling the Scabin·NAD·DNA complex ........................................... 145
   A.1.11: Spectral decomposition of the Scabin Trp emission fluorescence .......... 145
A.2: Biological studies of Scabin ............................................................................... 146
   A.2.1: Reactive oxygen response induced by transient expression of Scabin ... 147
   A.2.2: Scabin treatment of Arabidopsis thaliana seedlings ............................... 148
CHAPTER 1: INTRODUCTION
1.1: Antibiotic resistance

The occurrence of resistance in pathogenic bacteria and the growing cost of developing a drug from a lead compound to a therapeutic have reduced enthusiasm for the design of novel antibiotic treatments for infectious diseases (1). Resistance is due to many causes: poor hygiene in developing countries, uncontrolled use of antibiotics in agriculture and animal husbandry, and a lack of continuing education for doctors/public (2). Multi-drug resistant (MDR) bacteria have arisen due to the overuse of antibiotics, with ESKAPE (Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter spp.) pathogens being notable as a result of their prevalence in nosocomial (hospital-acquired) infections (3).

Resistance may be acquired through several mechanisms, such as de novo mutation and horizontal gene transfer of plasmids or integrative conjugating elements (ICEs) (1, 4). Bacterial cells can take up new genetic material in their environment, accelerating evolution. This is quite a common occurrence in the case of biofilm formation, which poses a threat for development of resistance mechanisms. In the case of Vibrio cholerae, the bacterium will lyse non-immune cells within a biofilm, using their type VI secretion system (T6SS). Genetic material from these cells will be released into the extracellular medium and can be taken up by competitor cells, which can acquire multiple genes conferring resistance (5).

If a bacterium acquires resistance through mutation or horizontal gene transfer, the variation can be metabolically costly to the cell (1). The initial mutation often leads to secondary mutations, which counters the metabolic disadvantages from the primary mutation. For example, in Escherichia coli, a mutation in the gene encoding the ribosomal subunit S12 leads to streptomycin resistance; however, this slows peptide elongation significantly. The initial mutation
is followed by a secondary, beneficial mutation in the genes encoding subunits S4 and S5, effectively returning elongation back to satisfactory levels.

As stated previously, the inevitable spread of resistance to bacterial progeny decreases the effective shelf-life of antibiotic therapeutics, coinciding with a decrease in the incentive to discover and develop novel antibiotic compounds. As an alternative, we can target the characteristics that allow bacteria to be virulent rather than inhibit their growth (1). Using anti-virulence drugs, we can effectively ‘disarm’ bacteria rather than kill, reducing the selective pressure that antibiotics place on bacterial cells (1, 4).

1.2: Anti-virulence strategy

Anti-virulence compounds – also known as anti-infectives – can be developed as inhibitors against virulence factors or mechanisms that are not vital for bacterial cell metabolism (6, 7). Anti-virulence factors often target secreted effectors and secretions systems used to deliver effectors to the host organism. Factors can be deemed a virulence mechanism based on the development of deletion mutants lacking the factor, which cannot successfully colonize a host organism (6). Anti-virulence compounds reduce the selection pressure on bacterial colonies to develop resistance. Strains that do not produce the virulence factor that can be inhibited by an anti-virulence compound would avoid the metabolic costs of producing non-beneficial metabolites (1). The main advantage of anti-virulence compounds is that they can transform a population of pathogenic bacteria to a harmless, less virulent state, without decreasing the population; by disarming the bacterial cells, the host immune system is able to overtake the population of avirulent bacteria (1). Another advantage in using anti-virulence compounds to replace antibiotics is the specificity that these compounds have towards their target (7). Treatment with antibiotics can cause secondary infections or affect surrounding tissue due to non-specific activity, often killing commensal
bacteria within the host organism (6, 7). It is also important to note that anti-virulence compounds would provide an alternative therapeutic option for patients who have serious allergies to antibiotics, such as penicillin. One of the best anti-virulence factors that is used in modern day medicine is the diphtheria antitoxin vaccine (1). The vaccine contains antibodies that specifically target the secreted diphtheria toxin, produced by Corynebacterium diphtheriae, thus preventing entrance into the host-cell (8). Anti-virulence compounds can be found in nature, such as guadinomine compounds produced by Streptomyces (9). In a study by Iwatsuki et al. (2008), guadinomine compounds could inhibit type 3 secretion system (T3SS) at sub-nanomolar concentrations and prevent hemolysis of erythrocytes by enteropathogenic Escherichia coli (EPEC) (9).

1.3: mART toxins

Mono ADP-ribosyltransferase (mART) toxins are a class of enzymes produced by pathogenic bacteria as virulence factors, and thus are potential targets for developing anti-virulence compounds against their activity (10). Pathogenic bacteria that use mART toxins to inflict toxicity can infect humans (Vibrio cholerae, Bordetella pertussis), insects (Pieris rapae) and plants (Pseudomonas syringae) (11–14). mART toxins catalyze the scission of the glycosidic linkage between nicotinamide and ADP-ribose of nicotinamide adenine dinucleotide (NAD\(^+\)). The second step proceeds with transferring the ADP-ribose moiety to a target molecule within a host cell (10, 15). Glycohydrolase (GH) activity is characteristic of most mART toxins, whereby in the absence of a transferase substrate, an OH\(^-\) molecule in solution acts as a nucleophile to accept the ADP-ribose moiety from NAD\(^+\) (10). The adduct from the transferase reaction will result in a change in function of the target molecule to either activate, repress or completely inactivate the molecule. The target molecule is most often an essential protein for eukaryotic cellular function,
such as eukaryotic elongation factor 2 (eEF2), Rho and actin (16–18). Labelling these proteins will often lead to host cell death, as the cell cannot survive without the proper function of these proteins (10). Within the target protein, the mART toxin will label a specific nucleophilic residue such as diphthamide (a post-translationally modified His residue), Arg, Asn, Cys, Thr or Gln (19–22). The protein and residue that is labelled is specific to each mART toxin.

There are several classic examples of mART toxins, such as Diphtheria toxin from *C. diphtheriae* and Cholera toxin from *V. cholerae* (23, 24). Diphtheria toxin labels a post-translationally modified histidine residue, diphthamide, on eEF2 at the site involved in binding the peptidyl (P) site of the ribosome (20, 24). This is proposed to induce steric hindrance on eEF2, preventing efficient binding on the eukaryotic host cell ribosome and halting protein synthesis (10). Cholera toxin can bind ganglioside GM1 on intestinal epithelial cells, stimulating endocytosis and modifying Gα of the heterotrimeric G-protein with ADP-ribose (23). This binding locks the G-protein in a GTP-bound state, perpetually activating the adenylyl cyclase cell-signalling pathway. This results in efflux of ions and water into the intestinal lumen, and ultimately death by extreme dehydration where patients can lose up to 1 L of water per hour (4, 10, 23).

1.4: Classification

mART toxins are divided into two unique groups: Cholera toxin (CT) –like and Diphtheria toxin (DT) –like (15). The CT group is further subdivided into C2 and C3-like subgroups. These classifications are based on the eukaryotic target macromolecule that the mART toxin uses to invoke toxicity, as well as the structural and domain organization of the apo-enzyme. The details of these subgroups will now be discussed.

1.4.1: DT-like –

The DT-like group consists of only a few members: Diphtheria toxin, as briefly described
above, ExoA toxin from *P. aeruginosa*, and Cholix toxin from *V. cholerae* (13, 25). All the members accept eEF2 as their transferase substrate, labelling a post-translationally modified histidine residue called diphthamide. This ultimately halts protein synthesis within the cell, leading to host-cell death. DT binds to heparin-binding epidermal growth factor (HB-EGF), facilitating docking with the host cell (26), while ExoA and Cholix use low density lipoprotein receptor-related protein 1 (LRP1) (13, 27). DT-like mART toxins are approximately 60 kDa, organized as a single chain AB toxin with catalytic A domain and receptor-binding B domain (Fig. 1.1A) (24).

The active-site cleft has three distinct regions that are involved with binding and catalysis of the NAD$^+$ substrate (15). First, the scaffold of the cavity is composed of a β-strand followed by a 12-residue conserved α-helix for the DT-like group (Fig. 1.1B). The β-strand and α-helix represent the upper and lower faces, respectively, of the binding pocket for the nicotinamide ring of NAD$^+$. A conserved Tyr-$X_{10}$-Tyr motif is located in this region, important for forming the scaffold of the binding pocket. Next, there is a conserved His residue that is important for NAD$^+$ binding. Located N-terminal to the His are conserved hydrophobic (Leu or Val) and aromatic (Tyr or Phe) residues. Lastly, a highly conserved Glu residue located opposite the His residue on a β-strand is critical for catalysis. Substitution of this residue leads to a several hundred-fold loss in enzymatic activity and cytotoxicity (13, 28).

1.4.2: CT-like –

The CT-like group is highly diverse, with many members falling into this category (10). They act on a variety of proteins and use one of several different receptors on cell surfaces that mediate their entry into the host cell (10). The CT-like group can be subdivided into C2-like and C3-like, depending on their transferase substrate and structural organization. CT-like members include heat-labile Enterotoxin from *E. coli* and Pertussis toxin from *B. pertussis* (11, 29). Iota
FIGURE 1.1: Structural comparison of DT- and CT-like toxins. (A) Diphtheria toxin crystal structure at 2.3 Å resolution exhibiting the catalytic (blue), receptor-binding (green) and translocation (teal) domains shown as a cartoon diagram. (B) Close-up of the Diphtheria toxin active site containing the catalytic residue, Glu148. Other important residues in the reaction mechanism, Tyr54, Tyr65, His21 are also highlighted. The important β-strand (yellow) and α-helix (red) that form the upper and lower faces of the active site are shown (PDB ID: 1SGK). (C) Cholera toxin crystal structure at 2.4 Å resolution exhibiting the catalytic (blue), receptor-binding (green) and A2 translocation (red) domains shown as a cartoon diagram. (D) Close-up of the Cholera toxin active site containing catalytic residues Glu110 and Glu112. Other important residues in the reaction mechanism, Arg7 and Ser61, are also highlighted (PDB ID: 1XTC).
toxin from *Clostridium perfringens* and Photox from *Photorhabdus luminescens* are members of the C2-like subgroup and C3bot from *C. botulinum* and HopU1 from *P. syringae* are members of the C3-like subgroup (18, 21, 30, 31). The CT-like subgroup has an AB₅ domain organization, with a 28 kDa catalytic A domain that is non-covalently bound to the receptor binding B domain that consists of a pentamer with non-covalently associated 12 kDa proteins (Fig. 1.1C) (32). For C2-like members, the enzyme exists as an AB dimer with a 50 kDa A domain and 80 kDa B domain (30). C3-like mARTs exist as a single 25 kDa monomer A chain with no innate B domain (33). From a structural point of view, the active sites of CT-like members are similar to the DT-like group, with the exception of a longer α-helix within the active site (15). For example, in Pertussis toxin, this α-helix motif is 21 residues rather than 12 in DT-like members. The catalytic signature of CT-like members is very different from DT-like (Fig. 1.1D) – the motif involved in forming the active site scaffold is a conserved Ser-Thr-Thr/Ser motif; as replacement of a catalytic His in DT-like, the CT-like members have a catalytic Arg; along with the highly conserved Glu residue that is found in both groups, a second Gln or Glu is present at the -2 position from the catalytic Glu that forms the Glu/Gln-X-Glu motif in the CT-like group (10, 15).

1.5: *Mechanisms of secretion*

There are a few ways that mART toxins can be secreted from inside the pathogenic cell into either the extracellular medium or directly into the host cell (34, 35). A common method of secretion used by Gram-negative pathogens that exploit C2 and C3-like mART toxin members is a bacterial secretion system (10). Examples of mART toxins that use this mode of secretion include SpvB from *Salmonella spp.* and HopU1 from *P. syringae* (31, 36). Molecules that are secreted by the type III mechanism are deemed ‘effectors’ because they alter physiological functions within host cells, such as inflammatory responses and intracellular trafficking (31). Based on solid-state
NMR studies of the type-III secretion system for *S. typhimurium*, the complex resembles a needle-like injector (37). The apparatus is 8 nm in diameter, with a 2.5 nm axial lumen. Secretion is continuous through the membrane, as there is no periplasmic intermediate, and thus proceeds as a single-step mechanism. Prior to the exotoxin entering the needle, chaperones bind the protein and promote a more extended, unfolded state, allowing the toxin to pass easily through the needle, travelling into the host-cell cytoplasm and refolding into an active conformation to induce toxicity. The function of the needle-like apparatus is triggered by contact of the host and pathogen cells, ultimately activating the secretion of thousands of exotoxin molecules through the needle (35). This mechanism allows for both secretion and delivery of the effector molecules into the host cell.

Toxins that contain N-terminal signal peptides are often targeted to be secreted by the cell into the extracellular medium (38). Two common pathways are the Sec- and Tat-mediated pathways. The signal peptides contain conserved consensus motifs that can be used to identify whether a protein will be secreted. In the Tat-mediated pathway, proteins are identified by the N-terminal sequence S/T-R-R-x-F-L-K; the two R’s are essential for export to the extracellular space. The Tat complex assembles at the membrane as the TatABC integral membrane proteins (Fig. 1.2); TatC recognizes the twin arginine sequence and interacts with the signal peptide to bind the protein to the TatBC complex. TatA is recruited to help form the translocation ‘pore’, allowing the virulence factor to be secreted into the extracellular medium (39). This mechanism only delivers the protein to the extracellular medium; to gain entry into the host cell, the toxin can interact with specific surface receptors. Examples include lipolysis-stimulated lipoprotein receptor (LSR; iota-toxin from *Clostridrium perfringens*), and low-density lipoprotein receptor-related protein (LRP1; ExoA toxin from *Pseudomonas aeruginosa*) (10).
FIGURE 1.2: Schematic of TatABC translocase complex interacting with a protein containing signal peptide (39).
1.6: **Examples of important mART toxins**

1.6.1: Mosquitocidal toxin from *Bacillus sphaericus* –

Mosquitocidal toxin (MTX) is a mART toxin produced by the mosquito pathogen, *B. sphaericus* (40). The mature protein is a 97 kDa, 870-residue, single-chain protein that is proteolytically activated within the host organism into the N-terminal, 27 kDa, catalytic A domain and C-terminal, 70 kDa, receptor-binding B domain (41). When the activated protein is introduced into HeLa cells, filopodia-like protrusions and rounding up of cells are observed. MTX is thought to be a unique member of the mART family because it labels many eukaryotic proteins rather than a specific target protein and it has a unique structure that contains an inhibitory peptide that must be proteolytically degraded before activation can occur (40, 41). Schirmer, Just and Aktories (2002) observed inhibition of the N-terminal catalytic domain in both the mature protein and proteolytically activated protein (41, 42). However, in the absence of the C-terminal receptor-binding domain, activity is restored for the catalytic domain. As well, a 5 kDa peptide (residues 265-285) possesses inhibitory activity when attached to the 27 kDa domain to produce a catalytically inactive 32 kDa protein. Upon proteolysis, the 5 kDa peptide is degraded and restores full activity to the catalytic domain. The catalytic domain has sequence identity with other mART toxins (32% pairwise sequence identity with Pierisin-1 toxin) and contains the conserved Glu-X-Glu motif that is consistent with the CT-like subgroup (41, 43). The B domain consists of four ricin-B-like repeat (QxW)₃ domains that are proposed to bind specific glycoproteins or glycolipids on the cell surface and facilitate entry of the catalytic domain into the host cell. The specificity for a target transferase substrate is very broad for MTX, unlike other members of the mART family. Normally, a mART toxin can only label one protein substrate; MTX can label both eukaryotic and prokaryotic substrates, including eEF2 and several proteins within HeLa cell lysates (41, 42, 44).
1.6.2: HopU1 from *Pseudomonas syringae* –

HopU1 is a type III effector produced by *Pseudomonas syringae*, which infects *Arabidopsis thaliana* and tomato plants (10, 31). The 30 kDa holoenzyme possesses the conserved catalytic signature found within mART toxins, and exists only as a catalytic A domain; this is consistent with C3-like mART toxins such as C3bot toxin from *C. botulinum*. Most gram-negative plant pathogens use the type III protein secretion system, which exploits molecular syringes that inject bacterial proteins, called effectors, directly into host cells (31). Once inside the host cell, HopU1 ADP-ribosylates glycine-rich RNA-binding protein 7 (GRP7); this protein is involved in mRNA processing and promotion of mRNA stability in plant cells (10). When HopU1 ADP-ribosylates GRP7 at the Arg49 position, a salt bridge is prevented from forming between Arg49 and the phosphate backbone of mRNA (14, 31). This results in a loss of function of GRP7, increasing the plant’s susceptibility to pathogenesis by *P. syringae*. GRP7 is involved in binding to FLS2 and EFR pattern-recognition receptors (PRR); these PRRs contribute to the plant innate immune system, and upon recognition of bacterial invasion, a pathway is triggered that leads to the production of reactive oxygen species and expression of innate immunity genes. The presence of HopU1 decreases PRR protein expression and suppresses plant innate immunity (14).

1.6.3: A new mART subfamily - Pierisin-1 from *Pieris rapae* –

Pierisin-1 is a 98 kDa mART toxin produced by the cabbage butterfly, *P. rapae*, as a defense mechanism against invading organisms (12). When transfected into human carcinoma cells, Pierisin-1 induces apoptosis (45). Within the catalytic core, Pierisin-1 has a 49% pairwise sequence identity with MTX, suggesting that these two toxins have a similar structural fold within the active site (12, 40). However, one important difference within the catalytic core of these two toxins is that the catalytic glutamate residue is a Gln-X-Glu motif in Pierisin-1 as compared to the
Glu-X-Glu motif in MTX. The Gln-X-Glu motif is characteristic of mART toxins that have Asn and Cys residue targets. But in the case of Pierisin-1, the transferase target is novel; Pierisin-1 exhibits transferase activity towards 2’-deoxyguanosine bases in DNA by labelling the exocyclic N-2 position of guanine (40, 46). Only a few mART toxins show DNA target specificity, all of which are of eukaryotic origin (47). Takamura-Enya et al. (2001) first identified the transferase target of Pierisin-1 by measuring the incorporation of $[^{32}\text{P}]$ from radioactively labelled NAD$^+$ into cell fractions containing nuclei and mitochondria. The fractions were analyzed through SDS-PAGE and subsequent autoradiography, which showed that the radioactive fractions could only be digested with DNase (46). The ADP-ribosylated DNA adducts can be repaired through the nucleotide excision repair (NER) system (48).

Similarly to MTX, Pierisin-1 is proteolytically activated and is inhibited by the presence of the B domain (40). It has been proposed that the autoinhibitory role of the B domain provides a protective mechanism for cabbage butterflies when they are producing the toxin within the cell (40). The B domain is very similar to MTX, having (QxW)$_3$ repeats characteristic of Ricin-like binding domains. Analyses of Trp variants show that substituting the Trp for any residue drastically reduces cytotoxicity, corroborating the important role the binding domain plays in pathogenesis. The binding-domain receptor consists of glycosphingolipids Gb3 and Gb4 located on sensitive human cancer cell lines (49).

1.7: Mechanism

mART toxins catalyze the scission of the C-N glycosidic linkage between nicotinamide and ADP-ribose of NAD$^+$ (17). The mechanism of catalysis depends on the transferase substrate; however, there is underlying conservation within the family (Fig. 1.3). The overall mechanism is an $S_{N1}$ type and can be simplified into two steps (17, 50). In the first part of the reaction, the
FIGURE 1.3: Schematic for the ADP-ribosylation mechanism of guanine nucleotides, modified from Jørgensen et al. (2005). The reaction mechanism proceeds as an $S_N 1$-type, with scission of the glycosidic linkage of NAD$^+$. The next step is attack by the nucleophilic guanine substrate and subsequent formation of the guanine-ADP-ribose adduct (17).
glycosidic linkage is cleaved to release nicotinamide and form an oxacarbenium intermediate that is stabilized by a hydrogen bond between the 2′-OH of N-ribose and the conserved Glu present in both DT- and CT-like members. NAD$^+$ is proposed to bind and strain the scissile glycosidic bond. Cleavage of the glycosidic bond relieves the strain and promotes the formation of the oxacarbenium intermediate (17, 50). The second part of the reaction mechanism proceeds with a nucleophilic attack at the C1 of N-ribose by the ADP-ribose acceptor molecule, which is activated by the second Glu/Gln residue (17). The end-product is an ADP-ribosylated acceptor molecule that can be either a water molecule, target protein, or DNA molecule. As stated previously, the formation of the ADP-ribose adduct within the host cell may lead to mutation or suppression/activation/inactivation of vital cellular functions, often leading to host-cell death (17, 18, 51).

1.8: Structural motifs of DNA-binding proteins

DNA-binding proteins often have conserved structural motifs that facilitate association with their target ligand (52). Well-characterized motifs include the helix-loop-helix, helix-turn-helix, helix-hairpin-helix, leucine zipper and zinc-finger motifs. The helix-hairpin-helix motif is most often found in non-sequence-specific DNA-binding proteins; nitrogens of the protein backbone within the hairpin region interact through hydrogen bonds with phosphates of DNA (52, 53). Helix-loop-helix and helix-turn-helix are both sequence specific and most often found within transcriptional regulatory enzymes (54, 55). In a few cases, the DNA-binding region of a protein may be intrinsically disordered and, upon ligand binding, will conformationally shift into a traditional motif, such as the case for the GCN4 leucine-zipper domain (56). The leucine-zipper domain of GCN4 contains a C-terminal, leucine-rich, helical domain that allows for dimerization, and an N-terminal basic region. Upon DNA binding, the basic region – which has residues with
the propensity to form helices – will transition into an ordered helical structure that fits into the major groove of DNA (56). Poly [ADP-ribose] polymerase (PARP) -1 is an enzyme that is recruited during repair of DNA damage (57). PARP-1 catalyzes a similar reaction to mART toxins, using an analogous catalytic motif (58). Despite these enzymes containing no significant structural similarities, PARP-1 is still a good model protein for studying the mechanism of mART toxins that interact with DNA rather than protein (47, 59). PARP-1 employs a zinc finger motif for recognizing single- or double-stranded DNA damage (60). Upon DNA binding, the enzyme will transfer multiple ADP-ribose groups from NAD⁺ (via auto-modification) or to histones, leading to the recruitment of DNA repair enzymes. Recently, the solution structure of PARP-1 bound to DNA containing a single-stranded break was modelled to reveal the protein:DNA interface (57). The single-strand nicked DNA adopts a twisted structure, exposing the 3’ and 5’ bases within the nicked duplex. As expected, the interacting residues are highly hydrophobic. Several valines contact the exposed bases, and a phenylalanine is involved in a π-stacking interaction with an exposed 3’ guanine (57).

1.9: Methods for structural characterization of macromolecular complexes

Relevant methods that will be used in this thesis – or in future experiments – for obtaining structural data on proteins with their macromolecular targets are discussed.

1.9.1: X-ray crystallography –

Obtaining crystals of a binary macromolecular complex requires consideration of several factors: (A) concentration ratio of each macromolecule required for a stable complex; (B) pH/temperature/buffer/salts/additives required for bringing both macromolecules out of solution, into an ordered state; (C) concentration and type of precipitant required to form a crystal (61). Often one must try hundreds of different combinations of conditions to get a crystal, if the protein
will crystallize at all. Several milligrams of protein is required for sufficient screening; crystals can form after several hours, or several months. An extra level of difficulty is added when dealing with protein:DNA complexes. DNA has several levels of diversity, and the crystallography conditions must consider length, single vs. double-stranded and nucleotide composition (61).

Once a crystal is formed, different cryoprotectants must be tested to properly flash-freeze the crystal in liquid nitrogen without allowing water to get into the crystal; proper selection of a cryoprotectant is extremely important for maintaining crystal integrity during data collection (62). The crystal can then be analyzed at an appropriate facility, and diffraction data can be collected and analyzed in programs such as Phenix, and models can be built in Coot (63, 64). Phenix is used for developing macromolecular structure solutions, which can be manipulated in Coot to assist with model building (63, 64). Obtaining an X-ray crystallographic data set has its benefits: atomic level resolution can be achieved, providing a wealth of information regarding protein structure and interactions with its cognate macromolecular-binding partner.

1.9.2: Hydrogen-deuterium exchange –

Though X-ray crystallography is the gold standard for obtaining structural data on proteins and their macromolecular complexes, hydrogen-deuterium exchange (HDX) can be used to map binding sites and study protein dynamics in solution when obtaining crystals is not a feasible option (65). This method exploits the ability of secondary amide hydrogens to exchange with deuterium in solution (66). Areas that are on the surface of the protein and exposed to solvent can exchange more readily than areas that are involved in stabilizing the core structure of the protein through strong hydrogen bonds (66, 67). To quantify the uptake of deuterium, mass spectrometry (MS) can measure the change in mass upon deuterium uptake. Global HDX-MS is used to determine the deuterium uptake of a protein in free and bound states. To obtain higher resolution, enzymatic
digestion allows for the generation of short peptides and their corresponding percent deuterium uptake can be calculated based on the mass spectra (66).

Higher spatial resolution can be obtained by coupling HDX with nuclear magnetic resonance (NMR), though protein quantity, size and stability can be problems with this analytical technique (66, 68). HDX-MS is not limited by protein or complex size, and very little sample is required at low micromolar concentrations. HDX-MS is thus a very useful tool for studying interactions between macromolecular complexes that are transient in solution and would thus be highly unlikely to crystallize. HDX is used to study folding intermediates of proteins in solution, an area of structural biology that has remained elusive due to the cooperative nature of protein folding (69). Roder and colleagues (2012) were able to pulse label cytochrome c during different stages of folding at a millisecond timescale, creating a ‘foot-print’ of each stage of folding that was then analyzed from the folded structure (69). This method offers solution-based structural information, which facilitates the study of dynamic loops and transient secondary structures often involved in macromolecular ligand binding (70). As is the case for PARP-1, local unfolding is observed of an ordered helix upon DNA binding, allowing catalysis to proceed (70).

1.9.3: Small-angle X-ray scattering –

Small-angle X-ray scattering (SAXS) is a useful tool for obtaining solution-based, low resolution data on biomolecules and their complexes (71). Dynamic changes and transient protein:protein/DNA interactions can be studied in solution, achieving 10Å resolution. A protein sample is first subjected to an incident beam and the corresponding scattered X-ray beam is measured on a detector. SAXS data can be affected by several factors, including protein aggregation and solubility, homogeneity of solution, contamination, and protein stability. Collection times often range from one to several days, thus requiring the protein(s) of interest to
be extremely stable and resistant to X-ray damage. For complex structures that would otherwise not crystallize, SAXS is a routinely used tool employed in conjunction with other structure-based studies such as NMR or electron microscopy (72, 73).

1.10: *Streptomyces scabies* and the common scab disease of potatoes

The plant pathogen *Streptomyces scabies* and its role in the common scab disease is discussed. The proposed virulence factor Scabin is introduced, and will remain the focus of this thesis.

1.10.1: *Streptomyces scabies* –

The development of the common scab disease in root and tuber vegetables is caused by the soil-dwelling, filamentous, Gram-positive bacterium, *Streptomyces scabies* (74). The disease is characterized by deep-pitted and corky lesions that can cover the entire surface of the vegetable, making it unattractive to consumers (Fig. 1.4). This leads to significant food waste and serious financial loss in the agricultural industry, especially for small-scale farms (75). The development of the common scab disease is poorly understood and no viable treatment strategy has been developed. Recently, the use of biocontrol agents has been studied to help suppress *S. scabies* growth and reduce the progression of the common scab disease (76, 77). Once a field is contaminated, it can be very costly to eradicate the *Streptomyces* organism, since it forms highly stable spores which infiltrate the soil (75).

The infection begins when germinating spores invade the plant through lesions and lenticels (cracks) in the vegetable (78). Only rapidly expanding areas of the tuber will be affected, allowing the scabs to grow as the tuber expands. Infection of roots is usually not as evident as the corky lesions observed on the surface of the vegetable; root stunting, browning and seedling death may occur.
FIGURE 1.4: *Streptomyces scabies* and the common scab. (A) Electron micrograph of *Streptomyces scabies*. (B) Potato that exhibits the characteristic symptoms of the common scab disease caused by infection of *Streptomyces scabies* (80).
Many proteins and small molecules are secreted into the external environment upon infection by *S. scabies*, as determined by Joshi *et al.* (2010). Thaxtomin A is a small molecule, plant cellulose synthesis inhibitor produced by *S. scabies* during infection; the synthesis of thaxtomin A is triggered by several plant materials, such as xylan degradation products, suberin (cell-wall associated poly-aromatic found on tubers) and cellobiose (79). Cellobiose is a product of cellulose degradation and directly activates thaxtomin synthesis through the cellulose utilization regulator, CebR. The cellobiose ABC type transporter, Ceb, is recruited to the plasma membrane to bring cellobiose into the cell and stimulates production of cellulases, like thaxtomin A, upon the presence of infection triggers (81, 82). Briefly, CebE binds cellobiose on the periphery of the membrane, transporting it into the cell through integral membrane proteins CebF and CebG; cellobiose will bind to CebR and inhibit DNA binding, leading to expression of *txtA* and subsequent production of thaxtomin A. Thaxtomin A can then allow further degradation of the plant cell wall, increasing lesion density on the infected vegetable and allowing other virulence factors to invade (79). Deletion of *cebR* leads to hyper-virulence of *S. scabies*, indicating the important role of this transporter in controlling the pathogenesis of the bacterium (79). A detailed schematic presented by Jourdan *et al.* (2016) is shown in Fig. 1.5.

Many potential virulence proteins are secreted into the host environment, including a putative secreted protein with the gene annotation *scab_27771* (38). Utilizing bioinformatics techniques (43), this protein has been identified as a potential mART toxin, which we have named Scabin. We proposed that Scabin could play a significant role in disease development of the common scab in tuber and root vegetables (59).
**FIGURE 1.5:** Schematic of the cellobiose ABC transporter, CebEFG, as presented in Jourdan *et al.*, 2016.
Scabin, a putative mono-ADP-ribosyltransferase toxin –

Full-length Scabin protein consists of a single catalytic A-domain and a 29-residue N-terminal secretion signal. Based on the data presented by Joshi et al. (2010), it is known that Scabin is secreted into the extracellular medium, probably via the Tat secretion pathway, based on prediction models using the signal peptide sequence (38). Scabin was identified using a novel bioinformatics strategy as a putative mART toxin. Scabin exhibited no cytotoxicity (data not shown) to yeast when the gene was expressed under the control of a copper inducible CUP1 promoter, suggesting that the traditional method of screening for new mART toxins may not reach all possible transferase substrates (83). Scabin possesses the common catalytic features of mART toxins, including the catalytic Arg, catalytic Q-X-E and the S-T-T motif. Scabin is closely related to the DNA-acting Pierisin-1 mART toxin, as these two toxins share large conserved regions within the catalytic core of the protein.

Research rationale and objectives

Regarding their substrates, mART toxins are proving to be a diverse family of enzymes that can modify both proteins and DNA, creating detrimental adducts that are involved in disease progression in vertebrates, insects and plants. The goal of this research is to understand the catalytic mechanism of the DNA-acting mART toxin, Scabin, via kinetic and structure-guided studies.

First, the initial biochemical characterization of Scabin is presented, which details the first crystal structure of a DNA-acting mART toxin and the identification of inhibitors against its GH activity. A thorough kinetic analysis is performed, ending with the identification of the transferase substrate, DNA. Scabin is revealed as novel member of the DNA-acting mART toxin Pierisin-like subgroup. Next, catalytic variants are explored to identify residues comprising the DNA-protein interaction.
interaction surface and involved in the transferase reaction mechanism; a substrate analog, NADH, was co-crystallized with Scabin and a model of the Scabin-DNA complex is presented. Lastly, the first experimentally determined structure data of the DNA-protein interaction for this unique mART toxin is detailed. This was acquired using hydrogen-deuterium exchange coupled with mass spectrometry (MS) to resolve local residue interactions of Scabin with DNA.

The current research on Scabin only provides a small glimpse of the wealth of knowledge that further characterization of this unique mART toxin can reveal about the newly formed Pierisin-like subgroup of toxins. To shed more light on Scabin enzymatic features, a complete analysis of the important catalytic residues within the active site must also be performed. The DNA base sequence specificity for Scabin must be elucidated, as well as the binding affinity for specific oligonucleotide sequences.
CHAPTER 2: MATERIALS AND METHODS
The recombinant form of Scabin containing an N-terminal 29-residue truncation of the leader sequence (Scabin Δ29) will herein be referred to only as Scabin. All variants of Scabin contain the N-terminal 29-residue truncation.

2.1: Expression and purification of Scabin

The Scabin gene was cloned into a pET-TEV vector containing an N-terminal His$_6$ tag with a tobacco etch virus protease cut site. Site-directed mutants (Table 2.1) were prepared using the Quikchange mutagenesis method (84). Chemically competent *Escherichia coli* BL21 λDE3 cells were transformed with plasmid and grown overnight at 37°C on LB medium containing 30 µg/mL kanamycin. Half the colonies on each plate were scraped into 50 mL LB containing kanamycin and allowed to grow to an OD of 0.6 at 37°C with shaking; 25 mL culture was inoculated into 800 mL 2xYT medium containing kanamycin. Cells were grown to an OD of 1.2 at 37°C with shaking and subsequently induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 4 h. Wild-type (WT) and variants were all expressed using the above conditions, except for Q$_{158}$A/E$_{160}$A and N$_{110}$A, which yielded better expression when induced for 16 h at 16°C. Cells were pelleted at 4000 x g and resuspended in lysis buffer containing 25 mM Tris-HCl, pH 8.2, 200 mM NaCl, 50 µg/mL 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 120 µM phenylmethanesulfonyl fluoride (PMSF), 1 mM ethylenediaminetetraacetic acid (EDTA) and 100 µg/mL DNase. Resuspended cells were lysed using an Emulsifex-C3 high-pressure homogenizer (Avestin Inc., Ottawa, Ontario) and subsequently centrifuged at 14,000 x g for 50 min at 4°C. Protein was purified from whole cell lysate using immobilized-metal-affinity chromatography. Supernatant was passed over a HiTrap IMAC HP 5 mL column (GE Healthcare) equilibrated with 5 mM imidazole in binding buffer (50 mM TAPS, pH 8.5, 500 mM NaCl). The column was washed with 25 mM imidazole in binding buffer and the protein was eluted using a gradient from 25 to
TABLE 2.1: Summary of variants developed for studying Scabin toxin.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Role</th>
<th>&quot;Stable protein&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>R77A and R77K</td>
<td>Conserved catalytic Arg in CT-like mART toxins</td>
<td>×</td>
</tr>
<tr>
<td>Q158A-X-E160A</td>
<td>Conserved catalytic Gln-X-Glu motif; involved in carrying out the catalytic reaction mechanism</td>
<td>✓</td>
</tr>
<tr>
<td>N110A</td>
<td>Conserved Asn in Pierisin-like (DNA targeting) mART toxins; proposed to bridge the binding of NAD⁺ and DNA backbone</td>
<td>✓</td>
</tr>
<tr>
<td>S117A</td>
<td>Conserved Ser in CT-like group; forms the scaffold for NAD⁺ binding</td>
<td>✓</td>
</tr>
<tr>
<td>W68A and W68Y</td>
<td>Trpₙ₋₁ variant; used for Trp fluorescence/emission analysis</td>
<td>×/×</td>
</tr>
<tr>
<td>W128A and W128Y</td>
<td>Proposed Trp involved in carrying out transferase activity and binding DNA</td>
<td>×/✓</td>
</tr>
<tr>
<td>W155A</td>
<td>Proposed Trp involved in carrying out transferase activity</td>
<td>✓</td>
</tr>
<tr>
<td>W199A</td>
<td>Trpₙ₋₁ variant; used for Trp fluorescence/emission analysis</td>
<td>✓</td>
</tr>
</tbody>
</table>

" Variants that produced a stable protein are distinguished by a ‘✓’; if no stable protein was produced, a ‘×’ was used.
250 mM imidazole. Fractions containing the protein of interest were resolved on an SDS-PAGE gel to confirm identity; purified protein was pooled and dialyzed into 50 mM Tris-HCl, pH 8.2, and 50 mM NaCl (dialysis buffer 1).

Further purification was performed via anion-exchange chromatography. Briefly, a HiTrap Q-Sepharose HP column was equilibrated with dialysis buffer 1 and the protein sample was passed over the column; bound protein was subsequently eluted with a linear gradient from 50 to 500 mM NaCl in dialysis buffer 1. Fractions that contained purified protein, as confirmed by SDS-PAGE, were pooled and concentrated initially on a bed of PEG 20,000 at 4°C to reduce mechanical stress on the protein and for optimal recovery. Once the sample was concentrated to approximately 5 mL, the remainder was brought to 1.5 mg/mL using Millipore 0.5-mL 10-kDa spin columns at 4250 ×g for 10-min intervals in a microfuge; protein yield was approximately 1 mg per litre of culture. Protein used for H-D exchange experiments was not concentrated using the PEG 20,000 method. Protein was concentrated using Millipore 15-mL 10-kDa spin columns until approximately 10 mL of solution remained; the remainder was concentrated to 1.5 mg/mL using Millipore 0.5-mL 10-kDa spin columns as previously described. Protein yield was approximately 50% less than when concentrated on PEG 20,000.

2.2: Confirmation of Scabin identity by mass spectrometry

The molecular weight of Scabin was determined using an Agilent UHD 6530 Q-TOF mass spectrometer at the Mass Spectrometry Facility of the Advanced Analysis Centre, University of Guelph. The instrument was configured with the standard electrospray ionization source and operated in positive-ion mode. Data analysis was performed using MassHunter Analysis version B.06.00 (Agilent) software. Deconvolution of the m/z spectrum was achieved using the maximum entropy algorithm in the BioConfirm software (Agilent).
2.3: Circular dichroism spectra

Circular dichroism (CD) spectra were acquired for all Scabin variants using a JASCO J-815 CD spectropolarimeter (250–195 nm scan, average of 9 spectra). The protein concentration was 0.16 mg/mL in a buffer containing 20 mM Tris-HCl, pH 8.2, and 50 mM NaF in a 1-mm path length quartz CD cuvette.

2.4: Protein crystallography

Crystal conditions for Scabin were screened using the PEG II suite (Qiagen) in 96-well screening trays containing 1.5 mg/ml Scabin. Crystal hits were observed after 24 h in a condition containing 0.1 M MES, pH 6.5, and 15% PEG 400. The crystal condition was scaled up using 18-mm hanging drop trays with 200 µL of well solution. 1.5 mg/mL Scabin (1 µL) was mixed with well solution (1 µL) on a glass cover-slide to form the crystal drop. For the Scabin with NADH crystal, the condition contained 100 mM KCl, 50 mM sodium cacodylate trihydrate, pH 6.0, 16% PEG 1000 and 0.5 mM spermine. Inhibitors or NADH were co-crystallized by pre-incubating Scabin with either PJ34, P6-E or NADH (500 µM) for 30 min. After incubation at room temperature for 24–48 h, large single crystals appeared. Crystals were then routinely washed with mother liquor containing 30% PEG 400 to act as a cryoprotectant and subsequently were flash frozen in liquid nitrogen. In the case of co-crystals containing inhibitor, cryoprotectant also contained 500 µM inhibitor. Scabin variants W128Y, W155A, S117A and N110A (with and without NADH for W128Y and N110A) were crystallized in the same condition as inhibitor co-crystals and cryoprotected by washing step-wise in 12, 13.5 and 15% glycerol prior to flash freezing. X-ray diffraction data were collected at the Canadian Light Source in the Canadian Macromolecular Crystallography Facility (beamline 08ID-1).
2.4.1: Scabin-apo structure –

A total of 266 frames of diffraction images were collected with a 0.75° oscillation range. Images were processed and scaled with XDS in the C2 space group (85). The Scabin-apo structure was solved using molecular replacement in MOLREP (86). For MOLREP, the template was obtained by generating a model of Scabin-apo from a homologous *Bacillus sphaericus* mosquitocidal toxin structure (Protein Data Bank code 2CB4) (87) using MODELLER (88). The model was further optimized by truncating loops. The molecular replacement solution was subjected to iterative model building and refinement using COOT (63) and Refmac5 in the CCP4 Suite (89). The final model was refined to a resolution of 1.4 Å with $R_{\text{work}}/R_{\text{free}}$ (%) 16.3/19.7 (Table 3.3); the Scabin-apo crystal structure has been deposited in the Protein Data Bank database with code 5DAZ.

2.4.2: Scabin PJ34 and P6-E structures –

The data collected were processed in XDS (85). Molecular replacement on the crystal data set was conducted using Phenix (64) with the Scabin-apo structure as the model. Iterative cycles of model building in COOT (63) and refinement in Phenix were performed. The Scabin PJ34, and Scabin P6-E structures have been deposited in the Protein Data Bank database, with codes 5EWK, and 5EWY, respectively. Table 3.3 summarizes diffraction data and refinement statistics.

2.4.3: Scabin NADH and variant structures –

The data collected were processed in XDS (85). Molecular replacement on the crystal data set was conducted using Phenix (64) with the Scabin-apo structure as the model. Iterative cycles of model building in COOT (63) and refinement in Phenix were performed. Scabin-NADH, W128Y and W155A structures have been deposited in the Protein Data Bank database, with codes 5TLB, 5UVO and 5UVQ, respectively. Table 4.1 summarizes diffraction data and refinement

2.5: Hydrogen-deuterium exchange of Scbin

Microfluidic chip fabrication is shown in Appendix I; original fabrication was presented by Rob et al. (90). The HDX reaction took place inside the time-resolved electrospray ionization (TRESI) mixer, where either 35 µM Scbin, Scbin:140 µM DNA complex or Scbin:250 µM NADH in 500 mM ammonium acetate (C₂H₃O₂NH₄), pH 8.2 was flowed through the inner glass capillary (flow rate 1 µL/min) and deuterium through the outer metal capillary (flow rate 1 µL/min) (see Fig. 2.1). Total flow rate into the TRESI mixer (2 µL/min), along with intercapillary space and with the glass capillary pulled back 5 mm, was used to calculate the mixing volume of 124 nL and a reaction time of 4.14 s. The instrument was operated with a source voltage of 3000 V in positive-ion mode, and samples were scanned over a range of m/z 400–1500. Apo-Scbin HDX was performed using four biological replicates with each having four technical replicates, in which all data were then averaged; the ligand HDX experiments were performed using 2 (NADH) or 3 (DNA) biological replicates, with each having four technical replicates. A pepsin spectrum was collected to eliminate peptides within the Scbin digestion profiles that correspond to self-digestion of pepsin. A Scbin digestion profile was collected in the absence of D₂O to identify peptides. In the cases where peptides were overlapping, spectra were collected using ion-mobility spectrometry to separate peptides based on their size, shape and charge.

2.6: Hydrogen-deuterium exchange data analysis

Digested peptides were identified using mMass (91) and corresponding masses were analyzed using the FindPept tool on the ExPASy Proteomics server (mass tolerance ± 0.5 Da) to determine peptide sequences. Peptides that could not be confirmed due to multiple ‘high-certainty’ hits (ΔM ± 0.5 Da) in FindPept were subjected to MS/MS sequencing to confirm identification.
FIGURE 2.1: TRESI schematic. Glass capillaries (gold) were used for protein, D$_2$O and acetic acid. Pull back distance was 5 mm, corresponding to the distance between the sealed end of the glass capillary and entrance to T-mixer where acid quenching of reaction occurs. This distance created a reaction time of 4.14 s and mixing volume of 124 nL.
The experimental deuterium uptake was calculated using an in-house-developed FORTRAN software (Wilson Laboratory) that models the change in the isotopic distribution to determine the percent uptake of deuterium for each peptide of interest (92).

2.7: \(\beta\)-NAD and inhibitor binding

The dissociation constant, \(K_D\), for \(\beta\)-NAD and inhibitors with Scabin was determined by measuring the quenching of intrinsic tryptophan fluorescence using a Cary Eclipse fluorescence spectrophotometer (Varian Instruments, Mississauga, Canada). An excitation wavelength of 295 nm, emission wavelength of 340 nm, and band passes of 5 nm were used. In quartz UV cuvettes (0.5 × 0.5 cm), 1.25 \(\mu\)M Scabin in 25 mM Tris-HCl, pH 8.2, and 200 mM NaCl was titrated with \(\beta\)-NAD (0–1000 \(\mu\)M) or inhibitor (0–1000 M), and fluorescence intensity was monitored at 340 nm. Assays were performed in triplicate, and data were corrected to account for the increase in volume upon the addition of \(\beta\)-NAD or inhibitor. A blank titration with N-acetyl tryptophanamide (NATA) was performed to correct for inner filter effects. Data were analyzed using OriginPro version 8 software (OriginLab, Northampton, MA) to determine the dissociation constants; either a one-site binding or a two-site binding (only used for S117A variant) model in OriginPro was used to fit the data.

2.8: Glycohydrolase activity

GH assays were performed on a Cary Eclipse fluorescence spectrophotometer using an excitation wavelength = 305 nm, emission wavelength = 405 nm, and band pass = 5 nm. Scabin (WT 50 nM; Q158A-X-E160A 2.3 \(\mu\)M; S117A 0.5 \(\mu\)M; N110A 1 \(\mu\)M; W128Y 250 nM; W155A 50 nM; W199A 50 nM) was incubated with increasing concentrations of \(\varepsilon\)-NAD\(^{+}\) (0-500 \(\mu\)M) in GH buffer containing 20 mM Tris-HCl, pH 7.9, and 50 mM NaCl. Triplicate reactions were monitored for 10-min intervals, and the initial slope of each reaction was recorded. An \(\varepsilon\)-AMP
standard curve was derived to convert fluorescent units/min to [ε-ADP ribose] formed/min. A Michaelis-Menten curve was derived by fitting the data to the hyperbolic model in OriginPro version 8 software.

2.9: ADP-ribosyltransferase reaction

ADP-ribosyltransferase assays with deoxyguanosine (dG) were performed as described under “GH Activity” with a few modifications. ε-NAD was held at a concentration of 250 µM and was mixed with Scabin (WT 10 nM; Q158A-X-E160A 1.5 µM; S117A 0.1 µM; N110A 200 nM; W128Y 50 nM; W199A 10 nM) in GH buffer containing 2% dimethyl sulfoxide and various concentrations of dG (0–1250 µM). The data were fit to the sigmoidal dose response or biphasic model in OriginPro version 8. Activity toward other small nucleotides was also tested using the method described above, with the exception of holding the small nucleotide substrate at a constant concentration = 1 mM. Substrates tested include GDP, cyclic guanosine monophosphate (cGMP), and deoxyinosine (dI). Scabin activity toward genomic DNA was also tested using the method described above. The genomic DNA was extracted from various samples, as described previously (93), and was partially fragmented (nicked) with 1 unit of DNaseI/g of DNA (375 ng/µl, final concentration) in Tris-EDTA buffer containing 5 mM MgCl₂ for 10 min at room temperature.

2.10: Mass spectrometry of mART reaction products

Liquid chromatography-mass spectrometry (LC-MS) analyses were performed in the Mass Spectrometry Facility, Advanced Analysis Centre, University of Guelph. Reactions containing 1 mM nucleotide (GDP or cGMP) or 100 µM DNA (single-stranded pentamer of dG or double-stranded DNA with forward sequence 5’-ATTAAGTATT-3’ and reverse sequence 5’ TAATTCATAA-3’), 250 µM β-NAD, and 50 nM Scabin in GH buffer were run overnight at room temperature with gentle shaking. Samples were diluted 1:100 in water before injection into a
Dionex UHPLC UltiMate 3000 liquid chromatograph interfaced to an amaZon SL ion trap mass spectrometer (Bruker Daltonics, Billerica, MA). A Luna C18 column (5 µm particle size, 150 × 2 mm; Phenomenex) was used for chromatographic separation. The initial mobile phase conditions were 100% water (0.1% formic acid) isocratic for 5 min followed by a gradient to 60% acetonitrile (0.1% formic acid) in 20 min and then a linear gradient to 100% acetonitrile in 10 min. The flow rate was maintained at 0.4 ml/min. The mass spectrometer electrospray capillary voltage was maintained at 4.5 kV, and the drying gas temperature was 280 °C with a flow rate of 10 liters/min. Nebulizer pressure was 40 p.s.i. Nitrogen was used as both nebulizing and drying gas, and helium was used as collision gas at 60 p.s.i. The mass spectrometer was set on enhanced resolution positive-ion auto-MS/MS mode and scanned across 50–1500 m/z. The smart parameter setting was used to automatically optimize the trap drive level for selected precursor ions. The instrument was externally calibrated with the ESI TuneMix (Agilent).

2.11: Inhibitor assays

Testing of potential inhibitors against Scabin GH activity was performed using compounds designed to act competitively with β-NAD. P6 series inhibitors were obtained from Sinova Inc. (Bethesda, MD). Initial testing was performed on a Cary Eclipse fluorescence spectrophotometer using the same settings as described under “GH Activity.” Initial testing was performed in a reaction vessel containing 50 nM Scabin, 250 µM β-NAD, 100 µM inhibitor, and GH buffer. Because some of the inhibitors were dissolved in DMSO, a control (2% DMSO) was performed to correct for the effect of DMSO on enzyme activity. Triplicate reactions were monitored for 10-min intervals, and the initial slope of each reaction was normalized to control for the effect of DMSO.
2.12: IC\textsubscript{50} and K\textsubscript{i} determination

Inhibition assays were used to determine the IC\textsubscript{50} values for selected inhibitors, and the experiments were performed on a Cary Eclipse fluorescence spectrophotometer with an excitation wavelength = 305 nm, emission wavelength = 405 nm, and band passes = 5 nm. Scabin (50 nM) was mixed with 250 µM ε-NAD, various concentrations of inhibitors, and GH buffer. Triplicate reactions were monitored for 10-min intervals, and the initial slope of each reaction was normalized to a 2% DMSO control. An ε-AMP standard curve was used to convert fluorescence units/min to [ε-ADP ribose] formed/min. The data were fit using the dose-response function in OriginPro version 8 to determine the corresponding IC\textsubscript{50} value. Experimental IC\textsubscript{50} values were used to calculate \( K_i \) values using the Cheng-Prusoff equation (94),

\[
K_i = \frac{IC_{50}}{1 + [S] / K_M},
\]

where [S] is the concentration of ε-NAD and \( K_M \) is for the ε-NAD substrate. The \( K_i \) was determined experimentally for PJ34 from a Lineweaver-Burk plot of the inhibition data. This was developed by incubating 50 nM Scabin in GH buffer and varying the concentration of ε-NAD (12.5, 25, 50, and 100 µM) but holding the concentration of PJ34 at either 0, 6, 12, or 24 µM. Data were collected on a Cary Eclipse fluorescence spectrophotometer as described above.

2.13: DNA binding of Scabin with cyanine-3 tagged oligomer

Synthetic dsDNA oligomers (oligomer 1: 5’-GGAAGAGAGAGAGAAAGAGAG-3’; oligomer 2: 5’-CTCTCTTTCTCTCTCTCTTTCC-3’; oligomer 3: 5’-CCTCTCTTTCTCTCTCTC TCTCC-3’) with a 5’ cyanine-3 tag on oligomer 1 were purchased from Sigma-Aldrich. Oligomers were mixed in equal molar amounts (oligomer 1 with 2 as a blunt-ended substrate; 1 with 3 for a one-base overhang on either primer 5’-termini) and annealed by heating to 90°C followed by cooling to 20°C at a rate of 1°C/min in a Techne TC-512 PCR instrument (Burlington, NJ). In an ultra-micro quartz cuvette (3 mm x 3 mm), 5 µM of Cy3-dsDNA in 25 mM Tris-HCl, pH 8.2, 100
mM NaCl was titrated with increasing concentrations of either WT, N110A, W128Y or W155A Scabin protein.

Data was collected in a ‘T-format’ detection using a PTI-Alphascan-2 spectrofluorometer (Photon Technologies Inc., South Brunswick, NJ) equipped with a thermostated cell holder and temperature was held at 22°C. T-format detection simultaneously compares the fluorescence intensities of both the vertical \( I_{VV} \) and horizontal \( I_{VH} \) polarized emitted light, while exciting with vertically polarized light. Fluorescence anisotropy is then calculated using equation 1:

\[
\frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}}
\]

where the ‘G’ instrumental factor is measured as \( I_{HV}/I_{HH} \), the vertical \( I_{HV} \) and horizontal \( I_{HH} \) polarized emitted light when excited with horizontally polarized light. Changes in anisotropy \( r \) of the cyanine-3 fluorophore were measured for 20 s intervals, using excitation and emission wavelengths of 550 and 570 nm, respectively. Excitation and emission band-passes were set to 4 nm.

For all titrations, the ratio of the quantum yield of Cy3-dsDNA in protein-bound \( Q_b \) and free \( Q_f \) states was calculated by exciting at 550 nm and scanning emission from 560 to 620 nm in 1 nm increments. Data were corrected for increasing volumes upon addition of protein. The concentration of bound DNA \( [DNA]_b \) at each point of the titration was calculated using equation 2:

\[
[DNA]_b = \frac{([E]_t + [DNA]_t + K_D) - \sqrt{([E]_t + [DNA]_t + K_D)^2 - 4[E]_t[DNA]_t}}{2}
\]

where \([E]_t\) is the total concentration of enzyme, \([DNA]_t\) is the total concentration of DNA and \(K_D\) is the binding constant for Cy3-dsDNA. A full derivation is presented in Appendix I. Anisotropy values were corrected for changes in the quantum yield of the Cy3 fluorophore upon addition of
protein using equation 3:

\[ r = \frac{Q_f ([DNA]_t - [DNA]_b)}{F_t} r_f + \frac{Q_b [DNA]_b}{F_t} r_b \]

where \( r_f \) and \( r_b \) are defined as the anisotropy of free and bound states of Cy3-dsDNA. \( F_t \) is defined as the total fluorescence at each point, defined by equation 4:

\[ F_t = Q_f ([DNA]_t - [DNA]_b) + Q_b [DNA]_b \]

2.14: Absorbance and emission of single Trp variants

Absorbance of single Trp variants (W128Y, W155A and W199A) was performed in a Cary50 UV-Vis spectrophotometer (Varian Instruments, Mississauga, Canada); 40 µM NATA was used as a control. Each variant was run in conjunction with WT Scabin at 3.3 µM in 25 mM Tris, pH 8.2, and 100 mM NaCl in an absorbance ultra-micro quartz cuvette (3 mm x 3 mm).

Emission spectra of single Trp variants (W128Y, W155A and W199A) were collected on a PTI-Alphascan-2 spectrofluorometer with excitation wavelength of 295 nm and scanning emission from 310 to 450 nm in 1 nm increments; band-passes were set to 4 nm. Forty µM NATA was used as a control. Each variant was run in conjunction with WT Scabin at 3.3 µM in 25 mM Tris, pH 8.2, and 100 mM NaCl in a fluorescence ultra-micro quartz cuvette (3 x 3 mm). Fluorescence quantum yield measurements were calculated using equation 5:

\[ Q = Q_r \frac{I_r OD_r}{I_r OD} \]

where \( Q_r \) and \( Q \) represent the quantum yields of reference (NATA, \( Q_r = 0.14 \)) and sample, respectively; \( I_r \) and \( I \) are the integrals of emission scan (310 – 450 nm) for reference and sample, respectively; OD \(_r\) and OD are the optical densities for reference and sample, respectively.
CHAPTER 3: INITIAL CHARACTERIZATION OF SCABIN, A MONO-ADP RIBOSYLTRANSFERASE PRODUCED BY *STREPTOMYCES SCABIES*
Abstract

A bioinformatics strategy was used to identify Scabin, a novel DNA-targeting enzyme from the plant pathogen 87.22 strain of *Streptomyces scabies*. This putative toxin has nearly 40% pairwise sequence identity with Pierisin toxins, and possesses the characteristic RSQXE motif found in the mART toxin family. Scabin was purified to homogeneity as a 22-kDa single-domain enzyme and was shown to possess high NAD\(^+\)-glycohydrolase \((K_M = 68 \ \mu\text{M}; \ k_{\text{cat}} = 94 \ \text{min}^{-1})\) activity. The target substrate was identified as deoxyguanosine, with Scabin exhibiting sigmoidal enzyme kinetics \((K_{0.5} (\text{deoxyguanosine}) = 302 \ \mu\text{M}; \ k_{\text{cat}} = 83 \ \text{min}^{-1})\) for transferase activity. Mass spectrometry analysis revealed that Scabin labels the exocyclic amino group on guanine bases in either single-stranded or double-stranded DNA. Several small molecule inhibitors were identified, with the most potent compounds found to inhibit activity at \(K_i\) values ranging from 3 to 24 \(\mu\text{M}\). PJ34, a well-known inhibitor of poly [ADP-ribose] polymerase, was shown to be the most potent inhibitor of Scabin. The crystal structure of Scabin represents the first 3-dimensional model of a DNA-targeting mono-ADP-ribosyltransferase enzyme; the structures of the apo-form (1.45 Å) and in complex with two inhibitors (P6-E, 1.4 Å; PJ34, 1.6 Å) were solved.

*This work has been published unless otherwise stated.*

3.1: Results and Discussion

3.1.1: Identification and expression of Scabin –

Scabin – scab_27771 gene – is found in the *S. scabies* pathogenic strain 87.22. The full-length protein consists of a single catalytic mART domain and a 29-residue N-terminal secretion signal. As shown from the multiple-sequence alignment depicted in Fig. 3.1A, Scabin shares common catalytic features with other mART toxins. This includes the catalytic Arg that is essential for NAD\(^+\) binding to the active site, the catalytic Gln-X-Glu motif, which is essential for the reaction mechanism, and the Ser-Thr-Thr motif that forms the scaffold of the NAD\(^+\) binding pocket (15) (Fig. 3.1A). The multiple-sequence alignment clearly shows that Scabin is closely related to Pierisin mART toxins, as evident by the large conserved regions shared by Scabin and the Pierisin members (Fig. 3.1A). A percentage identity matrix of core sequences from several key members of the mART toxin family, including Scabin, revealed that Scabin shares nearly 40% sequence identity with almost all Pierisin proteins (Fig. 3.1B). The Pierisin group of mART toxins is distinguished by its unique target specificity, where ADP-ribose is transferred to a guanine base in DNA, leading to host-cell death via apoptotic pathways (12, 45) (Fig. 3.1B).

Before purification, an N-terminal 29-residue truncation of the Scabin gene was cloned for expression into *E. coli* BL21 λDE3 cells. The truncated sequence corresponds to the signal peptide of Scabin, which facilitates its secretion from *S. scabies* via the twin arginine transport (Tat) pathway, as proposed by Joshi and colleagues (38). Scabin was shown to be secreted into the extracellular medium and was proposed to use the Tat secretion pathway based on prediction models utilizing the signal peptide sequence. The Scabin signal peptide sequence contains a twin Arg motif, which is indicative of Tat-mediated secretion (38). Scabin did not express well with the N-terminal signal sequence, so removal of this region for further experiments was required for
FIGURE 3.1: Multiple-sequence alignment of Scabin with various Pierisin-like mART toxins. (A) Sequence alignment of Scabin with Pierisin-like toxins; alignment was made using the T-Coffee Web server and ESPript to generate the alignment figure (95). Key catalytic regions are highlighted. Identical residues are highlighted in red, and similar residues are printed in red type. (B) Identity matrix showing the amino acid identity between the 100 core catalytic residues of the known ExoS-like, C2-like toxins and Vis. Colouring scheme is as follows: highly diverse sequences, salmon; a large amount of conservation, light green; an intermediate level of conservation between sequences, yellow. The identity matrix was generated using ClustalX2 (96) and coloured using Microsoft Excel. (C) Purification of Scabin from E. coli lysate. SDS-polyacrylamide gels showing the protein banding pattern for crude lysate (lane 1), immobilized metal affinity chromatography purification (lane 2), and FPLC ion exchange chromatography (lane 3). The arrow indicates the position of the Scabin protein. (D) MALDI-TOF mass spectrometry of purified Scabin protein showing a single peak at 21,691.9 Da, corresponding to the expected mass of recombinant Scabin (59).
adequate expression. Scabin was purified from whole-cell lysate via IMAC and a subsequent anion-exchange step. The purity level was assessed by inspection of the corresponding SDS-polyacrylamide gels after each purification step (Fig. 3.1C). Scabin protein yield of 1 mg/liter culture was obtained after purification to near homogeneity. Scabin was positively identified using MALDI-TOF mass spectrometry and gave the expected mass of 21,961.9 Da (Fig. 3.1D).

3.1.2: Folded integrity of Scabin WT and variant toxins –

To assess the folded integrity of Scabin and the catalytic variant Q158A/E160A, both DSF and CD spectroscopy were used (Fig. 3.2, A and B). Differential-scanning fluorimetry revealed a smooth transition between the folded and unfolded states, resulting in a melting temperature for Scabin and the catalytic variant Q158A/E160A of 46 ± 0.9 and 48 ± 0.4 °C, respectively (Fig. 3.2A). This suggested that the catalytic variant protein was slightly more stable than the WT protein. Scabin exhibited a less than standard CD spectrum for an α/β protein with a relatively higher helical composition than β-strand content (20.5% α-helix and 11.5% β-strand, respectively). The strong right-handed twist of the anti-parallel β-sheet core is most likely the reason for the strange spectrum, as well as the high content of loop regions of Scabin (97). There were no significant differences in secondary structure between the WT and variant proteins based on their CD spectra (Fig. 3.2B). This indicates that Scabin was properly folded into a stable and active conformation and that replacement of Gln158 and Glu160 residues did not significantly change the folded integrity of the enzyme (Fig. 3.2B).

3.1.3: NAD⁺ GH activity and NAD⁺ binding of Scabin –

GH activity is characteristic of most mART toxins, whereby in the absence of a transferase substrate, an OH⁻ molecule from solution acts as a nucleophile to accept the ADP-ribose group from NAD⁺ (10). The GH activity was characterized for Scabin in the absence of transferase
FIGURE 3.2: Folded stability and enzymatic activity of Scabin. (A) Folded stability ($T_m$) of WT Scabin (thick line) and Scabin Q158A/E160A variant (thin dotted line) as measured by the SYPRO Orange thermal shift assay (derivative of the raw data traces). The traces are representative scans of three replicates for each sample, and the apex of the minima shows the position of the $T_m$, where the protein is half-unfolded. (B) CD spectra of Scabin WT (thick line) and Q158A/E160A variant (thin dotted line) in 20 mM Tris-HCl, 50 mM NaF, pH 8.2, buffer. The concentrations of the proteins were both at 0.16 mg/mL, and each spectrum is the average of nine independent spectra. (C) $\beta$-NAD$^+$ binding by Scabin. The binding isotherm for $\beta$-NAD$^+$ with 1.25 $\mu$M Scabin was determined by quenching of the intrinsic protein fluorescence. The raw fluorescence quenching data were converted to relative values and are plotted against the $\beta$-NAD$^+$ concentration. The excitation was 295 nm, and the emission was 340 nm with excitation and emission band passes at 5 nm in 25 mM Tris-HCl, 50 mM NaCl, pH 8.2, buffer. (D) GH activity of Scabin WT showing the hydrolysis rate of the NAD$^+$ substrate by the protein (in 60 $\mu$L: Scabin, 490 nM; $\beta$-NAD$^+$, 0–450 $\mu$M. Error bars, S.D. (E) ADP-ribosyltransferase activity of WT Scabin. $\beta$-NAD$^+$ was held at a concentration of 250 $\mu$M and was mixed with 10 nM Scabin and buffer containing 1% dimethyl sulfoxide and various concentrations of dG (0–1250 $\mu$M). Error bars, S.D. (F) Inhibition plot of Scabin GH activity. Shown are Lineweaver-Burk plots for Scabin in the presence of various concentrations of inhibitor PJ34. The GH activity of Scabin was measured with 0 (filled circles), 6 $\mu$M (filled squares), 12 $\mu$M (filled triangles), and 24 $\mu$M (open circles). $V_o$ indicates initial velocity in $\mu$M/min (59). Error bars, S.D.
substrate using a well-established fluorescence-based assay (25). Scabin exhibited Michaelis-Menten behavior, and kinetic parameters (Table 3.1) were determined as $K_M = 68 \pm 3 \mu M$ and $k_{\text{cat}} = 94 \pm 2 \text{ min}^{-1}$ (Fig. 3.2D). This is consistent with other mART toxins, such as Certhrax toxin, which has a $K_M$ of $42 \pm 11 \mu M$ and a $k_{\text{cat}}$ of $8.7 \pm 1.0 \text{ min}^{-1}$ (51). The catalytic efficiency of Scabin ($k_{\text{cat}}/K_M = 1.4 \times 10^6 \text{ M}^{-1} \cdot \text{min}^{-1}$) reveals that it is a highly active enzyme (GH activity) when compared to other mART toxins within CT-like group ($k_{\text{cat}}/K_M$ for C3larvin = $11 \text{ M}^{-1} \cdot \text{min}^{-1}$ (16); MTX = $1.5 \times 10^5 \text{ M}^{-1} \cdot \text{min}^{-1}$ (98); C3cer = $2.1 \times 10^5 \text{ M}^{-1} \cdot \text{min}^{-1}$ (51)). Potential reasons for this will be explored in chapter 4.

An active site variant of Scabin was prepared that involved replacing the QXE catalytic signature with AXA, which has been shown previously to render mART toxins catalytically ineffective (16, 51). The catalytic Q158A/E160A variant was significantly less active, resulting in a $K_M = 53 \pm 9 \mu M$ (unpublished) and $k_{\text{cat}}$ of $0.31 \pm 0.02 \text{ min}^{-1}$ (Table 3.1). The Scabin Q158A/E160A variant exhibits the classic effect on enzyme activity observed for other mART toxins, demonstrating a 300-fold reduction in GH activity and efficiency decreasing by $10^3$-fold ($k_{\text{cat}}/K_M = 5.8 \times 10^3 \text{ M}^{-1} \cdot \text{min}^{-1}$).

The dissociation constant, $K_D$, of Scabin for $\text{NAD}^+$ was determined by exploiting the intrinsic fluorescence of tryptophan residues that are located near the binding pocket (Scabin has 4 Trp residues). A more detailed structural and biophysical analysis of these Trp residues will be presented in chapter 4. Scabin was titrated with $\text{NAD}^+$ and the $K_D$ was determined to be $70 \pm 3 \mu M$ (Table 3.1 and Fig. 3.2C). This is consistent with other mART toxins, such as Photox toxin, which has a $K_D$ for $\text{NAD}^+$ of $11 \pm 0.3 \mu M$ (18). The binding constant of the Q158A/E160A catalytic variant for $\text{NAD}^+$ was determined to be $86 \pm 7 \mu M$, which is comparable with the $\text{NAD}^+$ affinity of WT Scabin (Table 3.1). This suggests that the catalytic QXE motif in the enzyme is involved
**TABLE 3.1:** Kinetic parameters and NAD\(^+\) binding affinity of Scabin and Q158A/E160A variant.

<table>
<thead>
<tr>
<th>Parameter (^a)</th>
<th>WT</th>
<th>Q158A/E160A</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_{M(GH)}) (µM)</td>
<td>68 ± 3</td>
<td>53 ± 9</td>
</tr>
<tr>
<td>(k_{cat(GH)}) (min(^{-1}))</td>
<td>94 ± 2</td>
<td>0.3 ± 0.02</td>
</tr>
<tr>
<td>(K_{D(NAD^+)}) (µM) (^b)</td>
<td>70 ± 3</td>
<td>86 ± 7</td>
</tr>
<tr>
<td>(K_{0.5(dG)}) (µM)</td>
<td>302 ± 12</td>
<td>101 ± 12 (^c)</td>
</tr>
<tr>
<td>(k_{cat(dG)}) (min(^{-1}))</td>
<td>83 ± 5</td>
<td>0.5 ± 0.04</td>
</tr>
<tr>
<td>(k_{cat}/K_{M(GH)}) (M(^{-1})·min(^{-1}))</td>
<td>1.4 \times 10(^6)</td>
<td>5.8 \times 10(^3)</td>
</tr>
<tr>
<td>(k_{cat}/K_{M(dG)}) (M(^{-1})·min(^{-1}))</td>
<td>2.6 \times 10(^5)</td>
<td>5.0 \times 10(^3) (^d)</td>
</tr>
</tbody>
</table>

\(^a\) Kinetic parameters were obtained as described under “Experimental Procedures” and represent the mean ± S.D. All results represent triplicate measurements for at least three separate experiments.

\(^b\) The \(K_D\) value for the NAD\(^+\) substrate represents the binding affinity of NAD\(^+\) for the active site of Scabin and was measured by the Trp quenching caused by NAD\(^+\) docking within the active site of the enzyme (see “Materials and Methods”). All results represent triplicate measurements for three separate experiments, mean ± S.D.

\(^c\) Q158A/E160A transferase activity was best modelled as a biphasic sigmoidal curve; the first constant represents the \(K_{0.5}\) of the higher affinity transition, second constant is the lower affinity transition.

\(^d\) Q158A/E160A transferase activity was best modelled as a biphasic sigmoidal curve; the first constant represents the catalytic efficiency of the higher affinity transition; second constant is the lower affinity transition. Table content was altered from Lyons *et al.*, 2016.
in the transferase mechanism and that it does not play a role in substrate binding, consistent with other mART toxins (15).

3.1.4: Scabin acts on substrates containing guanine –

mART activity proceeds when a transferase substrate is present to act as a nucleophile in the reaction, replacing the endogenous OH⁻ nucleophile from water. Using the fluorescence-based assay for measuring GH activity, mART activity was monitored for several potential substrates. Previous studies involved testing whole cell lysates from potato, yeast, and mammalian cells for Scabin transferase activity against target proteins. However, there was no activity or labelling of any proteins from these lysates (data not shown). As demonstrated previously by Wakabayashi and co-workers (46), Pierisin-1 labels guanine-containing substrates, such as small nucleotides or DNA. Consequently, potential substrates that were tested for Scabin activity include GDP (2’-deoxy) and cGMP (2’-deoxy) for small nucleotides as well as a single-stranded oligonucleotide consisting of 5 dG bases and a double-stranded oligonucleotide containing a single guanine base at the fifth position (from the 5’ end) of the forward strand. A sharp increase in activity was observed for all substrates as compared with the baseline GH activity. To provide solid evidence that Scabin is indeed labelling these substrates, mass spectrometry was performed on reactions that were allowed to run overnight. This showed that Scabin could label GDP and cGMP with an ADP-ribose moiety, giving products with masses of 983.0 and 885.0 Da, respectively (Fig. 3.3A,B). Scabin was also able to label the single-stranded 5-nucleotide poly(dG) substrate with one, two, and three ADP-ribose moieties, as shown by the masses of 2123.5, 2834.3, and 3205.6 Da, respectively (Fig. 3.3C).

In addition, Scabin labelled the DNA strand containing a single deoxyguanosine base with an ADP-ribose group, yielding a total mass of 3579.3 Da (Fig. 3.3D). Importantly, Scabin did not
FIGURE 3.3: Mass spectrometry of Scabin with oligonucleotide substrates. (A) product ion spectra (singly charged, positive mode) after LC separation of the reaction products from the incubation of Scabin with 0.5 mM GDP. (B) product ion spectra (singly charged, positive mode) after LC separation of the reaction products from the incubation of Scabin with 0.5 mM cGMP. (C) product ion spectra (singly charged, positive mode) after LC separation of the reaction products from the incubation of Scabin with annealed poly(5)-deoxyguanidine/deoxycytidine oligonucleotide. Peaks corresponding to unlabelled (1582.5 Da), singly labelled (2123.5 Da), doubly labelled (2834.4 Da), and triply labelled (3205.6 Da) oligonucleotide were clearly resolved. (D) product ion spectra (singly charged, positive mode) after LC separation of the reaction products from the reaction of Scabin with annealed oligonucleotide with forward sequence 5’-ATTAGTATT-3’ (10-dG) and reverse 5’-TAATTCAAA-3’ (10-dC). Peaks corresponding to unlabelled (3038.6 Da) and singly labelled (3579.3 Da) oligonucleotide are shown. (E) histogram showing the relative activity of Scabin against the following substrates: none (baseline GH activity only), cGMP, GDP, and 2-deoxyinosine-5-monophosphate at 1 mM concentrations in GH buffer. (F) histogram showing the relative transferase activity of Scabin against genomic DNA from the following organisms: none (GH background activity), S. scabies, P. aeruginosa, and Solanum tuberosum (potato) (59).
label oligonucleotides that lacked guanine bases, as seen by the lack of 10-dC+ADPr peak in Fig. 3.3D.

Pierisin-1 was previously shown to label the exocyclic N2 group on guanine bases of DNA substrates (46). To test whether this was true for Scabin, dI, a nucleoside that is identical to dG with the exception that it is missing the exocyclic N2 substituent, was tested for activity with Scabin. Scabin exhibited no significant activity toward dI above the baseline GH as compared with cGMP and GDP substrates (Fig. 3.3E), demonstrating that Scabin may label the same position on deoxyguanosine as Pierisin-1 (46).

Further studies were performed with dG to determine kinetic parameters for this substrate. Notably, Scabin exhibited sigmoidal kinetic behavior in the presence of the dG substrate, with \( K_{0.5} = 302 \pm 12 \, \text{µM} \) and \( k_{\text{cat}} = 83 \pm 1 \, \text{min}^{-1} \) (Fig. 3.2E and Table 3.1). Concentrations of dG required to reach saturation are relatively high, and thus the possibility of dG forming ordered structures in aqueous solution cannot be ignored. However, as reported by Ababneh and colleagues, purine nucleosides form ‘stacked’ structures in aqueous solution at concentrations > 0.01 M; for kinetic experiments carried out with dG, concentrations were significantly less than the reported threshold. However, the addition of protein in solution may affect this threshold (99). Possible reasons for Scabin exhibiting monomeric kinetic cooperativity could include two nucleotide binding sites or more likely, two conformations of the protein.

One-site binding of monomeric enzymes that display cooperativity is described as ‘hysteric enzymes’, a term developed by Carl Frieden (100). Hysteric enzymes are characterized by their apparent time-lag in response to a rapid change in ligand concentration, often due to structural changes and sampling of different conformations in the enzyme. These alternate conformations can either (1) be induced by binding a substrate, termed mnemonic model (101, 102), or (2) be
inherently present in equilibrium, where one conformation is favored as ligand concentration increases, termed ligand-induced slow-transition model (LIST) (103, 104). The mnemonic model states that the enzyme exhibits two conformations, a low and high affinity state; in the absence of substrate, the enzyme strongly favors the low-affinity state, producing the apparent ‘lag’ in activity as the protein transitions to the high affinity after binding the substrate (101, 102). As substrate is added, the enzyme transitions to the high-affinity state and slowly returns to the low-affinity state when product is released; as substrate concentration increases, the conformation equilibrium is pushed to favor the high-affinity state and the enzyme will remain in high affinity as it releases product when substrate is in excess. In contrast, the LIST model states that there is already an equilibrium between low- and high-affinity states in the absence of substrate, where each conformation can complete a separate catalytic cycle; as concentration increases, the high-affinity state becomes more favored (103, 104). In the case of Scabin, Michaelis-Menten kinetics are observed for GH activity, but when the transferase substrate dG is added, an apparent lag and rapid change in activity is observed when the concentration of dG is increased; a proposed kinetic scheme is shown in Fig. 3.4 (unpublished). Distinct from WT Scabin, Q158A/E160A variant transferase activity (unpublished) was best described by a biphasic sigmoidal model; potential reasons for this are presented in chapter 4.

3.1.5: Scabin labels genomic DNA –

Scabin was also tested for its ability to modify genomic DNA of the producing organism, *S. scabies*, as well as a Gram-negative bacterium, *P. aeruginosa*, and the target plant, potato (Fig. 3.3F). Scabin showed weak activity against the genomic DNA of both *S. scabies* and *P. aeruginosa*. However, it showed much higher activity against potato genomic DNA, suggesting the presence of target genes or a preferred DNA conformation in the plant species. Further analysis
**FIGURE 3.4:** Proposed general kinetic scheme to describe Scabin’s observed hysteric activity. $E$ signifies transient 1 and $E'$ is transient 2, present only when NAD$^+$ ($S_1$) and DNA ($S_2$) are both present. $P_1$ is the GH product(s) and $P_2$ is the transferase product(s).
of Scabin’s proposed gene target is presented in Appendix I.

3.1.6: Inhibition of Scabin –

Several small molecules were shown to inhibit Scabin GH activity. These compounds, PJ34, P6-C, P6-D, P6-E, and P6-F (Fig. 3.5), are known inhibitors of other mART toxins (7, 51). Lineweaver-Burk analysis was performed for kinetic inhibition data of Scabin with PJ34 inhibitor to determine the mechanism of inhibition for the P-series inhibitors. As expected, it was clear that PJ34 acts as a competitive inhibitor against Scabin GH activity (Fig. 3.2F).

Another small in-house library of known mART inhibitors was also tested against Scabin, but only those that exhibited significant levels of inhibition were further characterized by determining their respective IC$_{50}$, pIC$_{50}$, $K_i$, and $K_D$ values (Table 3.2). These compounds bind tighter than NAD$^+$ with affinity in the range of low micromolar ($K_D < 50$ µM); the dissociation constant of P6-F could not be measured due to technical difficulties inherent to the estimation of the binding affinity by means of quenching of the Trp fluorescence emission. Nevertheless, based on a competition model, the affinities ($K_D$ values) of a series of inhibitors must correlate with their inhibition constants ($K_i$ values). Since the $K_i$ is the thermodynamic parameter that characterizes the inhibition, and due to the lower relative uncertainties with respect to the nominal values, the $K_i$ values were considered as the representative parameter of the binding/activity of the inhibitors. Thus, PJ34 showed the most promise as a potential lead compound for inhibition of Scabin ($K_i = 3 \pm 0.2$ µM). For the P6 series, the activity ranges between P6-F with the highest activity ($K_i = 7 \pm 0.2$ µM) and P6-E with the lowest ($K_i = 24 \pm 0.3$ µM). Evidently, the chemical and topological constitution of the inhibitor “tail” is a determining factor for the inhibition of Scabin GH activity. In this sense, the hydrophobic character of the terminal piperidine moiety of P6-D and P6-F favors a stronger inhibition in comparison with the “tail-less” benzo- naphthyridinone ring system of the
FIGURE 3.5: *P-series inhibitors effective against Scabin GH activity.* PJ34, 2-[[3-(dimethylamino)-2-oxopropyl]amino]-5,6-dihydrophenanthridin-6-one; P6-C, 8-fluoro-1H,2H,3H,4H,5H,6H-benzo[c]1,6-naphthyridin-6-one; P6-D, 8-fluoro-2-[3-(piperidin-1-yl)propyl]-1H,2H,3H,4H,5H,6H-benzo[c]1,6-naphthyridin-6-one; P6-E, 4-[8-fluoro-6-oxo 1H,2H,3H,4H,5H,6H-benzo[c]1,6-naphthyridin-2-yl]butanoic acid; P6-F, 8-fluoro-2-[3 (piperidin-1-yl)propanesulfonyl]-1H,2H,3H,4H, 5H,6H-benzo[c]1,6-naphthyridin-6-one (59).
TABLE 3.2: Comparison of $K_D$, IC$_{50}$, and solubility parameter values for inhibitors of Scabin (59).

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>$K_D$ (µM)$^a$</th>
<th>IC$_{50}$ (µM)$^b$</th>
<th>$K_i$ (µM)$^c$</th>
<th>pIC$_{50}$ (µM)$^d$</th>
<th>$\Delta T_m$ (°C)$^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PJ34</td>
<td>14 ± 0.5</td>
<td>12 ± 1</td>
<td>3 ± 0.2</td>
<td>4.9</td>
<td>2.4 ± 0.3</td>
</tr>
<tr>
<td>P6-C</td>
<td>25 ± 1</td>
<td>89 ± 4</td>
<td>19 ± 1</td>
<td>4.1</td>
<td>2.2 ± 0.6</td>
</tr>
<tr>
<td>P6-D</td>
<td>42 ± 5</td>
<td>97 ± 7</td>
<td>18 ± 1</td>
<td>4.0</td>
<td>1.8 ± 0.4</td>
</tr>
<tr>
<td>P6-E</td>
<td>50 ± 6</td>
<td>119 ± 2</td>
<td>24 ± 0.3</td>
<td>3.9</td>
<td>2.6 ± 0.7</td>
</tr>
<tr>
<td>P6-F</td>
<td>ND$^f$</td>
<td>38 ± 2</td>
<td>7 ± 0.2</td>
<td>4.4</td>
<td>1.5 ± 0.5</td>
</tr>
</tbody>
</table>

$^a$ The binding affinity of inhibitors to Scabin was measured by the quenching of the intrinsic Trp fluorescence caused by the binding of the ligand to the enzyme active site.

$^b$ The IC$_{50}$ values were determined by fitting each dose-response curve to a Boltzmann sigmoidal function in MicroCal Origin version 8.0.

$^c$ The inhibition constant ($K_i$) was calculated from the experimentally determined IC$_{50}$ values according to the relationship, $K_i = IC_{50}/(1 + ([S_{NAD}]/K_{M(NAD)})$ using fixed value for $[S_{NAD}] = 250$ µM, and $K_{M(NAD)} = 68$ µM.

$^d$ The pIC$_{50}$ values were calculated from the IC$_{50}$ values as follows, pIC$_{50} = −\log IC_{50}$. The higher the pIC$_{50}$ value, the lower dose that is required for 50% inhibition of Scabin toxin activity.

$^e$ The $\Delta T_m$ value (°C) was determined by the expression, $\Delta T_m°C(Scabin\text{-}inhibitor) − \Delta T_m°C(Scabin\text{-}apo)$.

$^f$ Not determined.
P6-C compound as the reference structure. Between P6-D and P6-F, the presence of the sulfonyl group adjacent to the tricyclic ring system in P6-F favors inhibition of Scabin. Notably, the anionic terminal carboxylate of P6-E is detrimental to inhibitor activity (Table 3.2 and Fig. 3.7).

Using DSF, the change in $T_m$ when inhibitors were bound to the active site was studied. A higher $T_m$ than the Scabin apo-enzyme is characteristic of inhibitor binding to the active site and stabilizing the protein (105). However, the change in $T_m$ does not necessarily correlate with the affinity/activity for a dissimilar set of compounds (Table 3.2). Indeed, among the compounds tested, PJ34 exhibited the highest affinity but was not the best stabilizer of the Scabin folded structure (Table 3.2, last column). The inhibitory activity and effect induced by the backbone ring structure of P6-C, P6-D, and P6-F compounds showed greater activity yet reduced stability compared with the P6-E inhibitor. Due to the error and rather small shifts in $T_m$ values, we can only truly conclude that all inhibitors showed a stabilizing effect on the Scabin structure.

3.1.7: Scabin crystal structure –

The structure of recombinant Scabin was refined to a resolution of 1.50 Å in the apo-form (substrate-free) (Table 3.3, Fig. 3.6A). The Scabin structure displays the characteristic mART fold with predominant $\alpha/\beta$ structure (e.g. $\iota$-toxin catalytic, 52% (106); C3bot1, 55% (21); Vis, 55% (107); and Spvb, 59% (108)); notably, Scabin possesses only 32% $\alpha/\beta$ structure as determined by CD spectroscopy. Depicted in Fig. 3.6, the periphery of the toxin crystal structure clearly shows the abundance of loop regions, with the core of the protein containing the majority of ordered secondary structure. The abundance of loop regions in Scabin may provide the dynamics required for the observed hysteresis in transferase activity. Furthermore, Scabin has low sequence identity with most mART toxins except for the Pierisin subgroup (Fig. 1A and B) and significantly differs in topology from well characterized mART toxins like $\iota$-toxin and C3-group toxins, which are
FIGURE 3.6: Scabin-apo crystal structures. (A) close-up of the Scabin-apo crystal structure, coloured green, shown as cartoon diagram. Catalytic residues Gln158 and Glu160 are coloured blue. Other important residues in the reaction mechanism, Arg77 (pink) and STS motif (red), are also highlighted. Disulfide bridges are coloured yellow. (B) structural comparison of Scabin-apo (green) and the catalytic domain of the MTX (PDB ID 2CB6) toxin structure (blue) based on an iterative, three-dimensional alignment of protein backbone C atoms using PyMOL. PN-loop of Scabin (red) and MTX (black) are shown. (C) surface potential of the catalytic subunit of MTX (front view). Molecular surfaces are coloured by the relative electrostatic potential (red, negative or acidic; blue, basic or positive). Surface potentials were calculated using PyMOL APBS software. (D) surface potential of the catalytic subunit of MTX toxin (back view). (E) surface potential of Scabin-apo (side view). F, surface potential of Scabin-apo (opposite side view). Modified from Lyons et al., 2016.


<table>
<thead>
<tr>
<th>Diffraction data</th>
<th>Scabin-apo</th>
<th>Scabin-PJ34</th>
<th>Scabin-P6-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>CLSI-08-ID-1</td>
<td>CLSI-08-ID-1</td>
<td>CLSI-08-ID-1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97949</td>
<td>0.97949</td>
<td>0.97949</td>
</tr>
<tr>
<td>Unit cell parameters (Å)</td>
<td>(a = 87.7, b = 61.0, c = 37.8, ) (\alpha = 90.0, \beta = 99.64, \gamma = 90.0)</td>
<td>(a = 87.5, b = 61.2, c = 38.0, ) (\alpha = 90.0, \beta = 100.4, \gamma = 90.0)</td>
<td>(a = 87.2, b = 60.5, c = 37.9, ) (\alpha = 90.0, \beta = 100.3, \gamma = 90.0)</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Resolution range (Å)(^a)</td>
<td>50.0–1.45 (1.48–1.45)</td>
<td>43.02–1.6 (1.667–1.6)</td>
<td>42.89–1.40 (1.48–1.45)</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>94.4</td>
<td>99.7</td>
<td>99.5</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.3 (4.1)</td>
<td>3.8 (3.7)</td>
<td>3.7 (3.5)</td>
</tr>
<tr>
<td>Average (I/\sigma(I))</td>
<td>18.81 (2.15)</td>
<td>23.33 (2.56)</td>
<td>20.44 (1.69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular replacement program</td>
<td>MolRep</td>
<td>Phenix</td>
<td>Phenix</td>
</tr>
<tr>
<td>(R_{work}) (%)(^b)</td>
<td>16.3</td>
<td>16.6</td>
<td>16.3</td>
</tr>
<tr>
<td>(R_{free}) (%)(^c)</td>
<td>19.7</td>
<td>18.8</td>
<td>17.7</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1319</td>
<td>1326</td>
<td>1326</td>
</tr>
<tr>
<td>Waters</td>
<td>183</td>
<td>143</td>
<td>195</td>
</tr>
<tr>
<td>r.m.s. deviation from ideal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond length (Å)</td>
<td>0.024</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>bond angles (degrees)</td>
<td>2.21</td>
<td>1.05</td>
<td>1.16</td>
</tr>
<tr>
<td>(B)-Factors (Å(^2))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scabin</td>
<td>20.8</td>
<td>34.76</td>
<td>30.83</td>
</tr>
<tr>
<td>Water</td>
<td>25.0</td>
<td>39.25</td>
<td>44.78</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>98.14</td>
<td>97.00</td>
<td>97.00</td>
</tr>
<tr>
<td>Outliers</td>
<td>1.86</td>
<td>0.00</td>
<td>0.61</td>
</tr>
</tbody>
</table>

\(^a\) Values in parenthesis are for the highest resolution shell.

\(^b\) \(|\Sigma|F_o| - |F_c|\Sigma|F_o|\), where \(|F_o|\) and \(|F_c|\) are the observed and calculated structure factor amplitudes, respectively.

\(^c\) The \(R_{free}\) value was calculated with a random 5% subset of all reflections excluded from refinement.
dominated by high helical content at the N-terminus (109). Scabin shares good pairwise sequence identity with MTX toxin from *B. sphaericus* (110) and the apoptosis-inducing Pierisin-1 from the cabbage butterfly *P. rapae* (no structure available). Superposition of the backbone $C_a$ atoms of Scabin with the catalytic subunit of MTX toxin (Protein Data Bank entry 2CB6) reveals a low root mean square deviation (RMSD) of 0.671 Å (for 436 atoms). However, upon closer observation of the backbone traces, some structural differences in certain regions are observed (Fig. 3.6B). The phosphate-nicotinamide (PN)-loop is longer and more extended compared with Scabin (indicated by black colouring in Fig. 3.6B). Notably, the PN-loop motif in Scabin (red region) begins as a helical segment and later becomes coil structure. The PN-loop of mART toxins interacts with the phosphate and nicotinamide moieties of NAD$^+$ and has been proposed to play a role in recognizing the transferase substrate (18, 111). The characteristic EXE catalytic motif found in MTX and in other Arg-modifying ADP-ribosyltransferases is replaced by QXE in Scabin. The QXE motif is also found in the C3 group of toxins and in pertussis toxin, which modify Asn (16, 112) and Cys residues (113), respectively. In the case of Scabin, it modifies deoxyguanosine bases like Pierisin group members (46). Notably, the Scabin structure is also uniquely stabilized by two important disulfide bridges (Cys42-Cys72 and Cys176-Cys190).

Further analysis of the crystal structures of MTX and Scabin using electrostatic potential data from their corresponding solvent-accessible surface areas show distinct patterns from one another (Fig. 3.6, C–F), suggesting that these two toxins do not share a similar substrate. MTX has been shown to ADP-ribosylate elongation factor-Tu in *E. coli* cells, but its eukaryotic substrate has not yet been determined (114); due to the high pairwise sequence identity of MTX with Pierisin-1 (54%), it has been postulated that MTX may have some sort of DNA-binding role. MTX shows patches of both electropositive and electronegative potential throughout its structure (Fig. 3.6).
3.6, C and D), whereas Scabin is predominantly electronegative, with patches of electropositive surface potential near the PN-loop (Fig. 3.6, E and F). Given the propensity of DNA to associate with its corresponding DNA-binding protein via electropositive surfaces, and the lack of these positively charged surface patches in other regions of the Scabin structure, the structural data suggest that the PN-loop and nearby areas probably play a crucial role in DNA substrate binding; structural data supporting this observation is presented in chapter 5.

3.1.8: Scabin in complex with its inhibitor, PJ34 –

The PJ34-bound structure of Scabin was refined to a resolution of 1.60 Å (Fig. 3.7A). As expected, the catalytic residues within the ADP-ribosyl-turn-turn (ARTT)-loop orient toward the inhibitor (Fig. 3.7C) with the catalytic Gln158 shifting nearly 2.5 Å. The PJ34 molecule is stabilized within the active site of Scabin through hydrophobic interactions and hydrogen bonds. Trp128 from the PN-loop appears to play a key role in inhibitor binding and stabilization. The side-chain oxygen of Asn110 is situated 2.9 Å away from the tertiary amine of the R-group of PJ34, which forms another important hydrogen bond. As reported earlier in other mART toxin structures with PJ34 (7, 13, 28), the electron density of the hetero-ring system of the inhibitor is clearly defined; however, the electron density is weaker for the R-group (tertiary amine), suggesting some flexibility in the “tail” of the inhibitor.

3.1.9: Scabin in complex with its inhibitor, P6-E –

The structure of Scabin in complex with the P6-E inhibitor was also determined at 1.50 Å resolution (Fig. 3.7B). P6-E is a member of the P6 series of inhibitors (7), with this structure representing the first report of its crystal complex bound to a mART enzyme. As observed in the Scabin PJ34 structure, the catalytic Gln158 is shifted nearly 2.5 Å upon inhibitor binding, whereas the catalytic Glu160 and other conserved residues that interact with P6-E (Arg77, Asn110, Ser117,
FIGURE 3.7: *Scabin inhibitor crystal structures.* (A) Scabin:PJ34 complex structure shown as a cartoon diagram. PJ34 is coloured teal and represented in stick format. (B) Scabin:P6-E complex structure shown as a cartoon diagram. P6-E is coloured orange and represented in stick format. (C) Interacting residues with PJ34 inhibitor coloured blue, apo coloured plum. Structural differences among important catalytic residues (Arg77, Asn110, Ser117, Trp128, Gln158, and Glu160) are highlighted. (D) Interacting residues of Scabin with P6-E inhibitor (orange) coloured blue, apo coloured plum. Structural differences among important catalytic residues (Arg77, Asn110, Ser117, Trp128, Gln158, and Glu160) are highlighted. Scabin:PJ34 PDB ID: 5EWK; Scabin:P6-E PDB ID: 5EWY; Scabin apo PDB ID: 5DAZ. Adapted and altered from Lyons *et al.*, 2016.
Trp128) do not significantly change position upon inhibitor binding (Fig. 3.7D). P6-E is stabilized by hydrophobic interactions and two hydrogen bonds. The first hydrogen bond is between the main-chain guanidinium group of Arg77 and the oxygen of P6-E (3.4 Å). The second is between the main-chain nitrogen of Ser78 and the oxygen of P6-E (2.8 Å). P6-E bound in the active site closely mimics the PJ34 structural position and has a well-defined electron density throughout the molecule.

3.1.10: Scabin pocket analysis with PJ34 –

The X-ray structure of the Scabin PJ34 complex reveals that the nicotinamide site of the binding pocket is well formed in the Scabin apo-structure (Fig. 3.8A), exhibiting a cavity with primarily hydrophobic character (white spheres) and a deeper hydrophilic one (red spheres). The PJ34 ring system fits into the sub-pocket with no significant variation when compared to the apo-form. Nevertheless, the binding of PJ34 induces dehydration of the nicotinamide pocket, with displacement of two water molecules bound to the backbone of Ser78 (Fig. 3.8B). The backbone of Ser78 and the side chains of Arg77 and Asp79 (both conserved), Lys94, Trp128 (conserved), and Glu158 (conserved QXE motif) are the main interacting residues, with Arg77 and Trp128 constituting the flat volume of the nicotinamide sub-pocket (Fig. 3.8C). Fig. 3.8D depicts a two-dimensional representation of the residue interactions within the NAD⁺ binding pocket. The hydrogen bonds between the fused amide of PJ34 and the backbone atoms of Ser78 is a key feature of the binding mode for P series of inhibitors in several mART toxins (115, 116). However, a few side chains of neighboring residues to Ser78 reorient significantly upon PJ34 binding (Fig. 3.8E). The side chain of Asn110 is shifted to bind the O-carbonyl of the PJ34 “tail,” whereas the conformational variation in the side chain of Ser78 and Ser131 may align with changes in the water structure. The shift in Ser131 may be due to an “artifact” of the modeled coordinates as it is not
**FIGURE 3.8:** *The Scabin·PJ34 complex.* Shown are details of the interaction between Scabin and the PJ34 inhibitor in the context of the apo-form and the modeled complexed with NAD$^+$ and dsDNA$_{10}$. (A) binding pocket on the Scabin-apo. The binding pocket was calculated based upon the X-ray coordinates of the Scabin-apo. The small dots correspond to the center of hydrophilic (red) and hydrophobic (white) spheres into the N-site of the binding pocket. The PJ34 molecule was superposed for reference. (B) binding of PJ34. Shown is the binding pose of PJ34 on the Scabin·PJ34 complex. The crystallographic waters are depicted in yellow from the Scabin-apo structure for reference. (C) Scabin-PJ34 ligand interactions; van der Waals interaction surfaces (vdW-interaction surfaces) around pocket residues, coloured by the electrostatic potential. (D) two-dimensional ligand interactions. Shown is a two-dimensional depiction of the bound PJ34 and interacting residues. The arrows represent the hydrogen bonds, with backbone atoms (blue) and side-chain atoms (green). The degree of exposure is shown by the purple sphere. (E) pocket variation upon binding. Shown are pocket residues with major side-chain variation upon PJ34 binding. Depicted are ribbons and carbon atoms for Scabin (in green) and for Scabin·PJ34 complex (in orange). (F) comparison between Scabin-bound forms. Shown is superposition of the Scabin·PJ34 complex (in orange ribbons and carbon atoms) and the modeled Scabin$_{in}$·NAD$^+$·dDNA$_{10}$ complex (in gray ribbons and carbon atoms), with NAD$^+$ not shown. The backbone of the inward facing DNA strand is shown in green. In all panels, PJ34 is shown as cyan carbon atoms. Figures were rendered by MOE.
well resolved in the X-ray structure.

3.2: Conclusion

Scabin was identified as a mono-ADP-ribosyltransferase toxin produced by the plant pathogen *S. scabies*. Structural and biochemical data suggest that Scabin is a member of the DNA-acting Pierisin subgroup in the mART toxin family, being only the second classified member and first prokaryotic toxin member. Scabin was shown to label the exocyclic amine of guanine nucleobases, similar to that of Pierisin-1 toxin (46). Scabin exhibited hysteresis during transferase activity with the model substrate deoxyguanosine, suggesting alternate forms of the enzyme may exist during catalysis. Mutagenesis studies of the PN-loop may reveal key residues involved in the interaction of Scabin with its DNA substrate, and will be a focus of the following chapter.
4.0: Abstract

The crystal structure of Scabin with NADH as a substrate analog was determined at 1.50 Å, which provided important insights into the active site structure and Michaelis complex of the enzyme. Scabin has four tryptophan residues; variants were made by site-directed mutagenesis to use tryptophan as a probe for the interactions with both NAD$^+$ and DNA substrates. Trp128 and Trp155 were shown to be important for binding and transferase activity, through kinetic analysis of Trp to Ala substituted variants. The crystal structures of two tryptophan variants, W128Y (with and without NADH) and W155A, were also solved, as well as an N110A and S117A variants, providing new insights into the role of these residues in NAD$^+$/DNA substrate binding and catalysis. The roles of catalytic residues (Arg77, Ser117, Gln158, Glu160) were also investigated and combined with a rigorous kinetic analysis of Scabin with model DNA substrates, furthering our understanding of the structure and function of the mART toxin Pierisin-like subgroup. A model of the Scabin-DNA complex was built, which accounts for the observed binding and kinetic interactions of Scabin with the DNA substrate.
4.1: Results and Discussion

4.1.1: Scabin crystal structure with NADH –

The high-resolution structure of Scabin with a substrate analog, NADH, and the structures of two Scabin Trp variants have been solved. Table 4.1 shows the refinement statistics for the Scabin-NADH, Scabin W128Y (with and without NADH), W155A, N110A and S117A structures.

Fig. 4.1A shows the structure of Scabin with NADH bound in the active site along with the key catalytic residues (Arg77, Ser117, Gln158, Glu160) and two important disulfide bonds (C42-C72; C176-C190). This represents the first crystal structure of the substrate-bound state for a DNA-acting mART toxin from the Pierisin-like subgroup. The NADH electron density map is shown as an inset to Fig. 4.1A, demonstrating that this substrate analog was well-resolved in the crystal structure (1.40 Å resolution) with an occupancy of 1.0. NADH is a potent inhibitor of mART toxins and serves as an excellent NAD\(^+\) substrate analog. This competitive inhibitor/ligand was surprisingly stable when bound to Scabin and was not oxidized to NAD\(^+\) during the crystallization process. The disulfide bridge between Cys42 and Cys72 exhibited some unrefined electron density, suggesting an alternate, reduced conformation in the NADH-bound state; this unrefined electron density was not observed in the apo-form, but was also observed in the W128Y-NADH variant. Notably, we cannot discount the possibility that NADH (E\(^{\circ\prime}\) = -0.28V) could be acting as a reducing agent for this disulfide bridge (E\(^{\circ\prime}\) = -0.22V). Using AreaMol (89), the solvent accessible surface area was calculated for each atom of the Scabin-NADH complex. The sulfur atoms of the C42-C72 bridge are only slightly more solvent exposed than the C176-C190 sulfur atoms, with the latter not exhibiting unrefined electron density. The overall C\(_\alpha\) backbone did not shift in the crystal structure when aligned to the WT Scabin structure (PDB ID 5DAZ). Subsequently, we tested the hypothesis that Scabin could exist in both oxidized and
FIGURE 4.1: Scabin binds competitive inhibitor/substrate NADH. (A) Cartoon representation of Scabin (gray) bound to NADH (cyan; represented as sticks). Conserved residues are shown as sticks: Arg77 (purple), Ser117 (blue), and the catalytic Gln158-X-Glu160 (green). The two disulfide bridges (Cys42-Cys72 and Cys176-Cys190) are shown as sticks in red. Inset is the electron density map (2mF_o−nF_c) around the NADH ligand, contoured at 1σ. (B) Induced fit in pocket residues upon NADH binding. Depiction of the pocket residues that change conformation between the apo (dark gray C-atoms, PDB:5DAZ) and the complexed (green C-atoms, PDB:5TLB) forms of Scabin (gray molecular surface) with the NADH ligand (black C-atoms and molecular surface coloured by electrostatic potential). (C) Non-bonded interactions between NADH and Scabin. Active conformations of the Scabin- NADH complex showing (top) the network of direct H-bonds between pocket residues (black C-atoms) and NADH (cyan C-atoms) and (bottom) the steric interaction surface (E_{vdW} = 0 kcal/mol) spectrally coloured according to the electrostatic potential, around the NADH molecule (green C-atoms). (D) In-pocket water molecules. A 2D-diagram depicting the presence of 5 crystallographic water molecules (CWM) that either bridge the NADH molecule to pocket residues, or stabilize the active conformation of the ligand. (E) Role of CWMs in stabilizing NADH and the Scabin- NADH complex. Changes in the configuration of in-pocket water molecules between the apo-form (yellow molecules, PDB:5DAZ) and bound form (brown molecules, PDB:5TLB) upon NADH binding (cyan C-atoms). (F) Induced fit of Trp128 and HOH^{171}. Reconfiguration of Trp128 and HOH171 between the apo-form (green C-atoms, PDB:5DAZ) and the bound form (yellow C-atoms, PDB:5TLB) upon NADH binding. The rotation of the Trp128 indole group increases contacts with NADH (purple shells), and displaces HOH171 to intramolecularly bridge the NADH ligand. Hydrogen atoms shown are as modelled by Phenix.
**TABLE 4.1A:** Crystallographic data and refinement statistics for Scabin-NADH complex, Scabin W128Y, and Scabin W128Y-NADH complex.

<table>
<thead>
<tr>
<th>Diffraction data</th>
<th>Scabin-NADH</th>
<th>Scabin W128Y</th>
<th>Scabin W128Y-NADH</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>CLSI-08-ID-1</td>
<td>CLSI-08-ID-1</td>
<td>CLSI-08-ID-1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97949</td>
<td>0.97949</td>
<td>0.97949</td>
</tr>
<tr>
<td>Unit cell parameters (Å)</td>
<td>( a = 86.0, b = 62.3, c = 38.1, )</td>
<td>( a = 88.4, b = 60.8, c = 37.9, )</td>
<td>( a = 87.1, b = 61.5, c = 37.9, )</td>
</tr>
<tr>
<td></td>
<td>( \alpha = 90.0, \beta = 100.3, \gamma = 90.0, )</td>
<td>( \alpha = 90.0, \beta = 99.6, \gamma = 90.0, )</td>
<td>( \alpha = 90.0, \beta = 99.9, \gamma = 90.0, )</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>42.33–1.4 (1.45–1.4)</td>
<td>37.3–1.35 (1.4–1.35)</td>
<td>42.91–1.5 (1.55–1.5)</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>93.3</td>
<td>100.0</td>
<td>99.5</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.5 (2.3)</td>
<td>4.1 (3.4)</td>
<td>4.5 (3.7)</td>
</tr>
<tr>
<td>Average I/σ(I)</td>
<td>9.42 (1.19)</td>
<td>13.33 (1.22)</td>
<td>17.19 (1.65)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<table>
<thead>
<tr>
<th>Molecular replacement program</th>
<th>Phenix</th>
<th>Phenix</th>
<th>PHASER</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{\text{work}} ) (%)</td>
<td>17.41</td>
<td>15.32</td>
<td>14.43</td>
</tr>
<tr>
<td>( R_{\text{free}} ) (%)</td>
<td>20.60</td>
<td>18.87</td>
<td>17.34</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1299</td>
<td>1310</td>
<td>1328</td>
</tr>
<tr>
<td>Waters</td>
<td>140</td>
<td>156</td>
<td>141</td>
</tr>
<tr>
<td>r.m.s. deviation from ideal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond length (Å)</td>
<td>0.021</td>
<td>0.005</td>
<td>0.021</td>
</tr>
<tr>
<td>bond angles (degrees)</td>
<td>1.36</td>
<td>0.71</td>
<td>1.41</td>
</tr>
<tr>
<td>( B )-factors (Å²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scabin</td>
<td>35.77</td>
<td>26.52</td>
<td>27.52</td>
</tr>
<tr>
<td>Water</td>
<td>47.10</td>
<td>39.14</td>
<td>41.11</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>98</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>Outliers</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
**TABLE 4.1B:** Crystallographic data and refinement statistics for Scabin W155A, Scabin S117A, and Scabin N110A.

<table>
<thead>
<tr>
<th>Diffraction data</th>
<th>Scabin W155A</th>
<th>Scabin S117A</th>
<th>Scabin N110A</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>CLSI-08-ID-1</td>
<td>CLSI-08-ID-1</td>
<td>CLSI-08-ID-1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97949</td>
<td>0.97949</td>
<td>0.97949</td>
</tr>
<tr>
<td>Unit cell parameters (Å)</td>
<td>a = 88.1, b = 60.8, c = 38.4, a = 88.5, b = 60.9, c = 38.0, a = 88.6, b = 60.8, c = 38.0,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>α = 90.0, β = 99.4, γ = 90.0, α = 90.0, β = 99.6, γ = 90.0, α = 90.0, β = 99.4, γ = 90.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Resolution range (Å)(^a)</td>
<td>37.86–1.6 (1.66–1.6)</td>
<td>43.61–1.55 (1.605–1.55)</td>
<td>43.69–1.45 (1.50–1.45)</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>100.0</td>
<td>100</td>
<td>96.1</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.5 (4.4)</td>
<td>4.2 (4.1)</td>
<td>3.8 (3.8)</td>
</tr>
<tr>
<td>Average I/σ(I)</td>
<td>16.21 (1.65)</td>
<td>17.72 (1.64)</td>
<td>22.96 (2.6)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<table>
<thead>
<tr>
<th>Molecular replacement program</th>
<th>Phenix</th>
<th>Phenix</th>
<th>Phenix</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(_{\text{work}}) (%)(^b)</td>
<td>14.50</td>
<td>15.13</td>
<td>15.66</td>
</tr>
<tr>
<td>R(_{\text{free}}) (%)(^c)</td>
<td>18.60</td>
<td>18.7</td>
<td>18.11</td>
</tr>
<tr>
<td>No. of atoms</td>
<td>1311</td>
<td>1300</td>
<td>1272</td>
</tr>
<tr>
<td>Protein</td>
<td>129</td>
<td>156</td>
<td>121</td>
</tr>
<tr>
<td>Waters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r.m.s. deviation from ideal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond length (Å)</td>
<td>0.016</td>
<td>0.005</td>
<td>0.01</td>
</tr>
<tr>
<td>bond angles (degrees)</td>
<td>1.25</td>
<td>0.76</td>
<td>1.11</td>
</tr>
<tr>
<td>B-factors (Å(^2))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scabin</td>
<td>34.76</td>
<td>31.22</td>
<td>25.88</td>
</tr>
<tr>
<td>Water</td>
<td>46.49</td>
<td>45.40</td>
<td>40.12</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>97</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>Outliers</td>
<td>0</td>
<td>0.61</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\) Values in parenthesis are for the highest resolution shell.

\(^b\) |\(\Sigma||F_o|-|F_c||\)/|\(\Sigma|F_o|\)|, where |\(F_o|\) and |\(F_c|\) are the observed and calculated structure factor amplitudes, respectively.

\(^c\) The R\(_{\text{free}}\) value was calculated with a random 5% subset of all reflections excluded from refinement.
reduced forms by measuring the percentage of free thiols. WT Scabin was found to have 9.0% free thiols, notably in the absence of NADH or other substrates (Table 4.2).

Upon NADH binding, the overall substrate pocket architecture of the enzyme is preserved with respect to the apo-form, according to the small Cα-RMSD of 0.27 Å (for 24 residues). This might have its origin in the intrinsic dynamics of the protein since it has the same magnitude as in the entire protein (Cα-RMSD = 0.27 Å for 165 residues). However, the side-chains of pocket residues are conformationally shifted (RMSD = 0.94 Å, 24 residues), as observed mainly for Arg81, Lys94, Asn110, Trp128, and the catalytic Gln158 (RMSD = 1.67 Å for 5 residues) (Fig. 4.1B); this is also observed in the water-coordinated structure (see later). For the catalytic Gln158, two alternate conformations of this side-chain were observed that are different from the apo-structure. A significant conformational change upon NADH binding involves the Trp128 side chain; it appears to have rotated nearly 180°, shifting the nitrogen of the indole ring 4 Å. The minor shifts in side-chain location upon NADH binding without changing backbone orientation could explain the high catalytic efficiency (1.4 × 10⁶ M⁻¹·min⁻¹) (59). For example, upon NAD⁺ binding, the ARTT-loop of C3 toxins usually displays large shifts in conformation, signifying an ‘in’ and ‘out’ phase of the loop; reported catalytic efficiencies are orders of magnitudes lower than Scabin for GH activity (C3larvin = 11 M⁻¹·min⁻¹; C3cer = 2.1 × 10⁵ M⁻¹·min⁻¹), which could suggest that large structural changes within C3 toxins account for their catalytic inefficiency as enzymes (16, 51, 117).

NADH is coordinated in the Scabin active site by a network of hydrogen bonds (Fig. 4.1C, top) and steric contacts (Fig. 4.1C, bottom) with a similar pattern of interactions described for the NAD⁺ substrate in other mART toxins (107, 118, 119). Remarkably, the nicotinamide amide group is anchored by two reciprocal hydrogen bonds with the backbone of Ser78 (at β₁), which is a
TABLE 4.2: Percentage of free thiols present in different Scabin variants. Data represented as mean of two biological replicates, each with n=3; average ± S.D.

<table>
<thead>
<tr>
<th>Scabin variants</th>
<th>Percent Free Thiol (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>9.0 ± 2.1</td>
</tr>
<tr>
<td>W128Y</td>
<td>5.9 ± 0.9</td>
</tr>
<tr>
<td>W155A</td>
<td>7.6 ± 2.0</td>
</tr>
<tr>
<td>W199A</td>
<td>5.2 ± 1.0</td>
</tr>
<tr>
<td>S117A</td>
<td>5.9 ± 1.5</td>
</tr>
<tr>
<td>N110A</td>
<td>13.5 ± 1.9</td>
</tr>
<tr>
<td>Q158A-X-E160A</td>
<td>15.7 ± 1.8</td>
</tr>
</tbody>
</table>
unique substitution for this position into the CT and DT groups (usually Gly but Trp in the Pierisin-like and MTX toxins), and only found in Sco5461, a putative DNA-targeting mART from *S. coelicolor* (120). Residues of the S117-T118-T119 motif, Trp128 (in the PN-loop) and Leu124, together form a flat hydrophobic pocket for the nicotinamide moiety. Additionally, there is room for two ‘out-of-plane’ hydrogen atoms at the C3 of the pyridine ring due to the *trans* location of Thr118, which is expected to be similar in conformation for bound NAD⁺. The catalytic Q158-X-E160 motif interacts with the N (nicotinamide)-ribose, while both conserved Arg77 and Asn110 residues contact the PO₂-O-PO₂ linker, assisted by Lys94 by means of a salt-bridge (Fig. 4.1C). At the adenine side of NADH, Lys94 and Arg81 stack with the adenine moiety, and Ser80 hydrogen bonds the A (adenine)-ribose.

The high resolution of the Scabin·NADH structure facilitated the analysis of CWMs. Ten CWMs observed in the apo-form are displaced upon NADH binding to Scabin. Other in-pocket CWMs stabilize the ligand or bridge it to the protein (Fig. 4.1D). HOH39 (apo-Scabin numbering) shifts and rotates to intramolecularly bridge the NADH ligand, further stabilizing the binding pose (Fig. 4.1E). The highest strain energy of NADH is observed at the adenine moiety. In addition, the A-phosphate is bridged to Asn110 by the displacement of HOH181, and to Ser117 by the rotation of HOH41 (Fig. 4.1E).

A case that deserves special attention corresponds to HOH171 and Trp128 (Fig. 4.1F). Trp128 interacts weakly with neighboring residues in the apo-form (*E*_int= -1.5 kcal/mol), with a rotation observed upon NADH binding. This results in an increase in the steric contact with NADH (*E*_int= -6.3 kcal/mol, purple shells in Fig. 4.1F), inducing a shift in HOH171 and leading to a greater stabilization by coordination with NADH. In summary, the active conformation of Trp128 (which exhibits weak electron density in the NADH bound crystal structure) is compatible with the bonded
location of HOH171 and with the active pose of NADH, validating its conformation represented in the crystal structure.

4.1.2: Role of Trp residues in Scabin –

The primary sequence of Scabin reveals that there are four Trp residues, Trp68, Trp128, Trp155, and Trp199 (59). Trps 68 and 199 are outside the catalytic core domain of the Scabin enzyme and hence likely do not participate in substrate binding and catalysis (59). However, Trp128 and Trp155 are located in the active-site core and both are predicted to participate in transferase substrate interaction and enzyme activity. To understand the roles of the Trp residues in the structure and function of Scabin, site-directed mutagenesis was conducted to replace each individual Trp with either an Ala or Tyr. Substitution for a Tyr was performed if the Ala substitution did not yield a stable protein, as was the case for W128. To characterize these proteins, CD spectroscopy was conducted to determine the effect of Trp substitutions on the folded integrity of the Scabin variants (Fig. 4.2A). Although some effects were noted, the overall consequences of residue replacements for Trps 128, 155, and 199 were not significantly perturbing. This is corroborated by structural alignment for W128Y and W155A experimental models with WT, where $C_\alpha$-RMSD = 0.09 and 0.1 Å for W128Y and W155A, respectively. Notably, Trp68 could not be replaced with either Ala or Tyr without a significant effect on the expression and stability of the Scabin variant.

The UV absorbance spectra of WT Scabin and three Trp variants are all similar, but do show differences from the spectrum of the model compound, NATA (Fig. 4.2B). The fluorescence emission spectra of the WT and Trp variants are also quite similar in their overall emission envelope and maxima (Fig. 4.2C). However, differences are seen in the average fluorescence quantum yields (Table 4.3), reflecting variations of each residue in its electronic environment.
FIGURE 4.2: Spectral data of Trp variants. (A) CD spectra of Scabin W128Y (blue), W155A (red) and W199A (green) in 20 mM Tris, 50 mM NaF, pH 8.2 buffer. WT is shown in black (control). The concentration of each protein was held at 0.16 mg/mL with each spectrum representing the average of 9 independent spectra. (B) Absorbance spectra for Scabin W128Y (blue), W155A (red), W199A (green) and WT (black); NATA control (orange) was prepared at 40 µM. Absorbance was measured from 245-350 nm. (C) Deconvolution of the intrinsic Scabin fluorescence for Trp (n-1) variants. Recorded emission of the global fluorescence (brown dotted line) and fitted curved (yellow solid line) of WT, and W128Y, W155A, and W199A Scabin variants, and the simulated emission from the constituent Trps, coloured as: 68 (black), 128 (red), 155 (green) and 199 (blue). All plots were shown at the same scale for comparison. The reported concentrations correspond to estimated values (nominal ± 5%). (D). Hydrophobic residues surrounding Trp68. Conformation of Trp68 (in black C-atoms) in hydrophobic pocket formed by Pro48, 66 (not shown) and 164, and Val and Iso residues. The pink shells highlight the interacting atoms.
**TABLE 4.3**: Fluorescence parameters of WT Scabin and tryptophan variants.

<table>
<thead>
<tr>
<th>Scabin protein</th>
<th>Protein ( \lambda_{em}^{max} )^a</th>
<th>Protein Quantum Yield (Q)^b</th>
<th>Calculated ( \Delta\Delta G ) ( \text{fold-unfold}^c )</th>
<th>Trp residue</th>
<th>Calculated ( M_i ) value for each Trp residue^d</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>340</td>
<td>0.070 ± 0.02</td>
<td>0.00</td>
<td>W68</td>
<td>0.498</td>
</tr>
<tr>
<td>W128Y</td>
<td>340</td>
<td>0.084 ± 0.02</td>
<td>-1.32</td>
<td>W128</td>
<td>0.298</td>
</tr>
<tr>
<td>W155A</td>
<td>340</td>
<td>0.073 ± 0.02</td>
<td>-1.27</td>
<td>W155</td>
<td>0.291</td>
</tr>
<tr>
<td>W199A</td>
<td>341</td>
<td>0.132 ± 0.01</td>
<td>-1.90</td>
<td>W199</td>
<td>0.169</td>
</tr>
</tbody>
</table>

^a Measured protein fluorescence \( \lambda_{em} \) maximum value as calculated from the fluorescence emission spectrum of the Scabin WT or variant (see *Materials and Methods*).

^b Protein quantum yield (Q) measured as described in *Materials and Methods* using the equation: 
\[ Q = Q_r \left( \frac{I}{I_r} \right) \left( \frac{OD_r}{OD} \right) \] using NATA as the reference standard.

^c \( \Delta\Delta G \) \( \text{fold-unfold} \) for Scabin WT or variants was calculated as described in *Materials and Methods*.

^d \( M_i \) represents the molar fluorescence (AU/µM) for each Trp residue and was calculated as described in *Materials and Methods*.

Data represent the mean ± SD of n = 6 replicates.
and/or exposure to the aqueous solution. The WT protein has an average quantum yield (Q) = 0.07 ± 0.02 for four Trp residues. Assuming that the spectral property of each individual Trp emitter in the variants is similar in the WT protein (i.e., same local environment), then Trp128 and Trp155 have lower Q than the average observed for the WT protein. The replacement of Trp199 with Ala yields a protein with an 89% increase in fluorescence Q (0.132 ± 0.01), suggesting that Trp199 has a relatively low Q value. From the Q values for the three Trp variants, Trp68 is predicted to have a significantly higher Q than the other three Trp residues in Scabin. Nevertheless, for a quantitative assessment of the contribution of each Trp in the corresponding n or n-1 composite signal, a global fit was performed for the emission spectra of the WT and variant protein emission spectra, using the NATA emission as an elemental “waveform function”. This enabled the deconvolution of the recorded spectra for the participating Trp emitters (see Materials and Methods), where each was characterized by its own maximum wavelength, $\lambda_{\text{max},i}$, and molar fluorescence, $M_i$ (Table 4.3, right side). The deconvolution revealed that Trp199 and Trp68 have the lowest and highest molar fluorescence, respectively, while Trp128 and Trp155 are similar to one another (Table 4.3). The calculated average of the molar fluorescence of each composite spectrum, $\bar{M}$, follows the same pattern among each protein (i.e., $\bar{M}_{\text{W199A}} > \bar{M}_{\text{W128Y}} \geq \bar{M}_{\text{W155A}} > \bar{M}_{\text{WT}}$), as observed in the experimental average quantum yield values, $\bar{Q}$, (i.e., $\bar{Q}_{\text{W199A}} > \bar{Q}_{\text{W128Y}} > \bar{Q}_{\text{W155A}} > \bar{Q}_{\text{WT}}$). In addition, the deconvolution showed a ‘blue’ (solvent protected) environment surrounding Trp68 and Trp199 ($\lambda_{\text{max}}$ at 339.8 and 326.8 nm, respectively), in comparison with the ‘red’ (solvent-exposed) environment characteristic of Trp128 and Trp155 ($\lambda_{\text{max}}$ at 345.4 and 343.6 nm, respectively).

As previously stated, it was not possible to obtain a stable and folded protein for the Trp68 variant (located at the N-terminus), indicating its important role in the folding/stability of the
protein. Trp68 is contained in a “Trp-cage” formed by Pro48, Pro66 and Pro164 (Fig. 4.2D), which accounts for its high fluorescence Q and blue-shifted emission. The calculation of $\Delta \Delta G_{\text{fold-unfold}}$ for the W68A and W68Y variants (-2.87 and -2.27 kcal/mol; compared with values in Table 4.3) suggests highly destabilizing values for these proteins and reflects the key role of Trp68 as an anchor for stabilizing the N-terminal region that precedes the $\beta$-core of Scabin.

In contrast, Scabin tolerates changes at Trp199 (at the C-terminus) without an appreciable effect on the secondary-structure composition, as observed in the CD spectrum (Fig. 4.2A) and only a modest effect on the stability according to the $\Delta \Delta G_{\text{fold-unfold}}$ calculation. Accordingly, W199A showed a 2-fold enhancement with respect to WT in its transferase activity with deoxyguanosine (Table 4.5), which may be attributed to variations induced in the dynamics of the protein. Notably, there is an enhanced mobility of central residues in the $\beta_A$ sheet potentially related to the catalytic function of the toxin, where Trp199 might aid in coupling the C-terminal segment with the $\beta_5$ strand. Trp199 interacts directly with N-terminal residues and like Trp68, it is in a hydrophobic pocket (not shown) formed by Pro40, Pro43, Phe45, Cys72 and Phe171, resulting in its blue-shifted fluorescence emission. However, it may be quenched by the hydrogen bond of the indole-H with the backbone of Ala139 and by its location near the Asp138 side chain, resulting in its low quantum yield value (Table 4.3). Trp128 and Trp155 exhibit red-shifted emission spectra and low mean molar fluorescence (Table 4.3) along with low $\Delta \Delta G_{\text{fold-unfold}}$ values of the corresponding variants compared to WT; these observations suggest that they are in solvent-exposed loops within Scabin, which explains their proposed role in DNA substrate binding and catalysis.

4.1.3: Catalytic signature residues in Scabin GH activity –

The kinetic and binding parameters for WT Scabin with the NAD$^+$ substrate (GH activity)
are shown in Table 4.4. As reported previously (59), the WT enzyme bound NAD\(^+\) with micromolar affinity (70 µM) and showed a similar \(K_M\) value (68 µM), as well as \(k_{cat}\) and catalytic efficiency (\(k_{cat}/K_M\)) of 94 ± 2 min\(^{-1}\) and 1.38 \(\times 10^6\) M\(^{-1}\)·min\(^{-1}\), respectively. This indicates that Scabin is a highly efficient NAD\(^+\) GH enzyme (121). Replacement of the Q-X-E motif with Ala, reduced \(k_{cat}\) by 300-fold (Table 4.4). Careful analysis of the Scabin crystal structure reveals that the small Ala residue within the A-X-A variant (Q158A/E160A), allows a closer approach of the bound NAD\(^+\) substrate (Fig. 4.3A). However, the reduced hydrogen bond capability of the A-X-A variant, combined with the smaller contact surface for the variant, resulted in a significant reduction in the calculated interaction energy with NAD\(^+\) (\(E_{int}\) of -99.43 and -104.25 kcal/mol for A-X-A and Q-X-E, respectively), which agrees with the slightly lower binding affinity observed for the A-X-A variant (Table 4.4).

The S117A variant showed an increase in \(K_M\) and a lower \(k_{cat}\) value compared to WT Scabin (Table 4.4). Ser117 is an important residue in the mART toxin S-T-S/T motif known for its role in binding and positioning of the NAD\(^+\) substrate. Regarding the S117A variant, the side-chain of Ala117 preserves the weak CH-bond with the pyridinium ring (Fig. 4.3B, top), while maintaining the magnitude of the steric interaction. Nevertheless, the higher \(K_M\) of the variant is compatible with the lower total interaction energy of NAD\(^+\) in the S117A variant (\(E_{int}\) = -102.9 kcal/mol) in comparison to the WT (\(E_{int}\) = -104.3 kcal/mol), where the lack of hydrogen bond contacts with Tyr106 in the variant (Fig. 4.3B, bottom) might affect the stability of helix \(\alpha_2\) and --in turn-- the interaction of Asn110 with NAD\(^+\). Furthermore, the lower \(k_{cat}\) for the GH activity of the S117A variant in comparison to the WT protein, might be explained by the absence of the stabilizing effect of the Ser117 hydroxyl on the catalytic Glu160 (Fig. 4.3B, bottom).

The effect of the different substitutions on the GH activity can be evaluated by their
### TABLE 4.4: Kinetic parameters of Scabin WT and variants for GH activity and NAD$^+$ binding.

<table>
<thead>
<tr>
<th>Construct</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$K_D$ (µM)</th>
<th>$k_{cat}/K_M$ (M$^{-1}$·min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>68 ± 3</td>
<td>94 ± 2</td>
<td>70 ± 3</td>
<td>1.4×10$^6$</td>
</tr>
<tr>
<td>Q158A-X-E160A</td>
<td>53 ± 9</td>
<td>0.3 ± 0.02</td>
<td>86 ± 7</td>
<td>5.8×10$^3$</td>
</tr>
<tr>
<td>S117A</td>
<td>88 ± 14</td>
<td>1.3 ± 0.16</td>
<td>23 ± 3</td>
<td>1.5×10$^4$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>435 ± 10</td>
<td></td>
</tr>
<tr>
<td>N110A</td>
<td>66 ± 12</td>
<td>1.0 ± 0.05</td>
<td>ND</td>
<td>1.5×10$^4$</td>
</tr>
<tr>
<td>W128Y</td>
<td>17 ± 3</td>
<td>1 ± 0.01</td>
<td>10 ± 4</td>
<td>5.9×10$^4$</td>
</tr>
<tr>
<td>W155A</td>
<td>55 ± 10</td>
<td>10 ± 0.4</td>
<td>43 ± 24</td>
<td>1.8×10$^5$</td>
</tr>
<tr>
<td>W199A</td>
<td>40 ± 7</td>
<td>12 ± 0.12</td>
<td>38 ± 4</td>
<td>3.0×10$^5$</td>
</tr>
</tbody>
</table>

Data represent the mean ± SD of n = 3 replicates.
**FIGURE 4.3:** *Analysis of WT vs. variant Scabin structures*. Active conformations and interactions of catalytic residues of the WT Scabin and variants complexed with NAD$^+$ *in-silico*. (A,C) Effect of the QxE/AxA substitutions. Variations in the vdW and hydrogen bond interactions with bound NAD$^+$. (B) Effect of the S117A substitution. (top) Direct interaction of (Ser/Ala)117 with NAD$^+$. (bottom) Ser177-mediated Tyr106–Glu160 interaction. (D) Effect of the W155A substitution. Reduction of the vdW-contact with Gln158. (E) Effect of W128Y substitution. (F) Comparative effect of W155Y and W128A on Gln158. Combined rendering of (Trp/Tyr)128, (Trp/Ala)155, and WT Gln158 configuration. In all panels: vdW-surfaces were coloured according to their electrostatic potential as calculated for the WT protein. NAD$^+$ (cyan) and C-atoms for the variants (gray). WT residues are shown in black C-atoms and substitution residues in green C-atoms.
influence on the conformation/stability of the catalytic residues, Gln158 and Glu160, and their interaction with NAD$^+$ (Fig. 4.3C). In kinetic terms, establishing the residual GH activity of the double A-X-A variant as the baseline activity, and assuming that W128Y obliterates the catalytic role of Gln158 (conformationally shifted) (Fig. 4.3D,E), and likewise S117A on Glu160 (Fig. 4.3B), the participation of each catalytic residue can be assessed by the $k_{cat}(W128Y)/k_{cat}(A-X-A)$ ratio of ~3-fold for Glu160, and by the $k_{cat}(S117A)/k_{cat}(A-X-A)$ ratio of ~4-fold for Gln158. However, the combined action of Gln158 and Glu160 has a synergistic effect on the GH activity, as seen by the ratio $k_{cat}(WT)/k_{cat}(A-X-A)$ of ~300-fold. Similarly, Trp128 and Trp155 sandwich Gln158 in their active conformations (Fig. 4.3F). Notably, the smaller impact of the W155A substitution, assessed by the $k_{cat}(WT)/k_{cat}(W155A)$ ratio, when compared with the W128Y (assessed by the $k_{cat}(WT)/k_{cat}(W128Y)$ ratio) of ~10- and ~100-fold, respectively, can be accounted for by the smaller reduction in the van der Waals interaction with Gln158. This, in turn, impacts the stability of the Gln158 catalytic residue. In other words, the major interaction surface of Trp155 with Gln158 corresponds to the methylene side chain of Trp155, and is thus maintained in the W155A substitution.

4.1.4: ADP-ribosyltransferase activity –

The kinetic parameters for WT Scabin transferase activity with a model nucleoside substrate, deoxyguanosine, are shown in Table 4.5. The WT enzyme produced a $K_{0.5}$ value of 302 $\mu$M, with a $k_{cat}$ and catalytic efficiency of 83 min$^{-1}$ and $2.75 \times 10^5$ M$^{-1}$·min$^{-1}$, respectively (59). The relatively similar $k_{cat(mART)}$ (83 min$^{-1}$) with the $k_{cat(GH)}$ (70 min$^{-1}$) values (Table 4.5) reveal that the hydrolytic and transfer steps are kinetically concerted, without a rate-limiting step, at least for the ADP-ribosylation of the mononucleoside substrate. Deoxyguanosine was used for all transferase studies as it did not significantly affect the baseline fluorescence of $\varepsilon$-NAD$^+$, whereas
TABLE 4.5: Kinetic parameters of Scabin WT and variants for transferase activity with deoxyguanosine substrate.

<table>
<thead>
<tr>
<th>Construct</th>
<th>$K_{0.5}$ (µM)</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$k_{cat}/K_{0.5}$ (M$^{-1}$·min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>302 ± 12</td>
<td>83 ± 5</td>
<td>2.8×10$^5$</td>
</tr>
<tr>
<td>Q158A-X-E160A</td>
<td>101 ± 12</td>
<td>0.5 ± 0.04</td>
<td>5.0×10$^3$</td>
</tr>
<tr>
<td></td>
<td>357 ± 26</td>
<td></td>
<td>1.4×10$^3$</td>
</tr>
<tr>
<td>S117A</td>
<td>165 ± 20</td>
<td>2.3 ± 0.1</td>
<td>1.4×10$^4$</td>
</tr>
<tr>
<td></td>
<td>634 ± 14</td>
<td></td>
<td>3.6×10$^3$</td>
</tr>
<tr>
<td>N110A</td>
<td>177 ± 22</td>
<td>5 ± 0.2</td>
<td>2.8×10$^4$</td>
</tr>
<tr>
<td></td>
<td>814 ± 83</td>
<td></td>
<td>6.2×10$^3$</td>
</tr>
<tr>
<td>W128Y ($K_M$)$^b$</td>
<td>1399 ± 167</td>
<td>23 ± 0.4</td>
<td>1.7×10$^4$</td>
</tr>
<tr>
<td>W155A</td>
<td>ND$^a$</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>W199A ($K_M$)$^b$</td>
<td>641 ± 82</td>
<td>178 ± 7</td>
<td>2.8×10$^5$</td>
</tr>
</tbody>
</table>

$^a$ No detectable transferase activity was observed.

$^b$ Data displayed a hyperbolic response and thus are described by the Michaelis-Menten constant ($K_M$) and catalytic efficiency of $k_{cat}/K_M$.

Data represent the mean ± SD of n=3.
substrates like GDP, cGMP or DNA exhibited strong enhancement of the fluorescent signal at concentrations > 200 µM (122). Using a high-performance liquid chromatography (HPLC) based assay would offer an alternative to help characterize the kinetics of Scabin using a more biologically relevant substrate, such as DNA.

The kinetic plots for the Q158A-X-E160A, S117A and N110A variants were non-Michaelis-Menten and yielded two $K_{0.5}$ values for each (Table 4.5). These $K_{0.5}$ values consisted of a lower and higher value compared with the $K_{0.5}$ value of the WT. The presence of two $K_{0.5}$ values is most likely due to the occurrence of two independent conformations, resulting from the substitution of residues. The crystal structures of N110A and S117A variants (Table 4.1) were determined, revealing a reduced conformation of the C-terminal disulfide bridge (Cys176-Cys190), which in the final model exhibited an occupancy of approximately 20% (N110A) and 10% (S117A). This value approximately correlates with the experimentally determined percentage obtained from the mono-bromobimane assay (Table 4.2). These three variants exhibited an expected loss in turnover number. The Q158A-X-E160A variant gave a $k_{cat}$ of 0.5 min$^{-1}$, 166-fold less than WT Scabin (300-fold less for GH activity). The S117A variant showed a 36-fold loss in activity compared with the WT enzyme (72-fold loss for the GH activity). The N110A variant produced an enzyme with a 17-fold loss in activity (Table 4.5).

Notably, the Trp variants showed Michaelis-Menten kinetic behaviour (transferase) with the deoxyguanosine substrate. The W128Y enzyme showed a higher $K_M$ (1399 µM; 4.6-fold increase) with a 3.6-fold loss in turnover number (23 min$^{-1}$), and the W155A enzyme had completely lost its transferase activity. Ala replacement at Trp199 produced a mixed effect on the enzyme catalytic activity—impaired by 2-fold for the $K_M$ value (641 µM), but increased the $k_{cat}$ value (178 min$^{-1}$). Contrary to the direct effect produced by the Trp128 and Trp155 mutations
(W128Y and W155A) in the active conformation of ligand(s) and pocket residues, the effect of W199A may be allosterically controlled. The development of variants which have lost their cooperative nature provides evidence that the observed hysteresis in WT Scabin is due to inherent properties of the enzyme, and not to the nature of the dG substrate in aqueous solution.

4.1.5: NADH binding and inhibition of Scabin GH activity –

It was found that NADH is a strong competitive inhibitor against the NAD$^+$ substrate; the inhibition curve with an IC$_{50}$ value of 7.2 μM is shown in Fig. 4.4A. The binding of NADH to Scabin, monitored by Trp fluorescence quenching, is shown in Fig. 4.4B. NADH binds Scabin with an affinity of 17 μM (Fig. 4.4B). The 4-fold higher binding affinity of Scabin to NADH (non-hydrolyzable by Scabin) in comparison with the $K_D$ of approximately 70 μM for the natural substrate NAD$^+$ ($\Delta \Delta G_{\text{bind}} \approx 0.9$ kcal/mol) must reside in a stronger van der Waals interaction with the two out-of-plane H-atoms at the pyridine C3. This may be in addition to an attenuation of unfavorable electrostatic interactions with adjacent polar residues due to the lack of the charged pyridine N-atom in NADH. Based on the kinetic and thermodynamic parameters, the interaction of NADH is very similar to that of NAD$^+$, making NADH a model compound for crystallographic studies of the ES complex when the GH activity of the enzyme is too high for trapping intact substrate without hydrolysis.

4.1.6: Scabin-DNA binding –

The ability of WT Scabin to bind DNA substrate was measured with a synthetic double-stranded DNA tagged at the 5' termini with a cyanine-3 fluorophore. Affinity of W128Y, W155A and N110A variants for DNA was measured by employing a fluorescence anisotropy binding assay (Fig. 4.4C). WT Scabin produced a $K_D$ of 51 μM, whereas the W128Y variant exhibited 9 times weaker affinity for dsDNA with a $K_D$ of 458 μM (Table 4.6). The $K_D$ for the W155A variant was
**FIGURE 4.4:** NADH and DNA binding. (A) IC₅₀ curve for NADH with WT Scabin. Scabin (50 nM) was incubated with 250 µM β-NAD⁺ and varying concentrations of NADH (0-100 µM); reaction volume was 60 µL. (B) NADH binding curve for WT Scabin. Scabin (1.25 µM) was titrated with increasing concentrations of NADH (0-1 mM) in 25 mM Tris-HCl, pH 8.2, 200 mM NaCl. (C) Binding curves for Scabin with Cy3 tagged double-stranded DNA oligomer containing +1 overhang on either termini (see Materials and Methods). Scabin was preincubated with 1.5 mM of either nicotinamide (○), ADP-ribose (●), NADH (□) or buffer only control (▲). dsDNA was prepared at 5 µM in 25 mM Tris-HCl, pH 8.2, and 100 mM NaCl. In the case where ligand was used, 1.5 mM of ligand (nicotinamide, ADP-ribose or NADH) was added to DNA prior to titration of complex. The change in anisotropy (Δr) was measured for 20s intervals; band passes were set to 4 nm, with excitation and emission wavelengths of 550 and 570 nm, respectively. Data represented as n = 3 ± SD; collected using a fluorimeter. (D) Binding curves for Scabin variants with blunt-ended dsDNA. W155A (———), W128Y (······), N110A (······) or WT (—) was titrated into 5 µM Cy3-dsDNA.
determined as 167 µM, three times lower affinity than WT. Despite showing no transferase activity using dG as a substrate, W155A was still able to bind dsDNA. The N110A variant exhibited only slightly lower affinity as compared to WT, with a $K_D$ of 82 µM. These findings suggest that the binding mode of a DNA oligo substrate to Scabin is likely different than for a small nucleotide substrate (such as dG), possibly by multiple interactions with the protein at various sites. These two Trp residues represent the first evidence of a DNA-binding motif harbored within the Scabin protein. If tolerated by the enzyme, producing cumulative variants involving residues that may be involved in electrostatic interactions might reveal important roles in substrate recognition since it may be expected that DNA substrate binding reflects the concerted interactions of a group of active-site residues.

Scabin showed relatively transient binding with a synthetic, double-stranded DNA substrate. WT Scabin produced a $K_D$ of 51 µM with blunt-ended dsDNA; however, Scabin exhibited an increase in affinity for a dsDNA with a single-base overhang on either termini (Table 4.7). The $K_D$ of WT Scabin with dsDNA containing a single-base overhang was determined to be 15 µM. The addition of a second-base overhang did not show a significant increase in affinity from the single-base overhang substrate, indicating a preference for binding dsDNA with a single-base overhang. This observation is much like the mechanism of activation for PARP-1, which is employed by eukaryotic cells during repair of DNA single-strand breaks (60, 123). PARP-1 exhibits a higher affinity for damaged dsDNA with a single-stranded nick as compared to blunt-ended DNA (double-stranded break) (124). PARP-1 also binds DNA with a Zinc-finger motif; Scabin does not possess such a motif, which was experimentally validated, since it showed no dependence on any divalent cations for DNA binding (data not shown). Notably, Scabin possesses no sequence or structural similarities to PARP-1 beyond the classic catalytic motifs present in
TABLE 4.6: Binding constants of Scabin WT and variants with blunt-end double-stranded DNA.

<table>
<thead>
<tr>
<th>Scabin Protein</th>
<th>$K_D$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>51 ± 4</td>
</tr>
<tr>
<td>N110A</td>
<td>82 ± 5</td>
</tr>
<tr>
<td>W128Y</td>
<td>458 ± 119</td>
</tr>
<tr>
<td>W155A</td>
<td>167 ± 21</td>
</tr>
</tbody>
</table>

Data represent the mean ± SD of n = 3 replicates.
TABLE 4.7: Binding constants of Scabin WT for different synthetic double-stranded DNA.

<table>
<thead>
<tr>
<th>dsDNA</th>
<th>$K_D$ (µM)</th>
<th>$K_D$ (µM) with 1.5 mM nicotinamide</th>
<th>$K_D$ (µM) with 1.5 mM ADP-ribose</th>
<th>$K_D$ (µM) with 1.5 mM NADH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blunt end</td>
<td>51 ± 4</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>+1 overhang</td>
<td>15 ± 2</td>
<td>29 ± 1</td>
<td>21 ± 1</td>
<td>ND$^a$</td>
</tr>
</tbody>
</table>

Data represent the mean ± SD of $n = 3$ replicates.

$^a$ ND signifies no detectable binding was observed.
ADP-ribosyltransferases. Therefore, Scabin is clearly a unique enzyme among members of the ADP-ribosyltransferase family.

Binding of overhang dsDNA in the presence of various substrate fragments or analogues (ADP-ribose, nicotinamide or NADH) revealed Scabin’s inability to bind DNA in the presence of NADH (Fig. 4.4F). The $K_D$ of Scabin with dsDNA, when in complex with nicotinamide or ADP-ribose, was determined to be 21 and 29 $\mu$M, respectively (Table 4.7). However, no detectable binding was observed when Scabin was in complex with NADH. This suggests that there are structural differences in the Scabin·NAD$^+$ and Scabin·NADH complexes, likely due to the loss of charge in the nicotinamide ring in the latter. Therefore, an atomic 3D-structure of the Scabin-DNA complex is an important prerequisite for a structure-function analysis of these biochemical data.

4.1.7: Model of the Scabin-DNA complex –

Scabin active-site variants showed an impaired ability to bind the DNA substrate, suggesting their role in surface recognition, as proposed previously (59). The toxin coordinates of the Scabin·NADH structure (PDB: 5TLB) were used as the basis to construct an in-silico Scabin-DNA binary complex database to help understand/interpret DNA substrate binding data. The top-ranked models (decoys) reported a similar binding mode (i.e., location and relative orientation) of the dsDNA ligand over the Scabin toxin, which differ only in an offset of the contacting segment (Fig. 4.5A). A decoy depicting the intercalation of the electronegative dsDNA strand into the active-site cleft via a major surface contact—possessing an electro-positive potential—is shown in Fig. 4.5B. Further inclusion of the NAD$^+$ ligand and in-pocket water molecules produced a Scabin-NAD$^+$-DNA-water complex. This revised Scabin-DNA model reveals the important participation of Asn110, Trp128 and Trp155, among others, in agreement with the negative effect on the dsDNA binding affinity observed in single-point mutations.
FIGURE 4.5: Scabin-NADH-DNA model. (A) Decoys of the Scabin-DNA complex. Five top-ranked decoys of the Scabin (surface coloured by electrostatic potential) complexed with a 21-mer dsDNA molecule (coloured backbone), showing the relative shift in the location of the bound dsDNA. The arrows represent the direction 5’→3’ of the strand defined in Materials and Methods. (B) Model of the Scabin-DNA complex. Refined pose of the complex decoy #4 (see A), depicting the complementarity between the electrostatic contact surface of Scabin (top) and the innermost DNA strand (bottom). In both, molecular surface of Scabin is shown as gray and the DNA backbone in cyan. The NADH molecule in shown as black sticks for reference.
FIGURE 4.5: Continued from pg. 100: (C) Active conformation of Asn110 in the Scabin-NAD$^+$.DNA model. Depiction of the active conformation of Asn110 (green) in the modelled ternary complex, showing hydrogen bond contacts with both NAD$^+$ (in black) and two consecutive bases (Gua9 and Gua10) of the innermost 21-mer DNA strand. (D) Active conformation of Trp128 and Trp155 in the Scabin-NAD$^+$.DNA model. Depiction of the active conformation of the two active-site Trps, Trp128 and 155 (in green) in their interactions with three consecutive bases (Ade8, Gua9 and Gua10) of the innermost 21mer DNA strand. The pink shells highlight the interacting atoms. (E) Embedding of Tyr129 and Lys154 into DNA in the Scabin-NAD$^+$.DNA model. Top. Upper view of the modelled ternary complex among Scabin (in translucent gray surface), DNA (cyan) and NAD$^+$ (not depicted), showing the spatial location of the Scabin interacting residue (green), and making evident the embedding of Tyr129 and Lys154 into the DNA molecule. Bottom. Close view of Lys154 (left) and Tyr129 (right), both in green C-atoms and with molecular surfaces coloured by electrostatic potential, embedded in their active conformation into the minor groove. (F) Interaction of an overhang base. (Top) Hypothetical rotamers of a terminal guanine residue (spectrally coloured) illustrating the increased mobility of an unpaired base. (Bottom) Configuration of the unpaired guanine base (black C-atoms) being stabilized by interaction with Lys154 and water-bridged with two protein residues. In both, molecular surface of Scabin is shown as translucent gray and the DNA backbone and paired bases in cyan.
involving these residues (Table 4.6). Asn110 appears to bridge the innermost DNA backbone with the NAD$^+$ molecule (Fig. 4.5C). The dual role of this residue in binding both ligands might be related to its absolute conservation within the Pierisin-like subgroup of mART toxins. In the model, Trp128 and Trp155 share contacts with backbone atoms of three consecutive nucleotides (Fig. 4.5D); this includes direct (Trp155) and water-mediated (Trp128) hydrogen bonding (not shown). Importantly, Trp128 is absolutely conserved and Trp155 is highly conserved within the Pierisin-like group (Fig. 4.2A). Based on the model, Lys154 was predicted to interact with the DNA backbone, however a similar effect was observed for this substitution as was shown for N110A ($K_{D(DNA)} \approx 80 \mu M$; data not shown). Again, making cumulative variants would allow for identification of residues involved in electrostatic interactions.

The variable shift in the dsDNA ligand docked within the active site of Scabin raises the possibility of a close encounter between the terminal bases of the DNA substrate and Scabin residues. A DNA base overhang offers free functional groups that might contact protein residues (Fig. 4.5F, top). Effectively, the data in Table 4.7 show a modest increased binding affinity with respect to the blunt-end parental substrate, due to the presence of 1 or 2 overhanging bases. As an example, a dsDNA substrate with a single base overhang was superposed to the previous decoy, and a conformational search was performed on the terminal base and neighbor (< 9Å) residues. Fig. 4.5F (bottom) shows the lowest energy conformer depicting the additional interactions of the terminal and unpaired base with Scabin active-site residues.

4.2: Conclusion

Through site-directed mutagenesis combined with kinetic analyses, we have identified two residues, Trps 128 and 155, that are important for binding and catalysis of the DNA substrate. These residues represent the first insight of a potential DNA-binding motif of Scabin, unique from
the zinc finger of the DNA-acting ADP-ribosyl-polymerase, PARP-1 (60). Due to the transient nature of the Scabin DNA complex, crystallography was not an option for further structural studies. The focus of the following chapter will be on the use of hydrogen-deuterium exchange coupled with mass spectrometry to assist with mapping the Scabin:DNA interface to define the DNA-binding footprint on Scabin.
CHAPTER 5: TOWARDS THE ELUCIDATION OF THE SCABIN·DNA STRUCTURE: A TIME-RESOLVED HYDROGEN DEUTERIUM EXCHANGE COUPLED WITH MASS SPECTROMETRY STUDY.
5.0: Abstract

Hydrogen-deuterium exchange coupled with mass spectrometry was used to characterize the protein-DNA interface of the mART toxin, Scabin. Percentage of deuterium uptake was mapped onto apo-Scabin, revealing a partially protected core and some ordered loop regions. The N-terminus of Scabin was significantly protected in the solution, suggesting it occupies a more ordered conformation than depicted in the X-ray crystal structure. Deuterium exchange of the Scabin-NADH complex was performed to compare the dynamics of the complex in solution with the static crystal structure. The catalytic Gln158-X-Glu160 motif was significantly protected upon NADH binding, whereas the catalytic Arg77 exhibited a large increase in deuterium uptake. Deuterium uptake in the presence of DNA was mapped onto the Scabin-DNA model ternary complex, presented in chapter 4. The PN-loop exhibited a strong decrease in deuterium uptake upon DNA binding, and was proposed as an important interface for substrate recognition and binding.

The work presented in this chapter was performed in collaboration with Dr. Derek Wilson’s Laboratory at York University.
5.1: *Introduction*

5.1.1: Time-resolved hydrogen-deuterium exchange –

As discussed previously, hydrogen-deuterium exchange (HDX) was established on the principle that secondary amide backbone hydrogens that are ordered in secondary structures exchange less readily with hydrogens in solution than areas that are less ordered (92, 125). This equilibrium can be exploited to measure protein dynamics and ligand binding through deuterium exchange. Back-exchange of deuterium during the time course of the experiment is a problem and must be minimized. Optimal conditions that reduce back-exchange include temperature (0°C), acidic pH (2.5) and reduced processing time after the exchange reaction has occurred. The deuterium uptake can be quantitatively measured by coupling HDX with mass spectrometry. The Wilson laboratory at York University has developed a method using microfluidics to allow for millisecond-time-scale labelling and HDX detection (92).

A schematic of the current time-resolved electrospray ionization microfluidic device set-up is shown in Chapter 2: Materials and Methods. Time-resolved electrospray ionization mass spectrometry (TRESI-MS) can be used to study protein dynamics and enzyme kinetics on a millisecond time scale. Incorporating this method onto a ‘micro-chip’ allows for a concerted workflow, whereby all steps are coupled to decrease sample-handling and delays that can lead to increased back-exchange. HDX of Scabin was performed using a ‘bottom-up’ approach, whereby deuterated protein samples are quenched and subjected to limited proteolysis through treatment with an acid-stable protease—in this case, pepsin was used (92, 126). This approach yields a resolution of 4-10 residues, and produces a semi-quantitative analysis of protein dynamics and ligand binding; the ‘top-down approach’ can achieve single-residue resolution due to fragmentation of the protein in the gas phase (125). Though the top-down approach has significant
advantages over the bottom-up approach, top-down has a size-limit and specialized, high sensitivity equipment is required. The steps of the bottom-up approach incorporated onto the mini-chip are as follows: deuterium exchange of sample and mixing, followed by acid quenching, protein digestion and subsequent delivery to the electro-spray source (92). For the following experiments, a single time-point was selected to assess global changes in the deuterium uptake of Scabin reflecting the substrate footprint when Scabin is bound to either NADH or a 21-base, double-stranded DNA substrate containing a single-base overhang as a model for a single-stranded break (SSB).

5.2: Results and Discussion

5.2.1: HDX of Scabin –

HDX coupled with mass spectrometry (HDX-MS) was performed on apo-Scabin, using the TRESI set up (Fig. 2.1). A single time-point (4 s) was collected for all experiments to probe the structure of Scabin in solution (92). Four biological replicates were performed for apo-Scabin, with each replicate containing n = 3 or 4 individual runs. Peptides were identified using the online ExPASY FindPept tool and corresponding HDX values were calculated using in-house FORTRAN software developed in the Wilson lab (92). Peptides were collected for separation by LC and identity was confirmed by tandem MS. For the analysis, the first residue does not exchange, and the second residue is assumed to undergo rapid back-exchange; prolines lack a secondary amide hydrogen, and thus never participate in deuterium exchange (127). The isotopic distribution for each peptide was analyzed at a single-reaction time-point, resulting in a total sequence coverage of 85% (including expression tag); total sequence coverage including unconfirmed peptides was 98%. A cartoon representation of deuterium uptake mapped onto the apo crystal structure of Scabin is shown in Fig. 5.1A.
**FIGURE 5.1: HDX profile of apo-Scabin.** (A) Deuterium exchange profile mapped onto the apo-Scabin crystal structure. Colouring scheme is as follows: no data is black, 0-29% gray, 30-39% blue, 40-49% green, 50-59% yellow and 60-100% red. Regions that were not analyzed are coloured black. (B) Hydrogen-bonding network in a protected β1-α2 loop; residues are labelled and coloured based on the uptake range. Main-chains are represented as sticks with side-chains hidden from view; back-bone hydrogen-bonds are coloured black. (C) Deuterium uptake of representative peptides mapped onto the primary sequence of Scabin, black colouring indicates deuterium loss due to back exchange of the first two residues. Secondary structure is shown, with helices represented as blue cylinders and strands represented as green arrows; number indicates order in structure. PN-loop and ARTT-loop are both outlined. See *Materials and Methods* for details on the system.
Most of the structure showed relatively low deuterium uptake (0-29%) (Fig. 5.1A). These low-exchange areas within Scabin are surprisingly dominated by loop regions, which would be expected to be highly mobile and to exhibit the highest deuterium uptake. However, upon closer inspection, these regions are not classical, but consist of ordered loops involved in hydrogen-bonding that stabilize adjacent loop regions, as shown in Fig. 5.1B. Rob et al. (2012) showed this to be the case in cytochrome c, where low uptake was observed for a large loop region due to its involvement in hydrogen-bonding with neighbouring secondary structure elements (92).

Studying the dynamics of Scabin in solution and comparing the observed deuterium uptake to the known crystal structure (PDB ID: 5DAZ) clearly shows that there are differences in the depicted static model. The most obvious difference is in the extended N-terminus; in the crystal structure, the dynamic N-terminus sits in the binding cleft of the adjacent protomer in the crystal. This characteristic of Scabin prevents soaking crystals with ligands to obtain complexes; only co-crystallization of Scabin with the desired ligand allows the formation of binary complexes. The observed deuterium uptake of the N-terminus (residues 30-44) correspond to 20 ± 1%, which is significantly protected when compared to the PN-loop (83 ± 12%) of Scabin. Based on this observation, the N-terminus must fold and form a hairpin-like structure to reduce entropic costs of a dynamic, relatively hydrophobic segment in solution.

Scabin regions containing catalytic residues exhibited expected levels of uptake: R77 (HDX $x_{76-79} = 46 \pm 7\%$) located in $\beta 1$ is involved in intramolecular $\beta$-sheet stabilization of $\beta 3$; N110 (HDX $x_{103-119} = 28 \pm 3\%$); S117-T118-T119 (HDX $x_{116-119} = 34 \pm 4\%$). The catalytic Q158-X-E160 motif was not present in the apo-Scabin peptides, but adjacent regions showed deuterium uptake values of 26 ± 2% ($x_{161-166}$). As previously discussed, Scabin contains four Trps, with known solvent environments (see Chapter 4). Based on the Trp emission $\lambda_{\text{max}}$ values, the degree
of solvent exposure (least to most) was determined as follows: W199, W68, W155 and W128. The
HDX data reflect the previously observed solvent environments; the deuterium uptake percentages
of regions near the Trps are as follows: W199 ≈ 30 ± 3% (x_{196-200}), W68 ≈ 38 ± 5% (x_{68-72}) and
W128 ≈ 52 ± 5% (x_{128-131}). W155 was not present in any peptides (or within proximity to observed
peptides); however, W155 would expect to have an uptake percentage between W68 and W128
based on Trp spectral data (see Chapter 4).

HDX data in the vicinity of the two disulfide bridges (Cys42-Cys72 and Cys176-Cys190) show that the environment around the N-terminal bridge (HDX x_{42-47} = 42 ± 5%; HDX x_{68-72} = 38
± 5%) is slightly more exposed than the C-terminal bridge (HDX x_{169-175} = 42 ± 5%; HDX x_{189-197}
= 32 ± 5%), as previously determined by analysis of the accessible surface area (ASA) calculation
performed using ArealMol (89). A full primary sequence overview of representative peptides is
shown in Fig. 5.1C.

5.2.2: HDX of Scabin with NADH –

HDX-MS was performed on Scabin in complex with NADH. A single time-point (4 s) was
collected for all experiments; two biological replicates were performed for apo-Scabin, with each
replicate containing n = 3 or 4 individual runs, allowing for data to be averaged for replicate
peptides. Total confirmed sequence coverage was 99% (including the expression tag), with total
sequence coverage including unconfirmed peptides at 100%. A cartoon representation of the
change in deuterium uptake compared to the apo-state (%HDX_{free} - %HDX_{bound}) mapped onto the
Scabin-NADH crystal structure (PDB ID: 5TLB) is shown in Fig. 5.2A; negative changes in
deuterium uptake correspond to a decrease in protection (%HDX_{bound} > %HDX_{free}) and positive
changes correspond to an increase in protection (%HDX_{bound} < %HDX_{free}).

The difference map reveals expected protection from deuterium exchange in the ADP-
FIGURE 5.2: HDX of Scabin with bound NADH substrate analog. (A) The Scabin·NADH structure is represented as cartoon and coloured as shown. NADH is shown as sticks and carbon atoms are coloured green. (B) Primary sequence of Scabin with peptides obtained from HDX of Scabin (35 μM) in complex with NADH (250 μM) after 4.14 s reaction time. Deuterated sample was quenched with pH 2.5 acetic acid, with subsequent digestion performed in the microchip using immobilized pepsin crosslinked with agarose. Black colouring indicates deuterium loss due to back-exchange of the first two residues. Secondary structure is shown, with helices represented as blue cylinders and strands represented as green arrows; number indicates order in structure. PN-loop and ARTT-loop are both outlined. See Materials and Methods for details on the system.
ribosyl-turn-turn (ARTT) loop, consisting of the catalytic Q158-X-E160 residues. Though exact peptides were not present in both free and bound states, an approximation can be made using neighbouring peptides. An adjacent region exhibited a change in deuterium uptake of +17%, decreasing from $26 \pm 2\%$ ($x_{161-166}$) to $9 \pm 3\%$ ($x_{159-163}$). This decrease in deuterium uptake is due to the close approach of the nicotinamide-ribose, preventing accessibility of deuterium to the amide backbone of $\beta4$. Unfortunately, a section of the ARTT-loop ($x_{154-158}$) was not covered in both free and bound data sets and thus detailed conclusions of this region (coloured black in Figure 5.2A) were not possible. More replicates should be collected to obtain uptake values for this region.

The S117-T118-T119 motif – involved in stabilizing NAD$^+$ within the active site – exhibited no significant change in deuterium uptake ($\Delta$HDX $x_{115-125} = +2\%$). The main-chain of this region does not directly contact the NADH substrate; the only restriction involves the stabilizing effect of Ser117 through a weak CH-bond with the pyridinium ring of nicotinamide and a bridged hydrogen-bond between HOH41 and the A-phosphate (Fig. 4.1E).

Regions around the PN-loop ($x_{128-132}$) exhibited an increase in protection from deuterium exchange upon NADH binding. Regions around Trp128 of the PN-loop exhibited a change of $+10\%$ ($x_{128-131}$); downstream of Trp128, peptide 130-134 showed a change in deuterium uptake of $+16\%$. The PN-loop is involved in stabilizing NADH within the binding site, specifically the phosphate and nicotinamide moieties. Trp128 interacts weakly with neighboring residues in the apo-form ($E_{\text{int}} = -1.5 \text{ kcal/mol}$), with a side-chain rotation observed upon NADH binding. Trp128 side chain rotates nearly $180^\circ$ upon NADH binding, shifting the nitrogen of the indole ring 4 Å. This results in an increase in the steric contact with NADH ($E_{\text{int}} = -6.3 \text{ kcal/mol}$, purple shells in Fig. 4.1F), inducing a shift in HOH171 and leading to a greater stabilization by coordination with
NADH. This increase in steric contact would make the main-chain less accessible for deuterium uptake upon thermodynamic stabilization by NADH binding. The β-strand C-terminal to the PN-loop (β3) exhibited a slight increase in deuterium uptake (ΔHDX x_{132-138} = -6%), probably due to the strong increase in exposure of the adjacent strand in the β-sheet core (β1).

The most surprising change observed upon binding NADH was in β1, containing the catalytic Arg77. In Figure 5.2B, a large increase in deuterium uptake is shown for x_{69-77} = -34%; an overlapping peptide, x_{68-72}, shows a decrease in deuterium uptake (ΔHDX = +11%), suggesting that the region of the peptide x_{69-77} towards the C-terminal end exhibits the highest increase in exposure upon NADH binding. Comparing the crystal structure of the apo- and NADH-bound states reveals no obvious reason for such an increased exposure upon NADH binding. The increased exposure suggests that β1 has lost ordered secondary structure upon NADH binding. The data for the apo-Scabin HDX experiment have already revealed discrepancies between the crystal and solvent structures, specifically in the N-terminus; thus, it is difficult to analyze the HDX data without understanding the intrinsic dynamics displayed by the protein during binding. Future studies will focus on the use of molecular dynamic simulations of Scabin in the ligand-bound state to understand this significant change in deuterium uptake.

Helix α2 exhibited a decrease in deuterium uptake upon NADH binding (ΔHDX x_{80-94} = +13%; ΔHDX x_{83-87} = +24%). Despite the low RMSD of the entire protein upon NADH binding (Cα-RMSD = 0.27 Å for 165 residues), this region (residues 81-88) has a relatively higher Cα-RMSD of 0.32 Å; the side chains of the region are also conformationally shifted (RMSD = 0.7 Å). Arg81 is shown to shift in conformation and stack with the adenine moiety of NADH, forming a cation-π interaction (Fig. 4.1C).
5.2.3: HDX of Scabin with DNA –

HDX-MS was performed on Scabin in complex with a 21-base, double-stranded DNA oligomer containing a single-base overhang that was shown to be a good transferase substrate; Scabin exhibited the highest affinity for this oligomer (Chapter 4), and was used for further studies. A single time-point (4 s) was collected for all experiments; three biological replicates were performed for apo-Scabin, with each replicate containing \( n = 3 \) or 4 individual runs, allowing for data averaging for replicate peptides. Total confirmed sequence coverage was 89% (excluding expression tag), with total sequence coverage including unconfirmed peptides at 92%. A cartoon representation of the change in deuterium uptake compared to the apo-state (\( \%\text{HDX}_{\text{free}} - \%\text{HDX}_{\text{bound}} \)) mapped onto the Scabin-NADH crystal structure (PDB ID 5TLB) is shown in Fig. 5.3A. It is important to note that NADH was not present during experimentation; the crystal structure is only used for reference to the binding cleft.

The most significant increase in protection from deuterium-exchange was observed in the PN-loop; Trp128 is located within this region, which participates in DNA binding and transferase activity (Chapter 4). When Trp128 is substituted for a tyrosine, the enzyme exhibited a higher \( K_M \) (1399 \( \mu \)M; 4.6-fold increase), a 3.6-fold loss in turnover number (23 min\(^{-1}\)) and a 9-fold weaker affinity for dsDNA with a \( K_D \) of 458 \( \mu \)M. Two peptides covered this region, which exhibited a \( \Delta\text{HDX} \) for \( x_{128-131} \) and \( x_{130-134} \) of 8 and 30%, respectively. Trp128 is predicted to strongly interact with backbone atoms of consecutive nucleotides in the modelled ternary complex of Scabin-NAD\(^+\)-DNA through water-mediated hydrogen-bonding (Fig. 4.5D). Tyr129 is also located in the PN-loop and is predicted to intercalate into the minor groove of dsDNA, potentially allowing for target sequence specificity of Scabin (Fig. 4.5E). Again, a section of the ARTT-loop (\( x_{154-158} \)) was not covered in both free and bound data sets and thus specific conclusions of this region –
A

B

C

MGSSHHHHHHHSSENNLYFQGSHTATATSAAKAAA

PACPRFDDPVHAADPRVDVERITPDPVWRTTCGLYRS

DSRGPAVVVEQGFPLKDVIDGQYDIIESVLVQLNPSPYVISTT

YDHDLYKTWWYKSGNYVDAPGGVDNKTIQDRHKWADQ

VEVAPGGIRTEFVIGCPVDKTRTEKMSCVGNPHYPWELH

Δ -20%  
More exposed

Δ 0%  
No change

Δ 30%  
More obstructed

No data
FIGURE 5.3: HDX of Scabin bound with dsDNA substrate. (A) Scabin·NADH structure represented as cartoon and coloured as shown. NADH shown as sticks and carbon atoms coloured green. (B) Interaction surface of Scabin with DNA. Residues that approach DNA within 5 Å are shown as sticks with carbon atoms coloured based on deuterium uptake. DNA is represented as transparent surface, and is coloured pale yellow. (C) Primary sequence of Scabin with peptides obtained from HDX of Scabin (35 μM) in complex with DNA (140 μM) after 4.14 s reaction time. Deuterated sample was quenched with pH 2.5 acetic acid followed by subsequent digestion performed on a microchip using pepsin crosslinked with agarose. Black colouring indicates deuterium loss due to back exchange of the first two residues. Secondary structure is shown, with helices represented as blue cylinders and strands represented as green arrows; number indicates order in structure. PN-loop and ARTT-loop are both outlined. See Materials and Methods for details on the system.
including W155, involved in transferase activity – were not possible. More replicates should be collected to obtain deuterium uptake values for this region as the ARTT-loop confers protein substrate specificity in mART toxins (111). The α1 helix also shows some protection from deuterium exchange (ΔHDX x_{80-94} = +12%) upon DNA binding; this helix was not originally predicted to be involved in DNA binding, but should be considered for future molecular dynamic simulations.

In Figure 5.3B, residues that approach the dsDNA within 5 Å are shown as sticks. This Scabin-DNA model is the same as shown in Chapter 4, and does not consider the newly acquired HDX data. Residues in the PN-loop (W128, Y129, K130) are oriented to face the dsDNA substrate. Importantly, the dsDNA structure does not account for any perturbations of DNA conformation when bound to Scabin. A recently solved NMR solution structure of PARP-1 bound to DNA containing a single-stranded break was modelled to reveal the protein-DNA interface (57). The single-strand nicked DNA adopts a twisted structure, exposing the 3’ and 5’ bases within the nicked duplex. This facilitates several hydrophobic interactions with the exposed DNA bases, including a π-stacking interaction with an exposed 3’ guanine and Phe44 (57). Figure 5.4, adapted from Dawicki-McKenna et al. (2015), depicts the interaction of PARP-1 and its DNA binding partner. The observed protection from deuterium exchange of Tyr129 in the Scabin-DNA complex may suggest a comparable role to Phe44 in PARP-1 (57, 70). Since Scabin exhibits a higher affinity towards dsDNA containing a single-base overhang (a model for SSBs), future studies using MD simulations should incorporate the NMR structure of the SSB modelled in the PARP-1 complex.

Analysis of the crystal structure of Scabin reveals no typical DNA-binding motifs. From the HDX data, the PN-loop plays a significant role in the interaction of Scabin with DNA.
FIGURE 5.4: Interaction of PARP-1 with a single-strand break (SSB) DNA. (A) PARP-1 represented in cartoon, helices coloured green, strands coloured teal and loops coloured yellow; Zn represented as spheres, coloured gray. DNA backbone coloured orange. Dashed black box leads into (B) depicting Phe44 (carbon atoms coloured yellow) π-stacking with guanosine (carbon atoms coloured pink) base of exposed 3’ DNA; residue and base represented as sticks. Adapted from Dawicki-McKenna et al. (2015); PDB ID 2N8A (70).
However, secondary structures around this loop do not resemble a typical DNA-binding motif. Importantly, Scabin may alter its secondary structure upon interaction with DNA, as shown for the GCN4 transcription factor (56). The DNA-binding region of GCN4 is intrinsically disordered in the apo-form, and upon ligand binding, conformationally shifts into a traditional leucine zipper motif (56). The leucine zipper domain of GCN4 contains a C-terminal, leucine-rich, helical domain that promotes dimerization, and an N-terminal basic region. Upon DNA binding, the basic region – which has residues with the propensity to form helices – adopts an ordered helical structure that fits into the major groove of DNA (56). It is highly likely that the static crystal structure of Scabin does not reflect the true dynamic solution structure during catalysis, and thus molecular dynamic simulations of Scabin in solution are invaluable for further understanding of the DNA-binding and transferase mechanism for this novel mART toxin.

5.3: Conclusion

HDX experiments of apo-Scabin revealed discrepancies between the observed disordered N-terminus of the crystal structure and the ordered structure in solution, suggesting that the loop region in solution may fold to form a hairpin-like structure. Upon NADH binding, the catalytic Glu160 and PN-loop exhibit reduced deuterium uptake in support of the Scabin crystal structure (PDB ID: 5TLB). A large increase in deuterium uptake was observed for β1, which contains the catalytic Arg77. This suggests that β1 experiences local unfolding upon NADH binding, or possibly a reorganization of the β-sheet core. The PN-loop participates in DNA substrate binding and enzyme activity, representing the first clue of a DNA-binding motif in a mART toxin. Further structural studies in combination with computational chemistry (e.g., molecular dynamic simulations) are necessary to understand the dynamic properties of Scabin as a DNA-targeting enzyme, and to reveal the DNA-binding motif of this toxin.
CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS
6.1: Conclusions

mART toxins are proving to be a diverse family of enzymes that can modify both proteins and DNA, creating damaging adducts that are involved in disease progression in vertebrates, insects and plants. Only a few members of the mART toxin family are recognized as DNA-targeting, with Pierisin-1 being the first annotated member of this unique subgroup. No structural or in-depth kinetic study of Pierisin-1 has ever been published. It is therefore important to investigate Scabin toxin as a model for the catalytic mechanism of the DNA-targeting mART toxin subgroup. On a broader scale, characterizing these toxins may lead to novel therapeutic compounds and in turn, development of alternative strategies for treating bacterial infections.

In this thesis work, Scabin was identified and characterized as a novel DNA-targeting mART from the plant pathogen 87.22 strain of *Streptomyces scabies*. This putative toxin has nearly 40% pairwise sequence identity with Pierisin toxins, and possesses the characteristic RSQXE motif found in the mART toxin family. The Scabin gene was expressed and the 21.7-kDa Scabin protein was purified from *E. coli*, and both the GH and transferase activities were characterized. Several small molecule inhibitors were also identified; two inhibitor co-crystal structures were solved, providing important new insights into the development of antivirulence compounds against the Pierisin-like mART toxin subgroup. Apo-Scabin was crystallized and it diffracted to 1.5 Å; this structure was the first 3-dimensional model of a DNA-targeting mART enzyme, which revealed important insights into mART toxins that target DNA substrates.

The crystal structure of Scabin with NADH was also solved at high resolution; this represents the first in-depth analysis of the Michaelis complex for a DNA-targeting mART toxin. Through kinetic analysis of Trp to Ala variants, Trp128 and Trp155 in Scabin were shown to be important for binding and transferase activity with the DNA substrate. The crystal structures of
the Trp variants were solved, as well as N110A and S117A variants, providing new insights into the roles of these residues in both NAD$^+$ and DNA substrate binding and catalysis. The roles of catalytic residues (Arg77, Asn110, Ser117, Gln158, Glu160) were also investigated and combined with a rigorous kinetic analysis of Scabin with model DNA substrates, furthering our understanding of the structure and function of the mART toxin Pierisin-like subgroup. A model of the Scabin-DNA complex was built, which accounts for the observed binding and kinetic interactions of Scabin with the DNA substrate.

Lastly, hydrogen-deuterium exchange coupled with mass spectrometry was used to characterize the protein-DNA interface of Scabin. Initial experiments of apo-Scabin based on deuterium uptake revealed a relatively protected core with ordered peripheral loop regions. The N-terminus exhibited a relatively ordered conformation in solution, suggesting that the Scabin crystal structure may have some differences from the solution conformation. Deuterium exchange of the Scabin-NADH complex was used to compare the dynamics of the complex in solution and in the static crystal structure. The catalytic Gln158-X-Glu160 motif was significantly protected upon NADH binding, whereas the catalytic Arg77 exhibited a significant increase in deuterium uptake. This increase in exchange rate around Arg77 may reflect a reorganization of the β-sheet core or indicate a change in folding pattern of the N-terminus in solution, not seen in the crystal structure. Further studies will involve computational chemistry including molecular dynamic simulations to map conformation changes during substrate binding, which may better explain the HDX data. Deuterium uptake for Scabin in the presence of DNA was mapped onto the modelled ternary complex. The PN-loop exhibited a significant decrease in exchange rate upon DNA binding, and is likely an important interface for substrate recognition and binding. Unfortunately, the ARTT-loop was not found in any pepsin-generated peptides; future experiments will focus on
collecting exchange data on this region, which may participate in the protein-DNA interaction. No traditional DNA-binding motif could be identified in Scabin. However, the DNA-binding motif may adopt its active-conformation and structure only upon ligand binding, as reported for GCN4 (56). Future HDX experiments at different reaction time-points may improve the resolution of the putative DNA-binding domain of Scabin.

6.2: Future directions

Several possibilities exist for further characterization of Scabin as a potential virulence factor in *Streptomyces scabies*. Scabin represents the first bacterial mART toxin that targets DNA; further structural studies are required to understand this mechanism. Currently, we know little about the role of Scabin in *S. scabies* pathogenesis, or the potential gene target(s) of this enzyme within host cells. A few methods for future characterization of Scabin are listed below.

6.2.1: Characterizing the macromolecular complex of Scabin with DNA –

Obtaining structural data on macromolecular complexes is often a challenging task. The relatively low affinity of Scabin for DNA substrate poses an issue for obtaining crystals of the Scabin-DNA complex. The transient nature of the complex would likely prevent the formation of an ordered crystal lattice. Scabin would presumably exhibit a higher affinity for its potential target gene(s) within the host genome. However, unless the target for Scabin can be determined through methods such as selection and amplification binding (SAAB) assay (128) or the ChIP-sequence method (129) using *Solanum tuberosum* genomic DNA, then other means such as more comprehensive modelling of the protein-DNA complex should be pursued. As discussed in Chapter 1, small-angle X-ray scattering (SAXS) could be used in conjunction with hydrogen-deuterium exchange experiments. HDX and SAXS experiments on the Scabin-DNA complex may produce a low-resolution model (10 Å) of the protein-DNA complex.
Further HDX-MS studies of Scabin can reveal dynamic changes during ligand binding. Conformational dynamics can be monitored by measuring deuterium uptake for several D$_2$O reaction time-points (90). This method may facilitate the identification of a new DNA-binding motif within Scabin. Additionally, the apparent increase in disorder of $\beta$1 containing the catalytic Arg77 upon DNA binding can be further studied. Identifying a DNA-binding motif in Scabin would represent a unique feature within the mART toxin family, and it is therefore of great interest to pursue structural studies.

6.2.2: Identifying the target gene for Scabin –

Scabin binds DNA with rather low affinity. The oligomer used in the experiments described in this thesis represents an arbitrary sequence that contains at least one guanine base. Scabin should target a gene within the plant host, which would represent its enzymatic and pathogenic target. Identifying this gene may reveal the potential role that Scabin plays in *Streptomyces scabies* pathogenesis, and may also increase the likelihood that Scabin would crystallize with an oligomer containing this target gene sequence. As shown in Appendix I, Scabin alters the morphology of mitochondria within treated *Arabidopsis thaliana* seedlings in a dose-dependent manner. The clustering effect that is observed has been documented in the reactive oxygen-species response (130) of *Arabidopsis*, as well as in *fnt* (FRIENDLY mitochondria) knockouts (131). The *fnt* gene encodes a protein involved in mitochondria organization and distribution along the F-actin cytoskeleton of eukaryotes, including plants and yeast (131). Knockouts of *CLU1*, the *fnt* homolog in *Saccharomyces cerevisiae*, do not exhibit a decrease in growth rate, which would explain the observation that Scabin shows no cytotoxicity in our yeast model system (data not shown) (83, 132). In plants, *fnt* knockouts exhibit decreased growth in roots; after 28 days of growth, knockouts have a root length that is half of the wild-type length.
Using agro-infiltration, we have also shown that transient expression of Scabin leads to a reactive oxygen-species response (Appendix I), which could also explain the observed clustered phenotype.

Currently in the lab, a selection and amplification binding (SAAB) assay is being optimized to determine the binding site sequence for Scabin (134). Scabin will be expressed with a GST-fusion tag, purified from whole-cell lysate, and bound to a glutathione resin (128). A generated ‘library’ of short template sequences will be passed over the bound protein, in hopes that the specific sequence will bind with higher affinity and out compete non-specific binding by other sequences. The complex can then be isolated and the bound template sequence can be amplified via PCR. The selection cycle can be repeated with the PCR-amplified DNA to re-select the bound template. The amplified DNA can then be cloned into a vector for sequencing (128, 134). A global genome search of Arabidopsis for this sequence can be made, and potential target genes containing the sequence of interest can be identified.

6.2.3: The role of Scabin in Streptomyces scabies pathogenesis –

As shown in Appendix I, Scabin may affect both mitochondria morphology in Arabidopsis and mammalian cell morphology. Further characterization and quantification of these observations should be pursued. A collaboration with Dr. Jaideep Mathur’s lab at the University of Guelph provided preliminary characterization of the biological role of Scabin in S. scabies pathogenesis. Experiments involving transgenic Arabidopsis that expresses GFP-tagged mitochondrial ATPase has been performed; treating seedlings with purified Scabin at different concentrations produced the clustered mitochondria effect. Further studies involving other transgenic lines of Arabidopsis that involve tracking changes in organelle morphology should be pursued, as well as treatment with catalytic variants of Scabin to ensure the effects are due to Scabin enzyme activity.
Morphology changes of mammalian cells treated with Scabin have also been observed. In comparison to the negative control, the cells appear to be extended and cell division does not occur at later time points. Cytotoxicity assays should be pursued to quantify the activity of Scabin in mammalian cells; notably, Pierisin-1 kills HeLa cells via induction of apoptotic pathways (45). Overall, understanding the role of Scabin in *S. scabies* pathogenesis may provide a more tailored approach for the development of anti-virulence compounds.

6.3: Concluding remarks

The research documented in this thesis provides the first mechanistic glimpse of Scabin as a DNA-targeting mART toxin. Notably, the bioinformatic *in silico* strategies for identifying new mART toxins should be modified to include DNA-targeting mART toxins which may lead to a large-scale expansion of the Pierisin-like subgroup within this toxin family. Thus, it is now evident that further study of Scabin features and mechanisms will provide important new insights into the pathogenesis associated with DNA-targeting mART toxins and may prove helpful to understanding tuberous crop diseases such as the common scab in potatoes.
REFERENCES


94. Cheng, Y., and Prusoff, W. H. (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* 22, 3099–3108


APPENDIX
A.1: Methods

A.1.1: Microfluidic chip fabrication –

The microfluidic chip was made on a poly (methyl methacrylate) (PMMA) substrate with dimensions (8.9 cm × 3.8 cm × 0.6 cm); the proteolytic chamber and capillary chambers were etched onto the chip using a VersaLaser engraver (Universal Laser, Scottsdale, AZ). The time-resolved electrospray ionization (TRESI) mixer was made by inserting a glass capillary (152 µm outer diameter) into a metal capillary (inner diameter 178 µm) to create an intercapillary space of 26.8 µm. The end of the glass capillary was sealed using the VersaLaser, and a notch was cut 2 mm from the sealed end. For reaction quenching, a T-mixer with a dead volume of 51 nL was used to mix glacial acetic acid with deuterated protein. A schematic of the device is shown in Fig. 2.1 in Chapter 2. Pepsin-agarose beads were activated for 1 h with hydrochloric acid (pH 1.8), followed by 1 h incubation with acetic acid (pH 2.3). The proteolytic chamber was filled with pepsin-agarose beads, and a 33G metal capillary was used as an outlet to the electrospray ionization source. Hamilton syringes were used to deliver reagents through the glass capillaries using Syringe infusion pumps (Harvard Apparatus). The device was placed at the front end of a modified QSTAR Elite hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer (Sciex) for HDX-MS experiments.

A.1.2: Determination of the percentage of free thiols in Scabin variants –

To determine the percentage of free thiols of each protein, a fluorescence-based assay was developed using monobromobimane (MBrB). Samples consisting of 15 µM protein and 40 µM MBrB in 25 mM Tris-HCl, pH 8.2, and 200 mM NaCl were allowed to incubate at room temperature for 1.5 h. Fluorescence was measured using a FLUOstar Omega plate reader; excitation and emission filters were set to 350 and 460 nm, respectively. A standard curve was
prepared using varying concentrations of cysteine hydrochloride monohydrate (0-12.5 µM) and 40 µM MBrB.

A.1.3: Derivation for equation 2 used for representing DNA binding of Scabin –

Derivation for equation 2:

\[
[DNA]_t = [DNA]_f + [DNA]_b
\]

\[
[E]_t = [E]_f + [E]_b
\]

\[
[E]_b = [DNA]_b
\]

\[
K_d = \frac{([E]_t - [DNA]_b).([DNA]_t - [DNA]_b)}{[DNA]_b}
\]

solving for \([DNA]_b\), the exact bound concentration of DNA for each point of the titration is:

\[
[DNA]_b = \frac{([E]_t + [DNA]_t + K_D) - \sqrt{([E]_t + [DNA]_t + K_D)^2 - 4[E]_t[DNA]_t}}{2}
\]

where \([E]_t\) is the total concentration of toxin at each point and \([DNA]_t\) is the total concentration of DNA at each point.

A.1.4: Arabidopsis thaliana growth conditions –

Seeds of Arabidopsis thaliana (ecotype Columbia) were sterilized in 20% bleach solution and plated on Murashige and Skoog (MS) medium with 3% sucrose, buffered to pH 7.5 (135). Seeds were stratified at 4°C for 3 days and then grown at 21°C for 7 days at 70 µmol m\(^{-2}\)s\(^{-1}\) under a 16/8 h light dark cycle.

A.1.5: Treatment of G-mito Arabidopsis thaliana seedlings with Scabin toxin –

Seeds of Arabidopsis thaliana transgenic expressing mitochondrial targeted GFP-linked β-ATPase subunit were prepared and plated as described under ‘Arabidopsis thaliana growth conditions’. Seedlings were harvested after 7 days and transferred to petri-dishes with 2 mL of treatment, containing varying concentrations of Scabin toxin (0, 0.1 or 0.5 µM) in MS (pH 7.5)
with 3% sucrose. Seedlings were allowed to soak in treatment for 5 minutes, with the roots subsequently placed on filter paper dipped in treatment and cotyledons away from filter paper, as shown in Figure A.1. Petri dishes were kept vertical to allow diffusion of treatment up the filter paper (Fig. A.1A), with dishes placed back in the 21°C chamber with 16/8h light dark cycle. After 24 and 48 h, seedlings were imaged on a Leica TCS SP5 confocal microscope using an Argon (488 nm) excitation laser to visualize GFP (emission band 503-524 nm) while chlorophyll auto-fluorescence was collected between 660 and 750 nm. Experiments were performed with 5 replicates per time point.

A.1.6: Transient expression of Scabin in potato and tobacco –

Tobacco (*Nicotiana benthamiana*) and potato (*Solanum tuberosum*) plants were grown under greenhouse conditions (16 h light/24 °C and 8 h darkness/20 °C; and light intensity at 110 μmolm⁻² s⁻¹ (LI-250 A, LI-COR; Biosciences, Lincoln, NE, USA) for 4 weeks. For agro-infiltration experiments, we used the binary vector pGreen containing a fusion between the green fluorescent protein (GFP) and the Scabin or ExoA genes driven by the dual 35S promoter of *Cauliflower mosaic virus*. These binary plasmids were transformed into *A. tumefaciens* strain GV3101, which harbors the transformation helper plasmid pSoup (136). The Agro-infiltration of tobacco and potato by different gene constructs was conducted as described (137).

A.1.7: Visualization of H₂O₂ accumulation –

To visualize H₂O₂ accumulation in the infection site of infection, 3,3′-diaminobenzidine (DAB) staining was performed as described by (138). Briefly, leaf discs (1 cm²) around inoculation sites were harvested at 48, 96, and 122 hpi. For each time point, three discs originating from three different plants were used. For H₂O₂ accumulation, potato and tobacco leaf discs were placed for three h in a solution of 1 mg/ml DAB-HCl (pH 4) and then fixed in 100% ethanol before being
**FIGURE A.1:** Experimental set up for ‘Treatment of G-mito *Arabidopsis thaliana* seedlings with Scabin toxin’.
visualized by the bright-fielded microscopy.

A.1.8: Structure preparation and molecular mechanics calculations –

Protein preparation, molecular mechanics calculations, and protein rendering were performed using the computational suite MOE (Molecular Operative Environment) release 2014.08 (Chemical Computing Group Inc., Montreal, Canada). The force field employed was Amber12:EHT, with the AMBER12 parameter set (ff12) for protein and parameters calculated from the extended Huckel theory for P6E, PJ34, NAD+, and dsDNA molecules. The MOE Protonate3D module was used to assign the ionization states and tautomers of protein side chains at T = 300 K, pH 7.4, and 0.1 M ionic strength, along with the GB-VI solvation model and MMFF94 partial charges. The molecular surfaces are solvent-excluded surfaces obtained by rolling a probe sphere of 1.4-Å diameter (water radius) and coloured by several schemes. The van der Waals interaction surfaces correspond to zero-potential contours of the van der Waals potential, $E_{vdw} = 0$, between the considered set of atoms and a water oxygen atom as mobile probe, using a standard 12-6 Lennard-Jones definition.

A.1.9: Modelling wild-type and variant Scabin complexed with NAD+ –

Using the Scabin-NADH preparation as a template, the various Scabin proteins (WT, N110A, Q158AxE160A, S117A, W68Y and W68A, W128Y, W155A and W199A) were modelled by ‘mutating in silico’ the selected residue to the target residue with an optimum side-chain conformation obtained from an all-atom backbone-independent MOE rotamer library. In each case, the NADH molecule was crafted in situ to NAD+. The Protonate3D protocol was performed while protecting the oxidized state of the ligand, and new complexes were repacked by tethering the NAD+ molecule (10 kcal/mol, 0.25 Å buffer) and all the heavy-atoms (100 kcal/mol) except the mutated residue, all side-chains, and CWMs with atoms at $\leq 4.5$ Å distance from the
mutated residue or from the nicotinamide moiety. An energy minimization calculation was then conducted until a RMS gradient $\leq 0.001 \text{ kcal/mol/Å}^2$ was achieved.

A1.10: Modeling the Scabin·NAD·DNA complex –

A WT Scabin-apo structure with active conformation for pocket residues was generated in silico by stripping the Scabin-NADH structure (PDB: 5TLB) from NADH and CWMs. A 21-mer double stranded DNA (5’-GGAAGAGAGAGAGAAAGAGAG-3’, forward strand) was built in a B-helix conformation. The force-field used was again Amber12:EHT to parametrize the NAD$^+$ and the dsDNA molecule. A coarse-grained MOE protein-protein docking was performed using the dsDNA molecule as a substrate. For the five highest-ranked Scabin-DNA decoys, the NADH ligand and in pocket CWMs taken from the X-ray structure (PDB: 5TLB) were included. To each ternary decoy, NAD$^+$ was crafted from NADH and was subjected to the preparation protocol (see above). The ternary decoys were further restrained (10 kcal/mol, fixed NAD$^+$ molecule); however, free CWMs and contact (protein-DNA interface) side-chains and loop backbone atoms, were further energy-minimized until a RMS gradient $\leq 0.001 \text{ kcal/mol/Å}^2$ was achieved. The decoy with the lowest energy-score was selected for further analysis and rendering.

A1.11: Spectral decomposition of the Scabin Trp emission fluorescence –

A normalized emission spectrum of NATA centered at $\lambda_{\text{max},\text{NATA}}$ was used to obtain the set of parameters $\mathbf{p}$ that best fit an arbitrary $7^{th}$ polynomial $\mathbf{f}$ function of $\lambda-\lambda_{\text{max},\text{NATA}}$ as the independent variable, $\mathbf{f}_p(\lambda-\lambda_{\text{max},\text{NATA}})$. The parameters that defined the function $\mathbf{f}$ were then employed to develop four composite functions, $F_i(\lambda)$, that correspond to the normalized emission spectra of four $n$-1 Scabin Trp variants, as follows:

$$F_{\text{WT}}(\lambda) = M_{W68}\mathbf{f}_p(\lambda-\lambda_{\text{max},W68}) + M_{W128}\mathbf{f}_p(\lambda-\lambda_{\text{max},W128}) + M_{W155}\mathbf{f}_p(\lambda-\lambda_{\text{max},W155}) + M_{W199}\mathbf{f}_p(\lambda-\lambda_{\text{max},W199})$$
\[ F_{\text{W128Y}}(\lambda) = M_{\text{W68}} \cdot f_p(\lambda-\lambda_{\text{max,W68}}) + (\text{missing Trp signal, n-1}) + M_{\text{W155}} \cdot f_p(\lambda-\lambda_{\text{max,W155}}) + M_{\text{W199}} \cdot f_p(\lambda-\lambda_{\text{max,W199}}) \]

\[ F_{\text{W155A}}(\lambda) = M_{\text{W68}} \cdot f_p(\lambda-\lambda_{\text{max,W68}}) + M_{\text{W128}} \cdot f_p(\lambda-\lambda_{\text{max,W128}}) + (\text{missing Trp signal, n-1}) + M_{\text{W199}} \cdot f_p(\lambda-\lambda_{\text{max,W199}}) \]

\[ F_{\text{W199A}}(\lambda) = M_{\text{W68}} \cdot f_p(\lambda-\lambda_{\text{max,W68}}) + M_{\text{W128}} \cdot f_p(\lambda-\lambda_{\text{max,W128}}) + M_{\text{W155}} \cdot f_p(\lambda-\lambda_{\text{max,W155}}) \]

with \( M_i \) and \( \lambda_{\text{max,i}} \) representing the molar fluorescence and maximum emission wavelength length of each Trp probe, respectively. From a global fit of the four \( C_i \cdot F_i(\lambda) \) functions to the experimental emission spectra of the four proteins, with \( p \) as shared parameters and \( C_i \) the concentration of each protein, were obtained the spectral properties (\( M_i \) and \( \lambda_{\text{max,i}} \)) that describe each individual Trp were obtained.

### A.2: Biological studies of Scabin

Each of the following preliminary biological studies involving Scabin will be presented with a brief description of result/discussion within each figure title. Each figure has been referred to separately within the main body of the thesis, with reference shown in the respective figure title.
A.2.1: Reactive oxygen response induced by transient expression of Scabin –

**FIGURE A.2:** Scabin induces cellular stress response in *Solamum tuberosum*. (A) DAP staining at 96 hours post-injection (hpi); from left to right, negative control: pGreen vector, positive control: pGreen+ExoA, treatment 1: pGreen+Scabin, treatment 2: pGreen+Q^{158}AxE^{160}A. (B) DAP staining at 122 hpi. Agroinfiltration was used to deliver vectors into plant cells.

Results and Discussion: Transient expression of Scabin in tobacco and potato leaves resulted in a reactive oxygen species response. This effect proved to be due to the catalytic activity of Scabin, as evident by the decrease in response of the Q^{158}AxE^{160}A variant treatment. ExoA toxin (from *P. aeruginosa*) induced an ROS response, as expected. ExoA targets eEF2; disruption of this vital cellular process would lead to a stress response in the host cell (139). From these data, we can conclude that Scabin catalytic activity induces a stress response, with a greater response observed in potato leaves. See Chapter 6 for the in-text discussion.
A.2.2: Scabin treatment of Arabidopsis thaliana seedlings –
FIGURE A.3: Scabin induces a clustered mitochondria response in G-mito *Arabidopsis thaliana* seedlings. A-C is 24 h after treatment, D-F is 48 h after treatment. (A,D) Control treatment, MS (pH 7.5) with 3% sucrose. (B,E) 0.1 µM Scabin treatment, (C,F) 0.5 µM Scabin.

Results and Discussion: Scabin induced a clustered mitochondria response at nanomolar concentrations. This effect can be attributed to either an ROS response, or a change in expression levels of *fmt*. Scabin may be targeting the gene that codes for FRIENDLY, a protein that is involved in homogenous distribution of mitochondria throughout the cell (133). See Chapter 6 for the in-text discussion.
A.2.3: Summary for roles of residues in binding pocket –

**FIGURE A.4:** Proposed map of binding pocket for ternary complex of Scabin. Important residues are highlighted, including Arg77, Asn110, Trp128, Trp155, Q158 and E160. Hydrogen bonds are depicted as black dashes; π-stacking interactions are depicted by gray dashes. The N2 substituent that acts as the nucleophile to attack the C1 of N-ribose is highlighted. dA is an arbitrary nucleotide adjacent to dG in the bound oligomer. Trp128 is more important for interacting with the oligomer, whereas Trp155 is required for transferase activity to occur; Trp155 allows for dG to approach Q158, which further activates the nucleophile. Without this activation step, transferase activity cannot occur.