Exploring protein – ligand interactions: From mechanistic insight to inhibitor development

By

Scott Legare

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Chemistry

University of Manitoba

Winnipeg Manitoba Canada

Abstract

Biological processes required for life can be understood by studying the underlying molecular and chemical processes which give rise to these biological processes. Proteins are the primary functional molecules of biological systems where they serve as catalysts, cell signaling molecules, receptors and also serve structural and mechanical roles. In order to perform these various functions, proteins must physically interact with other molecules referred to as ligands. Therefore, developing a detailed understanding of protein – ligand interactions from a chemical and molecular perspective improves the understanding of the biological processes which these interactions give rise to. A detailed understanding of these protein ligand interactions also enables the development of drugs capable of disrupting or modulating protein – ligand interactions for the treatment of diseases. In other cases, the study of protein – ligand interactions can be used to develop new and useful biotechnologies. Many techniques can be useful for studying protein – ligand interactions including structural, computational, and biophysical techniques. Of these, biophysical techniques often provide the greatest deal of foundational insight into the nature of the protein – ligand interactions. In addition, a biophysical characterization of a protein – ligand interaction is often a prerequisite for the development of drugs and biotechnological applications for these protein- ligand interactions.

Three unique protein – ligand interaction systems are explored in this thesis, using a combination of structural, computational, and biophysical techniques. In the first part of this thesis the interaction between the Laminin N-terminal domain and Ca^{2+} is explored using a combination of structural, computational, and biophysical techniques. From this approach, a detailed understanding of the molecular basis and functional properties of the interaction between the Laminin N-terminal domain and Ca^{2+} is developed. In the next part of this thesis, the biophysical and enzymatic characterization of the SARS-CoV-2 Main protease is performed. Guided by this biophysical and enzymatic characterization, an improved substrate for high throughput screening is developed. The final part of this thesis performs a thorough examination of the interaction between Netrin-1 and UNC5B including the requirement for a heparin co-receptor. From this examination of the Nertrin-1 UNC5B interaction, peptide inhibitors of the interaction are rationally designed. Together the work presented in this thesis demonstrates the study of multiple aspects of protein - ligand interactions. It combines studies aimed at gaining an understanding of molecular basis and functional properties of protein – ligand interactions, developing methods to identify and test inhibitors of protein – ligand interactions, and leveraging the gained understanding to rationally design inhibitors of protein-ligand interactions.

Acknowledgements

I would like to thank my supervisor Dr. Jörg Stetefeld for his unwavering guidance, invaluable insights, and endless patience throughout this journey. I am thankful for the support and trust given to me throughout this process. His mentorship has been instrumental in shaping both my research and personal growth.

I would also like to thank my committee members for the support, expertise, constructive feedback and valuable time they have provided me throughout this process.

I am also thankful to my colleagues who have given me help, guidance and support in many forms. This includes mentorship, technical advice, and performing the many tasks required to keep the lab running, materials supplied and instruments in operating order. This also includes the support and encouragement in academic and research areas.

Lastly, I would like to thank my family and friends for their support throughout this process.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMC	7-amino-4-methylcoumarin
BIS-TRIS	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
BLI	Biolayer interferometry
CD	Circular dichroism
COVID-19	Coronavirus disease 2019
C-terminus	Carboxyl terminus
Da	Dalton
DABCYL	4-(4-dimethylaminophenylazo)benzoic acid
DCC	Deleted in Colon Cancer
DLS	Dynamic light scattering
DNP	2,4-dinitrophenyl
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EDANS	5-(2-Aminoethylamino)-1-naphthalenesulfonic acid
EDTA	Ethylenediaminetetraacetic acid
FAB	Fragment antigen binding
FAM	5-carboxyfluorescein
FRET	Fluorescence resonance energy transfer
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HRV-3C	Human rhinovirus 3C protease
HTS	High throughput screen
IC_{50}	Inhibitory concentration 50
IPTG	isopropyl-D-thiogalactoside
ITC	Isothermal titration calorimetry
k_a	Association rate constat
K_a	Association equilibrium constant
k_{cat}	Catalytic rate constant
K_D	Dissociation equilibrium constant
k_d	Dissociation rate constant
K_M	Michaelis constant
Lam-α	Laminin subunit-α
Lam-β	Laminin subunit-β
Lam-γ	Laminin subunit-γ
LE	Laminin epidermal growth factor like
LN	Laminin N-terminal
M	Molar
MCA	7-methoxycoumarin-4-acetic acid
MD	Molecular dynamics
MERS-CoV	Middle East respiratory syndrome coronavirus
MMPBSA	Molecular mechanics Poisson-Boltzmann surface area
M^{pro}	Main protease
MSA	Multiple sequence alignment
NEO	Neogenin
Net-1	Netrin-1
N-terminus	Amino terminus
PDB	Protein data bank
PL^{pro}	Papain-like protease
pp1a	Polyprotein 1a

pp1ab	Polyprotein 1ab
RFU	Relative fluorescent unit
R_g	Radius of gyration
R_H	Hydrodynamic radius
RMSD	Root-mean-squared deviation
SARS-CoV	Severe acute respiratory syndrome coronavirus
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide electrophoresis
TCEP	tris(2-carboxyethyl)phosphine
T_M	Melting point
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
TSA	Thermal shift assay
UNC5	Uncoordinated-5
V_{max}	Maximum catalytic rate
ΔG	Gibbs free energy change
ΔH	Enthalpy change
ΔS	Entropy change
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus

Amino acid abbreviations

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
2-Aminoisobutyrate	Aib	
Norleucine	Nle	

Chapter 1

Introduction

1.1

Protein biochemistry and protein – ligand interactions

1.1.1 Proteins and their role in biological function

Living organisms are capable of carrying out many diverse and complex processes such as cellular respiration, photosynthesis, cell growth, motor movements, and making synaptic connections essential for complex thought and memory. These life processes are the result of a complex series of well-regulated molecular interactions and chemical reactions. Understanding these intricate life processes therefore necessitates comprehension of these underlying molecular and chemical processes. The chemical species involved in these molecular and chemical processes can be classified in related groups of molecules: proteins, nucleic acids, carbohydrates, lipids, small organic molecules, inorganic ions and water. The chemical composition of a typical bacteria cell reveals that 15% of the cells weight is composed of proteins, while nucleic acids, carbohydrates and lipids make up 7, 3 and 2% respectively. The remaining components of the cell are 70% water and 3% small organic molecules and inorganic ions (1). The chemical composition of a bacterial cell demonstrates that proteins make up the major chemical component which gives rise to this complex biological function. Proteins can be incredibly diverse and serve many different functions in a cell. Enzymes are proteins that function as catalysts to facilitate and regulate many of the chemical reactions involved in biology. Proteins often function as signaling molecules which serve to relay environmental messages between cells and coordinate complex multicellular processes. As an example, figure 1.1 shows the many protein – ligand interactions involving the cell signaling molecule Netrin-1. Other proteins are required for mechanical purposes such as adhesion and structural support. Given the crucial role proteins play in life's chemical processes, it is essential to study molecular mechanisms through a protein-centric lens to grasp the foundations of life at the molecular level.

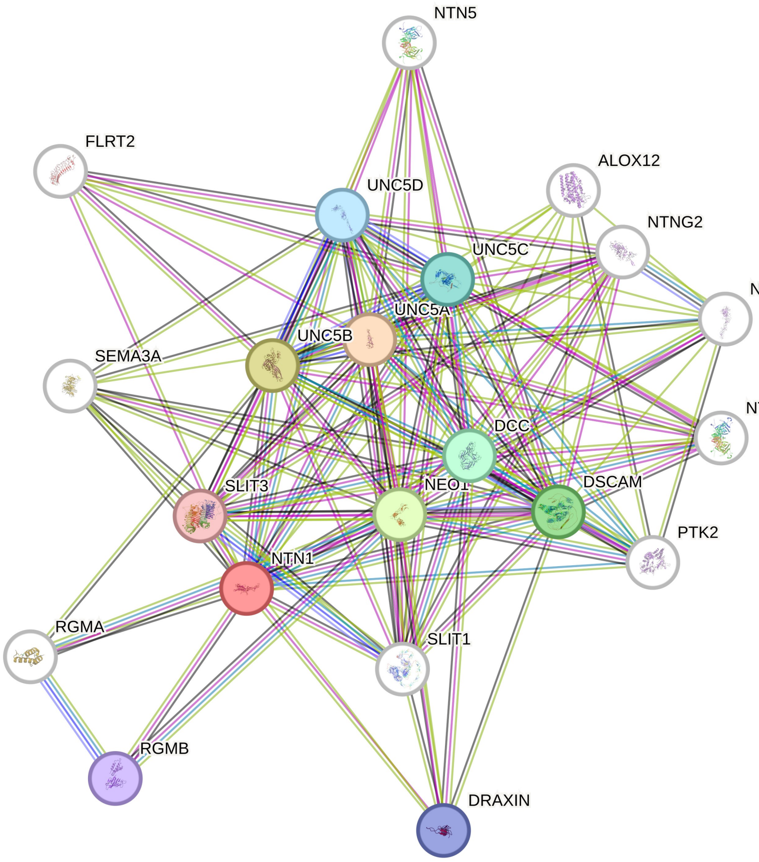


Figure 1.1. Functional association map of *Homo sapien* Netrin-1 (NTN1 red circle) protien from the STRING database(2). Map depicts proteins that are known or predicted to be functionally associated with Netrin-1. This includes protiens which physically interact in a protien – ligand interaction with Netrin-1 and proteins that are involved in common cellular processes.

1.1.2 Protein composition and structure

All proteins consist of a linear polymer of amino acids known as a polypeptide. Amino acids are a class of small organic molecules with an amino, carboxyl and a variable R-group bound to a central α-carbon. Amino acids form a linear polymer, which makes up the protein primary structure. The polypeptide can fold into locally ordered segments, such as various types of helices, loops, turns, β-strands and β-sheets, collectively referred to as secondary structure. Next these secondary structure elements arrange in three-dimensional space to form the tertiary structure of the protein. Finally, quaternary structure describes the arrangement of more than one individual polypeptide into an oligomeric complex (3). Interestingly, individual amino acids alone are generally not able to do any of the functions commonly associated with proteins such as chemical catalysis or regulate complex cellular processes. Amino acids can execute the diverse processes of life only when they are integrated into a protein that adopts a particular three-dimensional structure through folding. Because a proteins structure is critical to its biological function, understanding the structure of a protein can offer valuable insights into its biological function.

1.1.3 Protein – ligand interactions

Despite their varied forms and functions, all proteins share a common functional trait which is the necessity to physically interact with other molecules to participate in the intricate chemical events driving biological processes. These interactions, whether an enzyme binding to its substrate, a signaling protein binding to its receptor, or a protein interacting with a Ca^{2+} ion to support its three-dimensional structure, collectively fall under the umbrella of protein-ligand interactions. While structural information can provide a great deal of insight needed to develop an understanding of proteins and their interactions with ligands, structural information often falls short of describing the functional properties of protein – ligand interactions. Understanding protein-ligand interactions requires knowledge about the dynamic properties underlying these interactions which can be obtained using biophysical techniques. Biophysical techniques can be used to measure properties of protein – ligand interactions including binding affinity, specificity, cooperativity, allostery, binding kinetics and binding mechanisms, all of which provide useful insights about protein – ligand interactions (4, 5). The research presented in this thesis is largely focused on using biophysical techniques to understand these biophysical properties of protein – ligand interactions.

1.2

The importance of protein – ligand interaction in biological function

1.2.1 Enzyme – ligand interactions

Arguably the most extensively studied form of protein – ligand interactions are those involving enzymes. Enzymes are proteins which catalyze chemical reactions, and as such are critically important for chemical transformations required for life. Enzyme – ligand interactions take on many forms to accommodate the diverse functions enzymes perform. As catalysts all enzymes have a substate which is chemically transformed by the enzyme into a product, making enzyme – substrate and enzyme - product interactions the most numerous form of enzyme – ligand interaction. An interesting example of an enzyme - substrate interaction is the nitrogenase enzyme that uses N_2 as a substrate and produces NH_3, thereby serving as the only natural system capable of replenishing the earth's nitrogen cycle (6). Enzymes that hydrolyze the peptide bonds of other proteins are known as proteases. Proteases are involved in many important biological processes and are often critical to the life cycle of pathogens such as viruses, which makes them particularly interesting and important to study (7). Many enzymes utilize small molecules and ions as cofactors that aid in catalysis making this another form of enzyme – ligand interaction. Pyridoxal 5'-phosphate is an example of a very common cofactor used for transamination reactions important to metabolism (8). In many cases, the catalytic activity of enzymes can be modulated by enzymatic activators or inhibitors, making this an important form of enzyme – ligand interaction. Aspartate transcarbamoylase is a key enzyme in nucleotide biosynthesis and is allosterically activated by adenosine triphosphate and inhibited by cytosine triphosphate in order to regulate the balance of purine and pyrimidine nucleotide biosynthesis (9). Enzyme inhibitors are particularly valuable because of their use as therapeutics to treat various pathological conditions such as viral infections (7). Because of their extraordinary ability to catalyze chemical reactions, studying of enzyme – ligand interactions also enables the development of useful biotechnologies (10).

1.2.2 Receptor – ligand interactions

Receptor – ligand interactions are another intensively studied form of protein - ligand interaction, due to their importance in physiological and pathological processes and suitability as drug targets. A receptor is a protein which upon detecting a stimulus initiates a specific biological response, typically through a signaling cascade. The stimulus for most receptors involves the binding of a specific ligand, in a receptor – ligand interaction. The ligands recognized by receptors can take many different forms, such as ions, small organic molecules, or proteins. The calcium-sensing receptor is an example of a receptor whose ligand is an inorganic Ca^{2+} ion. Calcium is an element required for many biological processes and as such its concentrations in extracellular fluids are carefully regulated. Through binding with its Ca^{2+} ligand, the calcium-sensing receptor is able to detect changes in Ca^{2+} concentrations and initiates a signaling cascade that maintains Ca^{2+} homeostasis (11). Many commonly used pharmaceuticals are ligands of naturally occurring receptors. Morphine is a small organic molecule naturally produced by the plant species *Papaver somniferum* and is medically valuable for its use as an analgesic for treatment of pain. Morphine is an exogenous ligand of the μ opioid receptor, where binding of morphine to the receptor activates the natural pain-relieving response. The μ opioid receptor has several endogenously produced peptide ligands including β-endorphin, endomorphin-1 and endomorphin-2 which also trigger a pain-relieving response when bound the receptor (12). Many cell signaling ligands that regulate processes like cell growth, differentiation, survival, inflammation, tissue repair and the immune response are proteins. Fibroblast growth factors are a family of proteins that serve as ligands for fibroblast growth factor receptors (Figure 1.2). Binding of fibroblast growth factors to their receptors initiates cell signaling cascades which regulates processes like cell migration, proliferation, differentiation, and survival. Different members of the fibroblast growth factor family interact with various fibroblast growth factor receptors in different combinations to initiate different cell signaling responses (13). Because of the central role in regulating so many biological processes, understanding receptor – ligand interactions is important for the scientific understanding of physiological processes, as well as the development of therapeutics for treatment of disease or for biotechnology applications.

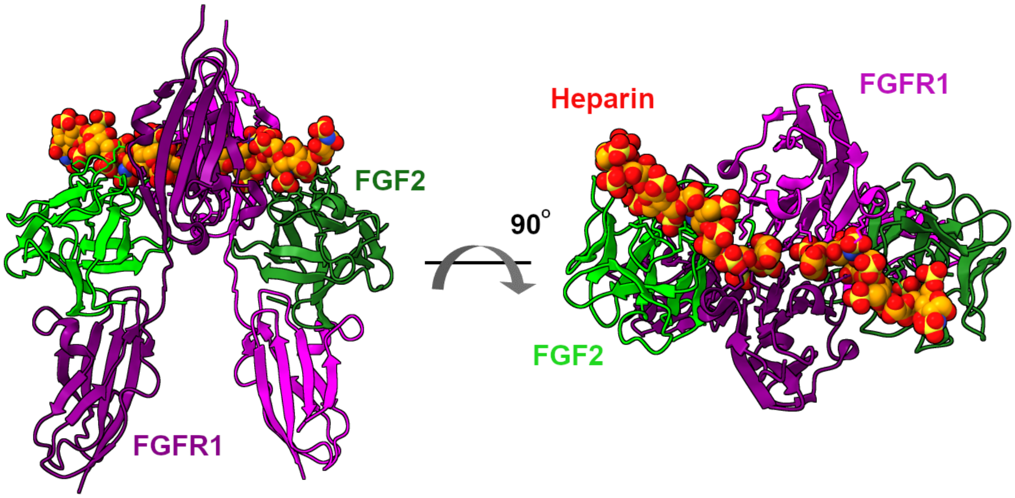


Figure 1.2. Crystal structure of the Fibroblast growth factor 2 (FGF2 shown in green and brown) in complex with fibroblast growth factor receptor 1 (FGFR1 shown in blue and purple) and heparin (shown in red) (PDB ID 1FQ9)(14). Complex includes a protien – protien interaction

between FGF2 and FGFR1 and a protien – glycan interaction between the FGFR1 - FGF2 complex and heparin, demonstrating the complexity of protein – ligand interactions.

1.2.3 Other protein – ligand interactions

Enzyme – ligand interactions and receptor – ligand interactions are two classes of protein – ligand interactions that are heavily studied. There are however many other examples of protein – ligand interactions that do not involve the catalytic action of enzymes or the stimuli sensing properties of a receptor yet serve an equally important role in biology. The interaction between hemoglobin and oxygen is a protein – ligand interaction of fundamental importance in all vertebrate organisms, enabling the efficient distribution of oxygen throughout the organism (15). Similar to enzymes which use co-factors to aid in catalysis, many proteins utilize co-factors to aid their function. An example of this are the integrins which are extracellular adhesion proteins which require Ca^{2+} and Mg^{2+} co-factors to carry out their adhesive function (16). Other protein – ligand interactions perform a storage function, like calsequestrin which binds and stores intracellular Ca^{2+} ions or ferritin which stores reactive Fe(II) ions (17, 18). A final form of protein – ligand interaction which deserves mention is the interaction between antibodies and antigens. Antibodies are proteins of a host’s immune system which recognize and bind specific features of pathogens known as antigens. Binding of these antigens allows for the neutralization and clearance of pathogens from the host (19). Due to their ability to bind specific antigens, antibodies also serve as important tools in biochemistry research, biotechnology applications and as therapeutics.

1.3

Thesis Structure

1.3.1 Thesis aims

This thesis presents the study of protein-ligand interactions using existing and newly developed biophysical methods. This biophysical approach aims to enhance the understanding of the physical properties and molecular mechanisms of these interactions in order to better understand the greater biological processes they give rise to. A secondary aim of this thesis is to utilize the understanding gained from this biophysical approach in order to guide the development of inhibitors targeting these interactions.

1.3.2 Thesis contents and chapter hypothesis

Chapter 1 provides an overview of protein biochemistry and highlights the critical role of protein – ligand interactions in life processes, to convey the fundamental importance of studying protein – ligand interactions. A summary of the following chapters is made, the aims and hypothesis of the research presented in each chapter are specified.

Chapter 2 provides an explanation of the theoretical and technical background of some of the core concepts and methods used in this thesis.

Chapter 3 presents work published in Biophysical Journal on the interaction between Ca^{2+} and the laminin N-terminal domain. This chapter utilizes multiple structural, computational, bioinformatic and biophysical approaches to characterize the interaction between Ca^{2+} and the laminin N-terminal domain. From this complete analysis the biophysical properties of the laminin N-terminal domain are characterized, and the molecular mechanism of the interaction is presented. Furthermore, this chapter demonstrates the use of a robust methodology for studying protein – ligand interactions that could be readily applied to other protein – ligand interactions.

Chapter 3 hypothesis: Characterization of the laminin N-terminal domain – Ca^{2+} interaction using structural, computational, bioinformatic, and biophysical techniques can expand the functional understanding of the laminin N-terminal domain - Ca^{2+} binding interaction.

Chapter 4 presents work published in the Journal of Biological Chemistry on the biophysical and enzymatic characterization of the SARS CoV-2 main protease and the development of an improved enzymatic substrate better suited for high throughput screening. The complete biophysical and enzymatic characterization serves as a valuable reference for protein quality and enzyme assay optimization. Several commonly used substrates and protease cleavage sites were tested, and their best use cases identified. Guided by this detailed analysis, a substrate with improved fluorescent properties was developed and shown to be better suited for high throughput screening. This chapter therefore presents an experimental approach used for the secondary aim of this thesis: the development of inhibitors targeting protein – ligand interactions.

Chapter 4 hypothesis: An improved SARS CoV-2 main protease high throughput screening assay can be developed through careful method optimization and use of a 5-carboxyfluorescin based substrate.

Chapter 5 presents work done on the interaction between Netrin-1 and its receptors UNC5B and heparin sulfate. This work presents a thorough biophysical characterization of the interaction between Netrin-1, UNC5B and heparin sulfate. Guided by the understanding gained from this biophysical characterization, peptide inhibitors of the Netrin-1 - UNC5B interaction are developed and shown to function by blocking the action of a heparin sulfate co-receptor. This chapter presents a substantial advancement to the biophysical and mechanistic understanding of the Netrin-1 - UNC5B interaction and demonstrates the first steps towards the development of peptide-based inhibitors targeting this interaction. In addition, this chapter presents the combined application of the methodology presented in chapter 3 for characterizing protein – ligand interactions and chapter 4 for developing inhibitors of these interactions.

Chapter 5 hypothesis: Biophysical characterization of the Netrin-1 - UNC5B interaction reveals the molecular mechanism of the interaction which can be targeted with peptide-based inhibitors.

Chapter 6 concludes this thesis with a discussion of the importance of understanding protein – ligand interactions. The larger implications and future directions of the work presented in this thesis are also discussed.

1.4 References

1. Nelson, D. L., and Cox, M. M. (2012) *Lehninger Principles of Biochemistry*, 6th Ed., W. H. Freeman and Company, New York, New York
2. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., and von Mering, C. (2022) The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646
3. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd Ed., W. H. Freeman and Company, New York, New York
4. Copeland, R. A. (2000) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd Ed., Wiley-VCF, Inc., New York, New York
5. Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.-L., and Liu, S.-Q. (2016) Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **17**, 144
6. Einsle, O., and Rees, D. C. (2020) Structural Enzymology of Nitrogenase Enzymes. *Chem. Rev.* **120**, 4969–5004
7. Leung, D., Abbenante, G., and Fairlie, D. P. (2000) Protease Inhibitors: Current Status and Future Prospects. *J. Med. Chem.* **43**, 305–341
8. Stetefeld, J., Jenny, M., and Burkhard, P. (2006) Intersubunit signaling in glutamate-1-semialdehyde-aminomutase. *Proc. Natl. Acad. Sci.* **103**, 13688–13693
9. Kantrowitz, E. R. (2012) Allostery and cooperativity in Escherichia coli Aspartate Transcarbamoylase. *Arch. Biochem. Biophys.* **519**, 81–90
10. Fasim, A., More, V. S., and More, S. S. (2021) Large-scale production of enzymes for biotechnology uses. *Curr. Opin. Biotechnol.* **69**, 68–76
11. Hofer, A. M., and Brown, E. M. (2003) Extracellular calcium sensing and signalling. *Nat. Rev. Mol. Cell Biol.* **4**, 530–538
12. Pathan, H., and Williams, J. (2012) Basic opioid pharmacology: an update. *Br. J. Pain*. **6**, 11–16
13. Xie, Y., Su, N., Yang, J., Tan, Q., Huang, S., Jin, M., Ni, Z., Zhang, B., Zhang, D., Luo, F., Chen, H., Sun, X., Feng, J. Q., Qi, H., and Chen, L. (2020) FGF/FGFR signaling in health and disease. *Signal Transduct. Target. Ther.* **5**, 1–38
14. Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., Linhardt, R. J., and Mohammadi, M. (2000) Crystal Structure of a Ternary FGF-FGFR-Heparin Complex Reveals a Dual Role for Heparin in FGFR Binding and Dimerization. *Mol. Cell*. **6**, 743–750
15. Ahmed, M. H., Ghatge, M. S., and Safo, M. K. (2020) *Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins*, 1st Ed., Springer Cham
16. Zhang, K., and Chen, J. (2012) The regulation of integrin function by divalent cations. *Cell Adhes. Migr.* **6**, 20–29
17. Wang, Q., and Michalak, M. (2020) Calsequestrin. Structure, function, and evolution. *Cell Calcium*. **90**, 102242
18. Arosio, P., Ingrassia, R., and Cavadini, P. (2009) Ferritins: A family of molecules for iron storage, antioxidation and more. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1790**, 589–599
19. Lu, L. L., Suscovich, T. J., Fortune, S. M., and Alter, G. (2018) Beyond binding: antibody effector functions in infectious diseases. *Nat. Rev. Immunol.* **18**, 46–61

Chapter 2

Biophysical methods for studying protein – ligand interactions

2.1 Protein – ligand binding equilibria and kinetics

Protein – ligand interactions are governed by the same thermodynamic principles which all physical processes occur by. As such, protein – ligand interactions can be understood by studying the physical properties of these interactions. The interaction between a protein (P) and a ligand (L) to form a protein – ligand complex (PL) can be described by the following chemical equilibrium, where k_a is the second order rate constant for complex formation and k_d is the first order rate constant for complex dissociation.

$$\mathrm{P} + \mathrm{L} \underset{k_\mathrm{a}}{\overset{k_\mathrm{d}}{\rightleftharpoons}} \mathrm{PL} \qquad \text{Eq.1}$$

The rate equations describing the rate of complex formation and complex dissociation are given by equations 2 and 3 respectively.

$$\frac{d[\mathrm{PL}]}{d\mathrm{t}} = k_\mathrm{a}[\mathrm{P}][\mathrm{L}] \qquad \text{Eq. 2}$$

$$- \frac{d[\mathrm{PL}]}{d\mathrm{t}} = k_\mathrm{d}[\mathrm{PL}] \qquad \text{Eq. 3}$$

At equilibrium the concentration of free protein, free ligand and protein – ligand complex are unchanging and therefore the rate of product formation (Eq. 2) is equal to the rate of complex dissociation (Eq. 3). The concentrations of protein, ligand and protein ligand complex at equilibrium determines the equilibrium constant for the reaction, given by equation 4. The convention in biochemistry is to describe this equilibrium in terms of the dissociation reaction, defined by the equilibrium dissociation constant (K_D) in units of M.

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_d}{k_a} \qquad \text{Eq. 4}$$

The dissociation constant of an interaction is a measure of binding affinity or strength of binding. A small dissociation constant indicates the protein – ligand interaction is strong, and the equilibrium favors the formation of the complex, while a large dissociation constant indicates the protein weakly interacts with the ligand and the reaction equilibrium favors the free protein and ligand. The dissociation constants for biological interactions can range from mM to fM (1). Understanding the affinity of a protein – ligand interaction is often important for understanding the concentrations of protein or ligand required to produce a biological response or to understand which reaction will prevail when multiple reaction equilibria are in competition.

Reaction rates are equally important for understanding biochemical processes. Studying reaction rates can often provide a great deal of information about the underlying mechanism of the chemical reaction being studied. Under biological conditions many protein – ligand interactions are not at equilibrium and therefore the protein – ligand interaction is under kinetic control. Under biological conditions the relative rates of complex formation and complex dissociation can be the factors which determine the biological effects of an interaction. For example it has been shown that drugs potency in many cases is determined by the dissociation rate of the protein – drug complex (2).

2.2 Measuring protein – ligand interaction kinetics and affinity with biolayer interferometry

Biolayer interferometry (BLI) is a technique that measures the real-time binding between a protein in solution and a ligand immobilized on a biosensor surface. With this technique binding affinity and kinetics can be measured for a wide range of protein – ligand interactions (3). The experimental set up for a BLI assay is simple and involves measuring the association and dissociation reactions of the protein to the surface immobilized ligand, at multiple protein concentrations typically ranging from 0.1 to 10 times the K_D of the interaction. An example of the resulting experimental data is shown in figure 2.1.

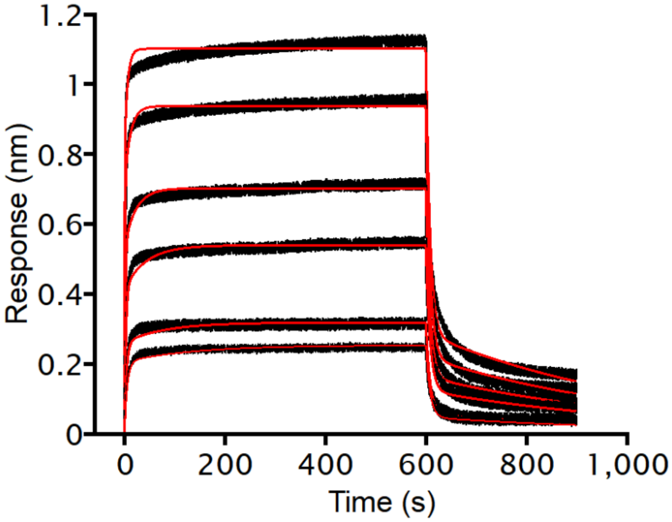


Figure 2.1. Example of data collected from a biolayer interferometry experiment. Black traces are the experimental binding response for various concentrations of protein, red traces are of the model that is fit to the experimental data to determine affinity and rate constants. The time from 0 to 600 seconds is of the association phase of the experiment, where protein in solution binds to the ligand immobilized on the biosensor, the response at 600 seconds is the equilibrium binding response. The time from 600 to 900 seconds is the dissociation phase of the experiment.

The ligand is immobilized on the biosensor surface using conventional immobilization chemistry and is immobilized such that the amount of ligand present is much less than the amount of protein present in solution(3). As a result, for an interaction such as that shown in equation 1, the association interaction between the surface immobilized ligand and protein in solution follows pseudo-first-order reaction kinetics. The integrated rate equation for this pseudo-first-order reaction can be expressed in terms of the experimentally measured binding response (R_t) and binding response at equilibrium (R_{eq}) as a function of time (t), and the known concentration of protein used in the experiment ([P]).

$$R_t = R_{eq}\,(1-e^{-(k_a[P] + k_d)t}) \qquad \text{Eq. 5}$$

Similarly, the integrated rate equation for the dissociation phase of the assay can be expressed in terms of the experimentally measured binding signal, where R_o is the binding response at the beginning of the dissociation step.

$$R_t = R_o\, e^{-k_d t} \qquad \text{Eq. 6}$$

Association (k_a) and dissociation (k_d) rate constants can be obtained by fitting the experimental data to these integrated rate equations and the equilibrium rate constant can be obtained according to equation 4. For more complex protein – ligand interactions that do not follow the simple binding model described by equation 1, differential rate equations describing the reaction mechanism can be generated, numerically integrated, and fit to experimental data to obtain rate and affinity constants (4).

Biolayer interferometry can also be used to perform an equilibrium binding experiment to measure the binding affinity of the interaction, without the use of kinetic information. This experiment is performed in the same way as the kinetic experiment however the reaction is left to progress until equilibrium binding signal is reached. The equilibrium binding signal is a relative measure of the concentration of protein – ligand complex for each protein concentration. To obtain the binding affinity (K_D) of the interaction, a plot of equilibrium binding response as a function of protein concentration can be fit to the following equilibrium binding model where R is the measured equilibrium binding response, R_{max} is the equilibrium binding response at infinite protein concentration, and [P] is the concentration of protein.

$$R = \frac{R_{max}[P]}{K_D + [P]} \qquad \text{Eq. 7}$$

Kinetics fitting measurements can often suffer from artifacts such as mass transport and analyte rebinding which can affect the accuracy of the resulting rate and affinity constants values obtained. Equilibrium binding measurements are less prone to these artifacts, thereby providing a complementary way of obtaining binding affinity from a BLI experiment.

2.3 Measuring affinity, stoichiometry, and thermodynamics of protein – ligand interactions with isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a technique that measures the heat absorbed or released when a protein and ligand interact. In a typical ITC experiment, a calorimeter cell is filled with a solution of protein, then ligand is titrated into the cell in small increments using a syringe. Under the constant pressure conditions of this experiment, the heat absorbed or released from the binding interaction between protein and ligand is the enthalpy change of the binding process (ΔH). Each injection of ligand into the cell will result in a decreasing heat signal as less free protein is available to bind the injected ligand, until eventually all protein binding sites are saturated with ligand (5). An example of the resulting data collected in an ITC experiment is shown in figure 2.2.

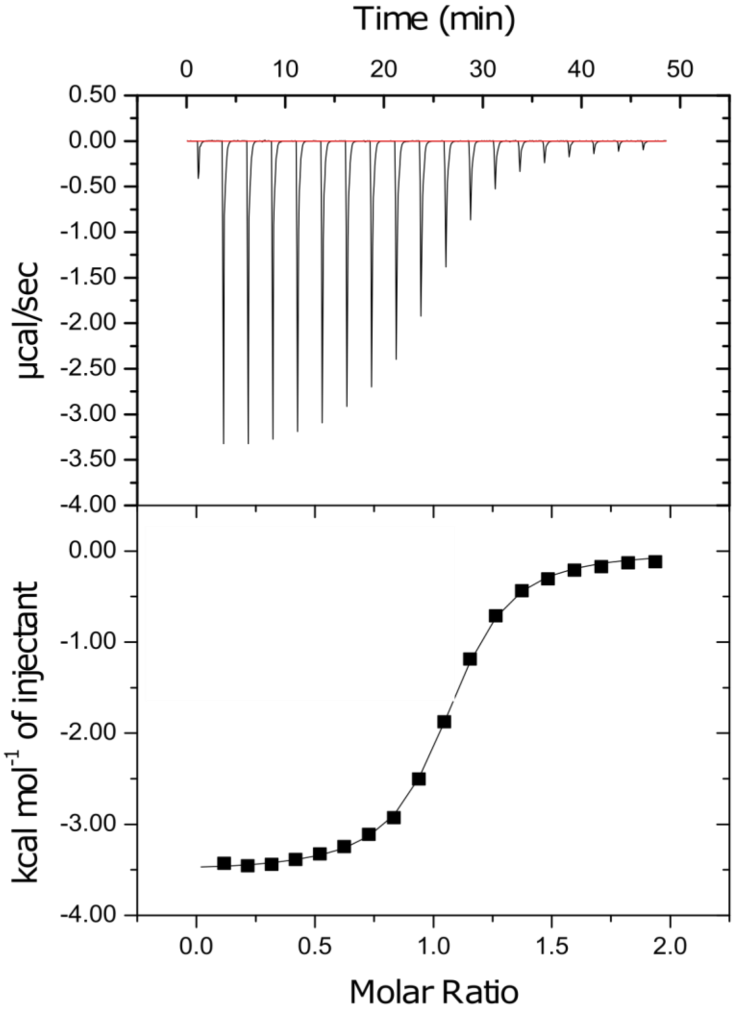


Figure 2.2. Example of data collected in an ITC experiment. Top panel shows heat signal produced by the titration of protein with ligand. Bottom panel shows the total heat signal per mole of injected ligand, for each injection of ligand in the titration series. This is plotted as a function of the molar ratio of ligand to protein.

By measuring the heat signal over the course of the titration experiment the binding affinity and stoichiometry can be determined. In this way ITC enables the measurement of binding affinity, stoichiometry, and enthalpy in one experiment. Typical ITC data analysis involves normalizing the total measured heat signal for each injection to the concentration of ligand injected. A plot of the heat absorbed or released per mole of ligand as a function of the mole ratio of ligand to protein can then be fit to the model described by Wiseman *et al.* (1989) to obtain values of stoichiometry, affinity, and enthalpy of binding. From these experimentally determined parameters, the free energy of binding (ΔG) and entropy of binding (ΔS) can be calculated using the following relation, where T is the temperature in degrees Kelvin and R is the ideal gas constant (6).

$$\Delta G = \mathrm{RT}\ \ln(K_{\mathrm{D}}) = \Delta H - \mathrm{T}\Delta S \qquad \text{Eq. 8}$$

The ability to obtain values of ΔH, ΔS, and stoichiometry is one of the primary benefits of ITC, as these parameters cannot easily be obtained by a conventional equilibrium binding experiment such as that described for BLI. ΔH and ΔS values can offer valuable insight into the nature of the protein ligand interaction. ΔH is the result of the change in energetically favorable interactions like hydrogen bonding, electrostatic and van der Waals interactions, whereas ΔS is the result of the change in disorder of the system due to things like hydrophobic effects and conformational changes (7). In addition to providing insight into the thermodynamics of the protein – ligand interaction, optimizing ΔH and ΔS values can aid in developing ligands with stronger protein binding affinity (8).

2.4 Ligand screening and measuring protein – ligand binding affinity with a thermal shift assay

A thermal shift assay (TSA) is a simple and widely applicable technique for the detection of protein - ligand interactions. The technique relies on measuring the thermal unfolding of a protein in the presence and absence of a ligand. Ligand binding to the protein typically increases the thermal stability of the protein which can be detected and measured as an increase in the midpoint (T_{M}) of the thermal unfolding curve of the protein (9). There are several ways to measure the thermal unfolding of a protein, some of the most common include measuring intrinsic protein fluorescence or the use of a SYPRO orange fluorescent dye. Thermal shift assays have several advantages including simplicity in assay design and data collection, requirement of small amounts of protein and ligand, and can easily be implemented for high throughput screening (10). A drawback of thermal shift assays is the general difficulty in quantifying binding constants from thermal shift assay data (11). Recently an isothermal analysis method for quantifying binding affinities from TSA data was developed (12).

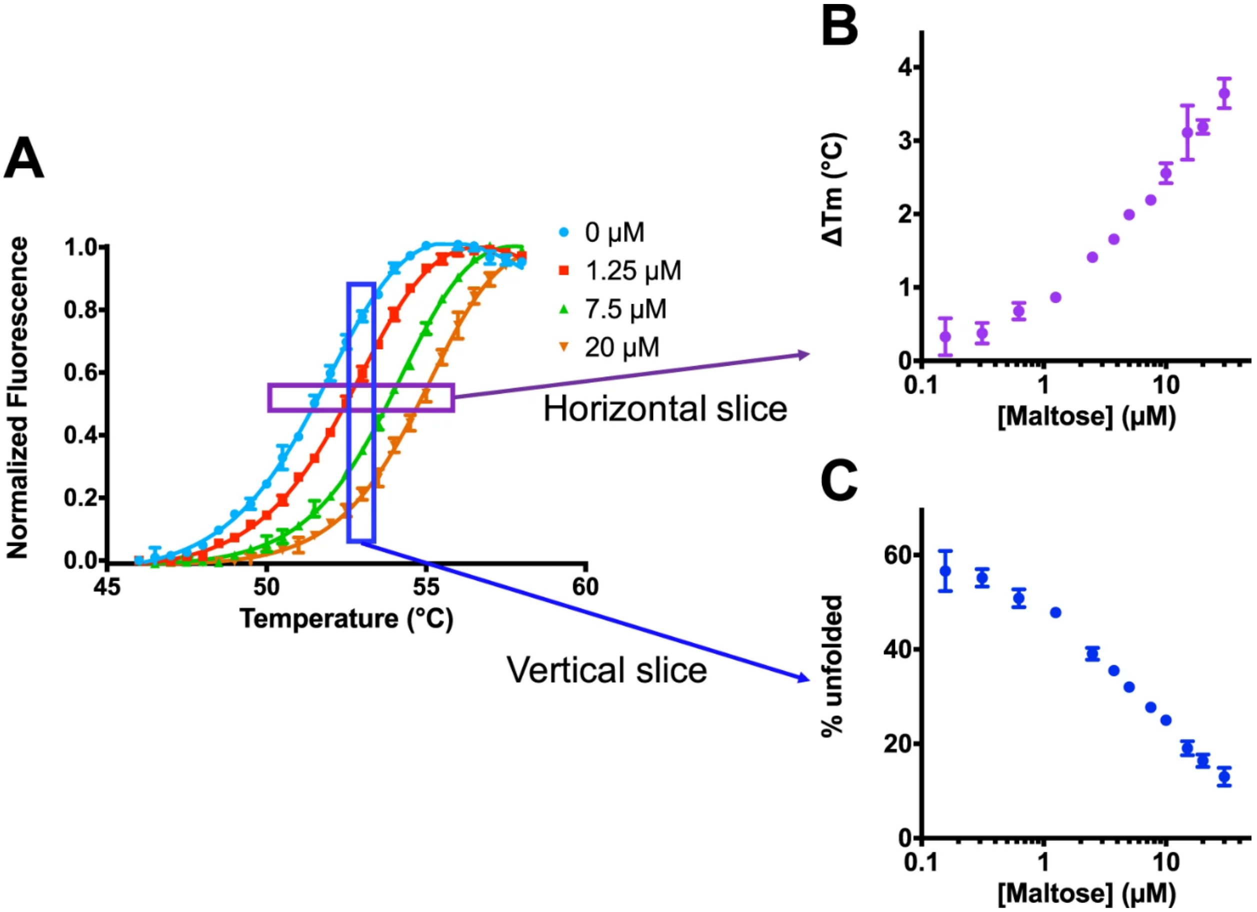


Figure 2.3. Example of TSA data used to determine protein – ligand binding affinity. Figure taken from Bai *et al.* (12) used under Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/deed.en). A) Protein melt curve obtained at various ligand concentrations, showing a ligand dependant increase in T_{M}. C) The fraction of unfolded protein at a particular temperature for each ligand concentration can be determined from the protein melt curves in panel A. The resulting data shown in panel C is fit to equation 9.

This method relies on measuring the melting curve of the protein in the presence of a dilution series of ligand as shown in figure 2.3. The fraction of unfolded protein at a particular temperature for each ligand concentration is then calculated and a plot of the fraction of unfolded protein (f_u) as a function of ligand concentration can be fit to equation 9 to determine the protein – ligand binding affinity (K_D). This method simultaneously fits the K_U which is the equilibrium constant for protein unfolding in the absence of ligand.

$$f_u = \frac{1}{1 + ((1/K_U) \times (1 + [L]/K_D))} \qquad \text{Eq. 9}$$

2.5 Methods to study enzyme – ligand interactions and the kinetics of enzyme catalyzed reactions

Enzyme – ligand interactions can often be studied by kinetic analysis of the chemical reactions they catalyze, in addition to the techniques used to study protein – ligand interaction described in the previous sections. Kinetic analysis of enzyme catalyzed reactions relies on measuring the reaction rate, which is achieved by monitoring the change in substrate or product concentration with respect to time. The method used to monitor this varies depending on the enzyme and the goals of the study. The enzyme catalyzed conversion of substrate to product can consist of a complex series of steps, however overall reaction can generally be described by the reaction scheme shown in equation 10. In this scheme E denotes the enzyme, S denotes the substrate and P denotes the product, and k_1, k_{-1}, and k_2 are the reaction rate constants (13).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad \text{Eq.10}$$

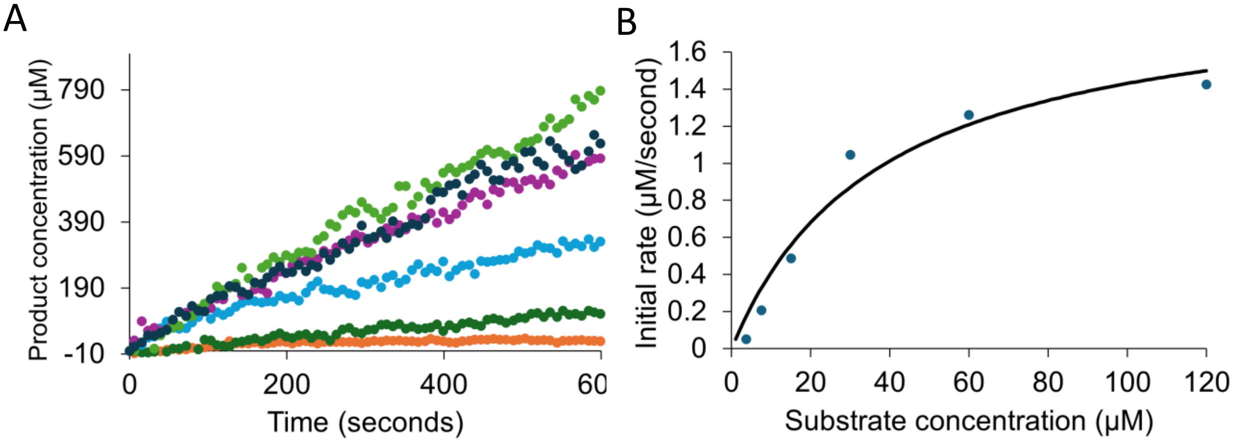


Figure 2.4. Kinetic analysis of initial rate of enzyme catalyzed reaction under steady state conditions. A) Initial rate of product formation for a range of substrate concentrations. B) Initial rate shown in panel A as a function of substrate concentration, fit to the Michaelis-Menten equation.

A common method to characterize the kinetics of an enzyme catalyzed reaction involves measuring the initial rate of the reaction for a range of substrate concentrations (Figure 2.4A). A plot of initial reaction rate as a function of substrate concentration shown in figure 2.4B can then be fit to the Michaelis-Menten equation (Eq. 11) to obtain values of V_{max} and K_{M}.

$$v = \frac{V_{\mathrm{max}}[\mathrm{S}]}{K_{\mathrm{M}} + [\mathrm{S}]} \qquad \text{Eq. 11}$$

V_{max} is the rate of reaction obtained at infinite substrate concentration, where all enzyme active sites are bound with substrate, and K_{M} is the substrate concentration at which half of the enzyme active sites are bound with substrate, resulting in a reaction rate half that obtained at V_{max}. If the concentration of enzyme known, then the value of k_{cat} can be obtained with the following equation.

$$V_{\mathrm{max}} = k_{\mathrm{cat}}[\mathrm{E}] \qquad \text{Eq.12}$$

k_{cat} is the first order rate constant describing the overall rate of the combined chemical steps involved in converting the enzyme – substrate complex into free enzyme and product. As such k_{cat} is useful for understanding the maximum rate at which the enzyme can convert substrate to product. However, given that under most conditions enzymes do not operate under biological conditions where substrate concentrations sufficiently high to fully saturate all available enzyme, the rate of conversion of substrate to product also depends on the K_{M}. When the concentration of substrate is much less than the K_{M} the rate of product formation can be described by the following equation.

$$v = k_{\mathrm{cat}}/K_{\mathrm{M}}[\mathrm{E}][\mathrm{S}] \qquad \text{Eq.12}$$

Where $k_{\mathrm{cat}}/K_{\mathrm{M}}$ is the apparent second order rate constant for product formation, and as such is a useful parameter for understanding enzyme catalyzed reactions.

2.6 References

1. Luong, J. H. T., and Vashist, S. K. (2020) Chemistry of Biotin–Streptavidin and the Growing Concern of an Emerging Biotin Interference in Clinical Immunoassays. *ACS Omega*. **5**, 10–18
2. Copeland, R. A., Pompliano, D. L., and Meek, T. D. (2006) Drug–target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **5**, 730–739
3. Concepcion, J., Witte, K., Wartchow, C., Choo, S., Yao, D., Persson, H., Wei, J., Li, P., Heidecker, B., Ma, W., Varma, R., Zhao, L.-S., Perillat, D., Carricato, G., Recknor, M., Du, K., Ho, H., Ellis, T., Gamez, J., Howes, M., Phi-Wilson, J., Lockard, S., Zuk, R., and Tan, H. (2009) Label-Free Detection of Biomolecular Interactions Using BioLayer Interferometry for Kinetic Characterization. *Comb. Chem. High Throughput Screen.* **12**, 791–800
4. Myszka, D. G., and Morton, T. A. (1998) CLAMP©: a biosensor kinetic data analysis program. *Trends Biochem. Sci.* **23**, 149–150
5. Fisher, H. F., and Singh, N. (1995) Calorimetric methods for interpreting protein—Ligand interactions. in *Methods in Enzymology*, Volume 259, Elsevier
6. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137
7. Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., Ji, X.-L., and Liu, S.-Q. (2016) Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **17**, 144
8. Gaucher, J.-F., Reille-Seroussi, M., and Broussy, S. (2022) Structural and ITC Characterization of Peptide-Protein Binding: Thermodynamic Consequences of Cyclization Constraints, a Case Study on Vascular Endothelial Growth Factor Ligands. *Chem. – Eur. J.* **28**, e202200465
9. Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., and Salemme, F. R. (2001) High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *SLAS Discov.* **6**, 429–440
10. Simeonov, A. (2013) Recent developments in the use of differential scanning fluorometry in protein and small molecule discovery and characterization. *Expert Opin. Drug Discov.* **8**, 1071–1082
11. Schellman, J. A. (1975) Macromolecular binding. *Biopolymers*. **14**, 999–1018
12. Bai, N., Roder, H., Dickson, A., and Karanicolas, J. (2019) Isothermal analysis of ThermoFluor data can readily provide quantitative binding affinities. *Sci. Rep.* **9**, 2650
13. Copeland, R. A. (2000) *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd Ed., Wiley-VCF, Inc., New York, New York

Chapter 3

Multidisciplinary approach to characterize protein – ligand interactions

Adapted from: Legare, S.[1], Heide F.[1], Gabir, H., Rafiei, F., Meier, M., Padilla-Meier, G.P., Koch, M., Stetefeld, J. (2024). Identifying the molecular basis of Laminin N-terminal domain Ca^{2+} binding using a hybrid approach. Biophysical Journal, https://doi.org/10.1016/j.bpj.2024.06.005.

[1] These authors contributed equally to this work

Contributions statement

All authors were involved in study design. J.S. and M.M. prepared Net-1 - Sm^{3+} crystals and collected diffraction data. F.H. processed the diffraction data, built the crystallographic model, and prepared figures. M.K. produced cell lines and established protein expression protocols. S.L., H.G., F.R. and G.P. produced proteins. S.L. removed Ca^{2+} from proteins and buffers. S.L. performed TSA, ITC and BLI experiments, analysed data and prepared figures and tables. F.H. performed MD and MMPBSA simulations, analysed data and prepared figures. F.H. performed the bioinformatic analysis of Net-1 sequences, analysed data and prepared figures. S.L. wrote the manuscript. All authors contributed to the editorial process of the final manuscript.

3.1

Unlocking the complexity: Exploring protein-ligand interactions through a multidisciplinary lens

The techniques and experimental approach used to study protein – ligand interaction depends on the research question being asked and the information the researcher is seeking to gain. A researcher seeking to understand the molecular basis of a protein – ligand interaction may utilize structural techniques like X-ray crystallography and computational techniques like molecular dynamics simulations. X-ray crystallography will allow a researcher to gain a visual understanding of the molecular basis of the interaction. Combining structural information with molecular dynamics simulations will enable the researcher to calculate dynamic and quantitative information about the protein – ligand interaction. This type of information is valuable for understanding the molecular mechanisms of biological processes and rationally designing drugs that target protein – ligand interactions.

Given the complex nature of protein – ligand interactions, structural and computational information cannot provide a complete understanding of protein – ligand interactions. Many proteins interact with a number of ligands with varying binding affinity, kinetics and specificity which combine to affect a biological outcome. Because properties such as binding affinity, binding kinetics and binding specificity cannot easily be gained from a structural or computational study, biophysical experiments are required for measuring these properties. Bioinformatic tools can be extremely useful for understanding protein – ligand interactions. With these bioinformatic tools a detailed understanding of protein – ligand interactions made from a particular system can be translated to similar systems. Often details like protein function, protein structure, and important residues involved in protein – ligand interactions can be identified with bioinformatic tools.

In this chapter we demonstrate the use of a multidisciplinary approach combining the use of structural, computational, biophysical and bioinformatic tools to gain a thorough and detailed understanding of molecular basis and functional properties of the interaction between the Laminin N-terminal domain and Ca^{2+}. The techniques and experimental approach used in this chapter could easily be applied to a variety of other protein – ligand interactions to gain a similar mechanistic and functional understanding of the interaction.

3.2

Identifying the molecular basis of Laminin N-terminal domain Ca^{2+} binding using a hybrid approach

3.2.1 Abstract

Ca^{2+} is a highly abundant ion involved in numerous biological processes, particularly in multicellular eukaryotic organisms where it exerts many of these functions through interactions with Ca^{2+} binding proteins. The Laminin N-terminal (LN) domain is found in members of the Laminin and Netrin protein families where it plays a critical role in the function of these proteins. The LN domain of Laminins and Netrins is a Ca^{2+} binding domain and in many cases requires Ca^{2+} to perform its biological function. Here we conduct a detailed examination of the molecular basis of the LN domain Ca^{2+} interaction combining structural, computational, bioinformatic and biophysical techniques. By combining computational and bioinformatic techniques with X-ray crystallography we explore the molecular basis of the LN domain Ca^{2+} interaction and identify a conserved sequence present in Ca^{2+} binding LN domains. These findings enable a sequence-based prediction of LN domain Ca^{2+} binding ability. We use thermal shift assays and isothermal titration calorimetry to explore the biophysical properties of the LN domain Ca^{2+} interaction. We show that the Netrin-1 LN domain exhibits a high affinity and specificity for Ca^{2+} which structurally stabilizes the LN domain. This study elucidates the molecular foundation of the LN domain Ca^{2+} binding interaction and provides a detailed functional characterization of this essential interaction, advancing our understanding of protein-Ca^{2+} dynamics within the context of the LN domain.

3.2.2 Introduction

Ca^{2+} is a critical element in biology where it serves a variety of structural, functional, and regulatory roles. Intracellular Ca^{2+} is a highly versatile second messenger playing roles in various processes from muscle contraction to signal transduction (1, 2). In the extracellular space Ca^{2+} is also know to play a critical role in diverse cellular processes like cell signaling, adhesion and blood coagulation (3–8). In all cases, Ca^{2+} exerts these effects through interactions with Ca^{2+} binding proteins. The extracellular matrix contains many Ca^{2+} binding proteins, some of which require Ca^{2+} for fixed roles such as structural stabilization while others require Ca^{2+} for dynamic roles such as cofactors in catalysis and receptor binding (9–13).

Laminin and related Netrin protein families are extracellular Ca^{2+} binding glycoproteins which serve structural, functional, and regulatory roles (14–16). Laminins are a major component of cell basement membranes, where they polymerize into a structural network of heterotrimers composed of the N-terminal short arms of α, β and γ Laminin subunits. Laminin polymerization occurs through Ca^{2+} dependent interactions between the Laminin N-terminal (LN) domains of the α, β and γ subunits (17). The Laminin subunit-γ (Lam-γ) LN domain is responsible for this Ca^{2+} dependent polymerization and is the only Laminin subunit LN domain known to bind Ca^{2+} (18). Members of the Netrin family arose from the Laminin family by gene duplication, domain shuffling and domain loss, and like the Laminins, they contain the LN domain (19, 20). The Netrin family is primarily involved in development of the nervous system through axon guidance and neurite outgrowth (21–26). Like Lam-γ, the LN domains of Netrin-1, Netrin-4, Netrin-G1 and Netrin-G2 are involved receptor binding and bind Ca^{2+} in a similar manner to the Lam-γ LN domain (27–30).

While the role of LN domain Ca^{2+} binding has been studied from a structural and biochemical perspective for some Laminins, the role of Ca^{2+} with respect to the Netrin LN domain has been under-researched, despite the availability of structural information for members of the Netrin family. In this work we use Netrin-1 (Net-1) as a model protein due to the broad availability of structural and functional information. By combining molecular dynamics simulations and bioinformatics analyses with X-ray crystallography we identify the molecular basis of the LN domain Ca^{2+} interaction, including the requirement of an uncommon bidentate coordination of Ca^{2+} by a conserved threonine. The structural analysis is combined with a biophysical characterization of the Net-1 LN domain Ca^{2+} binding interaction. Using isothermal titration calorimetry (ITC) and a differential scanning fluorimetry (DSF) based thermal shift assay (TSA) we explore the affinity, stoichiometry, and specificity of the Net-1 – Ca^{2+} interaction which demonstrates the role of Ca^{2+} in the stabilization of Net-1 and Lam-γ. Altogether, this study expands the structural and functional understanding of the role of Ca^{2+} in an important extracellular protein domain.

3.2.3 Methods

Protein expression and purification

Mus musculus Netrin-1 consisting of domains LN to LE3 (NCBI Reference Sequence: NP_032770, amino acids 24–457) used for crystallography was cloned, expressed and purified as previously published (31). *Gallus gallus* Netrin-1 domains LN to LE3 (NCBI Reference Sequence: NP_990750, amino acids 26 – 458), and laminin subunit-γ1 (NCBI Reference Sequence: XP_040561204.1, amino acids 27 - 391) used for biophysical studies was cloned, expressed and purified using previously described methods (32, 33).

Net-1 - Sm^{3+} complex crystallization

Net-1 domains LN to LE3 from *Mus musculus* was crystallized in a hanging drop vapor diffusion experiment at 293 K in 0.1 M HEPES (pH 7.7), 2.8 M NaCl, 0.2 M glycine at a concentration of 10.0 mg/mL. Crystals appeared after 1 month and grew to a final size of 0.1-0.5 mm after 3 months. Crystals were soaked with samarium(III) acetate hydrate (Hampton Research) and diffracted on the X06SA beamline at the Swiss Light Source with a wavelength of 1.0349 Å in 1.0 ° wedges at 100 K. The diffraction images were processed with XDS and the CCP4-package to final $P3_221$ space group (34, 35). Phases were calculated in Phaser using a Net-1 search model (PDB ID: 4OVE) (31, 36). The structure was built and refined crystallographically using the Coot suite and Phenix software packages (37, 38). The crystallographic data table (Table S3.1) is available in the supplemental information document. The Netrin-1 - Sm^{3+} complex crystallographic data set and model (PDB ID: 8SNP) is available on the RCSB Protein Data Bank (https://www.rcsb.org/).

Molecular dynamics simulations of apo and holo - Net-1

Molecular dynamics (MD) simulations for Net-1 with and without a Ca^{2+} ion were performed using the GROMACS molecular dynamics simulation package (39) with the CHARMM36 force field parameters (40) and a SPC/E water model (41). For both trajectories, Na^+ and Cl^- ions were added to counter the protein charges to maintain a neutral charge in the overall system. Both systems were subjected to a 50000 steps of steepest decent energy minimization, heated from 0 to 310 K for 100 ps in a canonical ensemble (NVT ensemble) and then equilibrated in an isothermal and isobaric ensemble (NPT ensemble) under a constant pressure of 1.0 Bar for 100 ps. The energetically minimized and stable systems were then placed on 100 ns trajectory with time integration steps of 2 fs. The MD trajectory used a Particle Mesh Ewald method with a short-range columbic interaction force cut-off at 12 Å and was temperature equilibrated at 310 K. The completed trajectory was post processed which included centering of the protein throughout the simulation and fixing rotational and translational trajectories. Protein stability as part of the overall system was confirmed by a backbone root-mean-square deviation (RMSD) and radius of gyration analysis. Further analysis included a molecular mechanics Poisson-Boltzmann surface area (MMPBSA) interaction analysis of the Net-1 - Ca^{2+} complex. The MMPBSA method used the GROMACS and APBS packages to calculate internal energies and decomposed interaction energies between the protein and calcium ligand with the gmx_MMPBSA toolkit(42). Interacting residues were cut off by a physical distance of 6 Å from the calcium ligand within the entire trajectory after which interaction energies were calculated across the 100 ns trajectory at 100 ps time steps, averaged and reported with its standard deviation. Finally, structural examinations and the generation of a movie covering 10 ns of the trajectory were performed using PyMOL 2.4.1 (Schrödinger, Inc.). These included removing solvent molecules and averaging residue trajectories to minimize high-frequency vibrations. MD simulation run parameters are listed in the supplemental information document (Table S3.2) and additional supporting data is available on the Open Science Framework (10.17605/OSF.IO/C9VQX).

Net-1 - Ca^{2+} binding region bioinformatics

The sequence of Netrin-1 from *Mus musculus* lacking the C-terminal domain (NP_032770, amino acids 24–457) was compared to other species' sequences through a database similarity search using BLAST (43, 44). This generated a multiple sequence alignment file which was further supplemented through the HHblits method and filtered in GREMLIN for overall residue coverage and gaps (45). The filtered multiple sequence alignment file containing 831 Netrin-1 related sequences from various species with a minimum sequence coverage of 75% on the target sequence. Further, individual sequences of the Ca^{2+} binding loops across Laminin family proteins were compared based on gene identity (NTN1, NTN3, NTN4, NTNG1, NTNG2, LAMA5, LAMB1, LAMC1). For each gene, 150 sequences from varying species were aligned using Clustal Omega (46). Lastly, the multiple sequence alignments were analyzed in WebLogo3 for residue position probabilities (47).

Preparation of apo protein and Ca^{2+} free buffer

Chelex-100 molecular biology grade resin, 200 – 400 mesh (Bio-Rad Laboratories Inc.), packed into a 1 cm diameter by 11 cm length cylindrical chromatography column was used to remove Ca^{2+} from proteins and buffers. The column was equilibrated in a two-step process, starting with 0.5 M HEPES pH 7.3 until the eluent reached a pH of 7.3, next the column was equilibrated with 20 mM HEPES pH 7.3, 150 mM NaCl until the eluent reached a stable conductivity. 1 mL of protein at a concentration of 1 mg/mL was injected on the column and the flow rate was set to 0.1 mL/minute. Fractions containing protein were pooled and concentrated, fractions not containing protein were used as buffer in ITC and TSA experiments.

Differential scanning fluorimetry based thermal shift assays

All DSF based TSA experiments measuring intrinsic protein fluorescence were performed on a Prometheus Panta (NanoTemper Technologies GmbH) using standard Prometheus capillaries (PR-C002), in Ca^{2+} free 20 mM HEPES pH 7.3, 150 mM NaCl buffer. Protein melts were performed from 25^{o}C until complete unfolding of the protein, using a temperature gradient of 1 ^{o}C/minute. Protein melting was measured by monitoring the change in 350/330 nm fluorescence ratio. Data analysis and T_m fitting was done with the Panta analysis software, version 1.2. To assess the complete removal of Ca^{2+} from proteins and to measure the T_m of the apo-protein, the protein melt before and after Ca^{2+} removal was recorded. To verify the complete removal of Ca^{2+} from the chelex-100 treated protein, between 1.6 mM to 50 mM EDTA was added to the treated protein and the resulting T_m was recorded. To assess the T_m of the holo-protein under physiological Ca^{2+} concentrations, 1.4 mM $CaCl_2$ was added to the apo-protein. To detect binding of divalent cations, the T_m of 10 µM apo - Net-1 incubated with 1 mM and 10 mM of the chloride salt (except for Pb^{2+} for which the acetate salt was used) of each cation was measured. Cations that showed a measurable dose dependant increase in T_m were tested further for binding affinity. To measure cation binding affinity by TSA, the T_m of 29 µM apo Net-1 or apo Lam-γ was incubated with a 1/3 dilution series of the chloride salt of the divalent cation ranging from 5.3 mM to 10 nM. Data fitting and analysis was performed using FoldAffinity available from the eSPC online data analysis platform (48–50). All binding affinity experiments were performed in triplicate, K_D values closest to the T_m of the apo protein and with the smallest confidence interval are reported.

Isothermal titration calorimetry

ITC was performed on a MicroCal iTC 200 titration calorimeter (Malvern Panalytical Ltd.) in Ca^{2+} free 20 mM HEPES pH 7.3, 150 mM NaCl buffer. 29.9 µM apo – Net-1 was loaded into the measurement cell and the reference cell was filled with water. The experiment was performed at 25 ^{o}C with a mixing speed of 600 rpm, a reference power of 6 µcal/sec in the high gain mode. The sample was titrated with 250 µM $CaCl_2$, with the first injection of 0.4 µL over 0.8 seconds, followed by 19 injections of 2 µL over 4 seconds and 200 seconds between injections. A reference experiment was performed under the same conditions, except the measurement cell was filled with buffer rather than apo – Net-1. Data analysis, reference subtraction and fitting to a one set of sites binding model was performed in the MicroCal supplied Origin 7.0 software package (OriginLab Corp.). Fitting values are reported as the mean ± 1 standard deviation of 3 replicate experiments.

3.2.4 Results

Structural examination of Net-1 and its Ca^{2+} binding site

In our investigation to understand the structural features of Net-1 and the binding of Ca^{2+} at the LN domain Ca^{2+} binding loop, we collected structural data through X-ray diffraction experiments. For this, a construct of Net-1 from *Mus musculus* consisting of the LN and the three LE domains was designed and crystallized. This construct is able to bind the Net-1 receptors, DCC, NEO, and UNC5B while being more amenable to crystallization (27, 31, 51). The crystals were then soaked with Samarium(III) acetate to produce the Net-1 - Sm^{3+} complex. Sm^{3+} is an electron-dense heavy atom used to bind at and displace Ca^{2+} from Ca^{2+} binding sites without altering the general protein structure (52). Moreover, the high electron density of Sm^{3+} allows for the detection of potentially low-occupied Ca^{2+} binding sites, making it an ideal choice for our structural examinations. The resulting structure was determined to have a resolution of 3.4 Å (Figure 3.1). As previously observed, the resulting structure contains N-glycans on the residues N95, N131, and N417, but not on N116 of the Ca^{2+} biding loop (31, 53). Additionally, a Net-1 dimer can be observed as a symmetry mate where various residue side chains stabilize the interaction. The general structure of this construct is characteristic of other Net-1 structures across various species with slight variations in multiple loop regions on the LN domain (27, 31, 54).

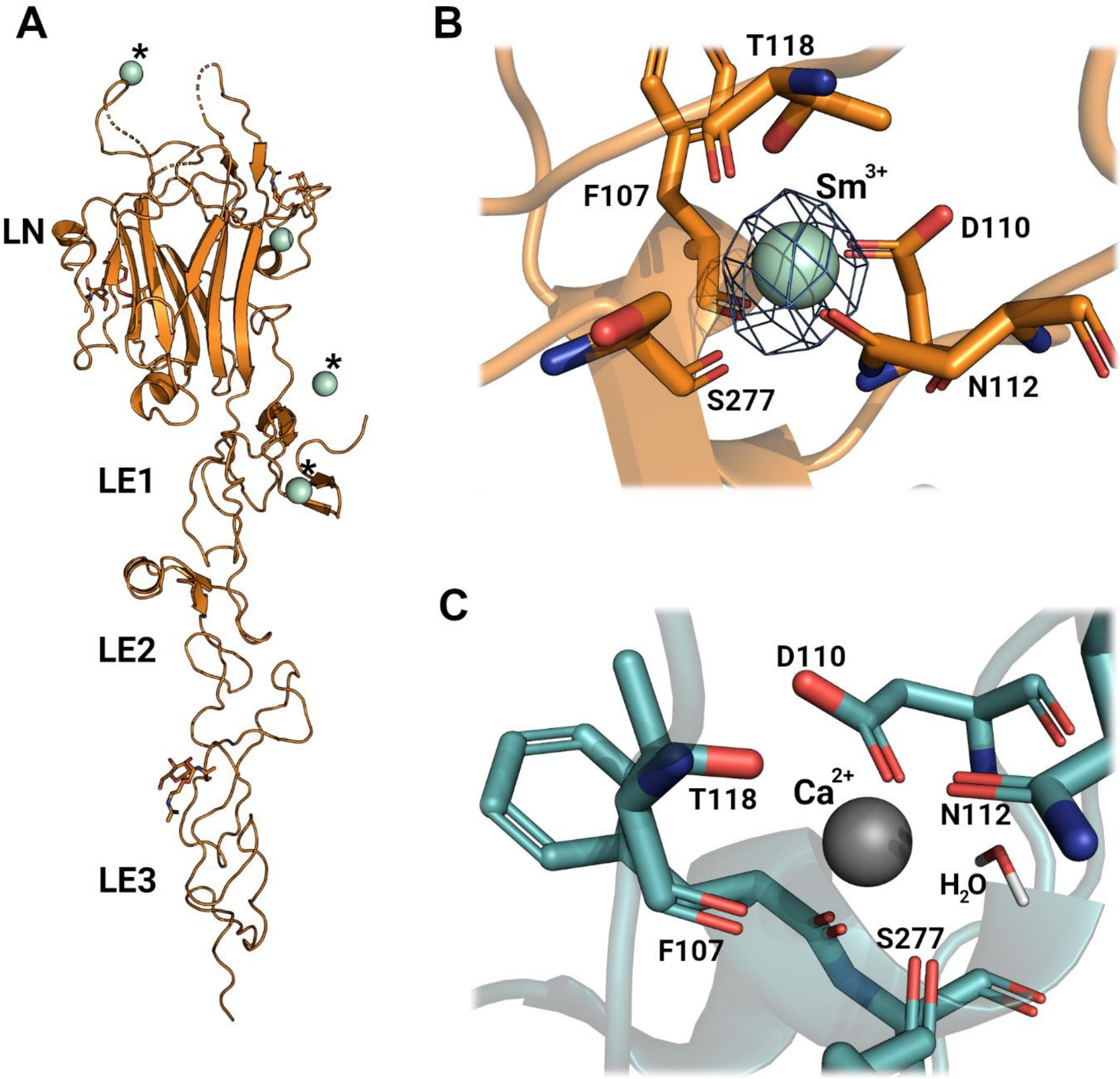


Figure 3.1. Structural examination of Net-1 in complex with metal ions. (A) X-ray diffraction structure model of Net-1 in complex with Sm^{3+} ions. Three Sm^{3+} ions are coordinated through crystallographic contacts (indicated with asterisk*) while one Sm^{3+} ion is bound to the Ca^{2+} binding loop of Net-1 (PDB ID: 8SNP). (B) Amino acid metal coordination of Sm^{3+} in the Ca^{2+} binding loop of Net-1. The ion is coordinated by side chain and backbone functional groups of the residues F107, D110, N112, T118 and S277. The electron density level of the 2Fo-Fc map

is set to 0.522 e/Å^3 (8 sigma contour level). (C) Ca^{2+} coordination in the Ca^{2+} binding loop of Net-1 (PDB ID: 4OVE) (31) with the indirect coordination of Ca^{2+} by D278 through a hydrogen bonded water molecule.

The crystal structure reveals four Sm^{3+} binding sites on Net-1 (Figure 3.1A). Three of the four Sm^{3+} ions have only a few coordinating ligands from Net-1 and rely on coordinating residues of Net-1 symmetry mates. For these reasons we conclude that these three Sm^{3+} ions are crystallographic artifacts rather than true Ca^{2+} binding sites.

The fourth Sm^{3+} ion is bound to the Net-1 LN domain at the equivalent Ca^{2+} binding location of other Netrin and Laminin structures (Figure 3.1C) (28–31, 55). This LN domain Ca^{2+} binding site consists of a short helix-loop segment packed against the edge of β-strand 8 of the LN domain β-sandwich. Most of the Sm^{3+} coordinating ligands come from the helix-loop segment, while a backbone carbonyl ligand from the β strand links the helix-loop segment to the β-sandwich core through the Sm3+ ion. The Sm^{3+} is directly coordinated by the side chains of D110, N112, T118 and backbone carbonyl oxygens from T118, F207 and S277. The side chain of D278 is oriented in a position to coordinate the Sm^{3+} through a hydrogen bonded water molecule, however this water is not observable in the 2Fo-Fc map due to the large overlapping signal from the Sm^{3+} ion and therefore was not added to the model (Figure 3.1B). The coordination geometry of Sm^{3+} is identical to the Ca^{2+} coordination geometry of the previously published mouse Net-1 crystal structure (PDB ID: 4OVE) (31) which includes a water molecule coordinating Ca^{2+} through D278 (Figure 3.1C).

Biophysical examination of the Net-1 - Ca^{2+} interaction

To evaluate the biophysical properties of the Net-1 - Ca^{2+} binding interaction we began by characterizing the thermal stability of apo and holo - Net-1 using a differential scanning fluorimetry (DSF) based thermal shift assay (TSA). The principle behind this assay is that holo - Net-1 will exhibit a higher melting point (T_M) compared to apo - Net-1, due to the stabilizing effects of Ca^{2+} binding. Measuring the Net-1 T_M in the presence and absence of Ca^{2+} can therefore be used to detect Ca^{2+} binding (48, 49). Ca^{2+} free buffer and apo - Net-1 was prepared using Chelex-100 cation exchange resin. The T_M of Net-1 after treatment with the cation exchange resin is 5.4 ^oC lower than the untreated Net-1, and the addition of EDTA up to 50 mM does not cause a measurable decrease in T_M of the treated Net-1 (Figure 3.2A). This demonstrates that the cation exchange treatment effectively removes Ca^{2+} from the buffer and protein to produce apo - Net-1. The same protocol was used to prepare apo – Lam-γ (Figure 3.2B). Next, we used this TSA to measure the thermal stability of holo - Net-1 at a physiological extracellular Ca^{2+} concentration of 1.4 mM (56). Under these conditions holo - Net-1 is considerably more thermally stable, with a T_M of 50.6 ^oC, compared to the 38.8 ^oC T_M of apo - Net-1 (Figure 3.2A). The same analysis for the Lam-γ reveals a T_M of 52.8 ^oC in the presence of 1.4 mM Ca^{2+}, compared to a T_M of 42.6 ^oC for apo – Lam-γ (Figure 3.2B). These results clearly demonstrate the important role of Ca^{2+} in stabilizing Net-1 and Lam-γ structure at physiological extracellular Ca^{2+} concentrations.

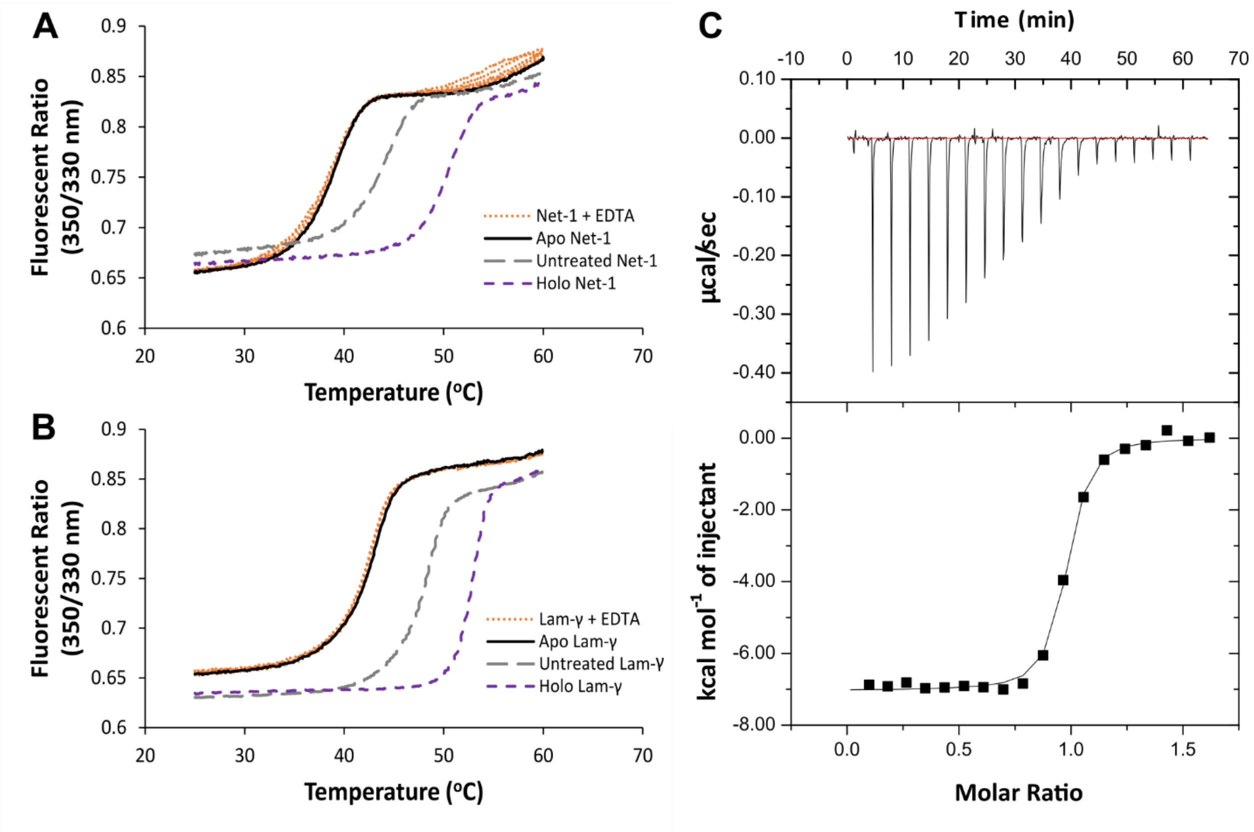


Figure 3.2. Characterization of the biophysical properties of the LN domain Ca^{2+} binding site. Thermal shift assay characterization of Net-1 (A) and Lam-γ (B) in 20 mM HEPES pH 7.3, 150 mM NaCl. Melt curve of holo – protein at a physiological extracellular Ca^{2+} concentration of 1.4 mM shown in purple short dashed line (Net-1 T_M of 50.6 °C, Lam-γ T_M of 52.8 °C). Melt curve of Chelex-100 treated apo – protein shown in solid black line and in the presence of 1.6 to 50 mM EDTA shown in orange dotted line (Net-1 T_M of 38.8 °C, Lam-γ T_M of 42.6 °C), demonstrating Chelex-100 treatment results in the complete removal of Ca^{2+} to produce apo - protein. Melt curve before treatment with Chelex-100 resin (Untreated) shown in grey long dash line (Net-1 T_M of 44.2 °C, Lam-γ T_M of 48.2 °C). Isothermal calorimetric titration of apo - Net-1 with Ca^{2+}(C). 29.9 μM apo - Net-1 in 20 mM HEPES pH 7.3, 150 mM NaCl was titrated with 250 μM $CaCl_2$ at 25 °C. The binding constant (K_a) is measured to be 1.46 ± 0.20 x 10^7 M^{-1} with a stoichiometry of 0.942 ± 0.022. The enthalpy change (ΔH) is -7040 ± 200 cal mol^{-1} and the

entropy change (ΔS) is 9.16 ± 0.96 cal mol^{-1} K^{-1}. ITC values are reported as the mean ± 1 standard deviation of three experimental replicates.

We then used the TSA to estimate the affinity of Net-1 for Ca^{2+} following a previously described method (48, 49). The resulting K_D of the Net-1 - Ca^{2+} interaction was determined to be 1.36 μM at 42 °C, assuming a 1 to 1 binding model (Table 3.1, Figure S3.1). Using this assay we also tested the affinity of the previously characterized Lam-γ - Ca^{2+} interaction and found the K_D of the interaction to be 22.4 μM at 47 °C, which is in appropriate agreement with the previously reported value (57) (Table 3.1, Figure S3.1). Finally, we used the TSA to screen a variety of divalent and trivalent cations for their ability to bind Net-1 (Table 3.2). Of the cations screened besides Ca^{2+}, only Sr^{2+}, Ba^{2+}, and Mg^{2+} exhibited measurable binding affinities (Table 3.1). Toxic metals like Pb^{2+} and Cd^{2+} exert their toxic effects, in part, by binding certain Ca^{2+} binding proteins (58, 59), however their binding to Net-1 was not detectable. Finally, Net-1 in the presence of Sm^{3+} did not produce a sigmoidal transition in the fluorescent signal that could be attributed to protein melting, indicating Sm^{3+} denatures Net-1 at the concentrations used in the assay (Table 3.2). These results show that Net-1 exhibits a high degree of specificity for Ca^{2+} over other divalent cations, including biologically relevant cations like Mg^{2+}.

Table 3.1. Divalent cation binding affinities for Net-1 and Lam-γ measured by a DSF based TSA.

Interaction	K_D of metal binding at given temperature	Asymmetric 95% CI [lower, upper]	T_M of apo protein
Net-1 – Ca^{2+}	1.36 µM at 42 °C	[0.925 µM, 1.94 µM]	38.8 °C
Net-1 – Sr^{2+}	25.9 µM at 42 °C	[12.7 µM, 53.3 µM]	38.8 °C
Net-1 – Mg^{2+}	981 µM at 42 °C	[151 µM, 491 mM]	38.8 °C
Net-1 – Ba^{2+}	1.03 mM at 42 °C	[119 µM, 513 mM]	38.8 °C
Lam-γ – Ca^{2+}	22.4 µM at 47 °C	[20.4 µM, 24.6 µM]	42.6 °C

Table 3.2. Thermal shift assay testing binding of various divalent and trivalent cations to apo Net-1.

Cation	ΔT_M (°C) at 1mM [a]	ΔT_M (°C) at 10 mM [a]
Ca^{2+}	10.4	12.6
Mg^{2+}	1.3	2.6
Sr^{2+}	4.8	9.8
Ba^{2+}	1.7	5.8
Mn^{2+}	-0.2	-0.7
Co^{2+}	1.5	-1.8
Ni^{2+}	-0.7	-3.5
Cu^{2+}	no melt [b]	no melt [b]
Zn^{2+}	no melt [b]	no melt [b]
Pb^{2+}	0.1	-12.9
Cd^{2+}	-0.6	-4.8
Hg^{2+}	no melt [b]	no melt [b]
Sm^{3+}	no melt [b]	no melt [b]

[a] T_M shift relative to apo Net-1

[b] monitoring the 350/330 nm fluorescence ratio did not produce a melting transition, indicating possible protein denaturation upon addition of cation to apo Net-1 solution.

To characterize the thermodynamics of the Net-1 - Ca^{2+} binding interaction we used ITC to measure the enthalpy change (ΔH), entropy change (ΔS), binding constant (K_a) and stoichiometry of the interaction (Figure 3.2C). ITC reveals that Net-1 strongly binds Ca^{2+} with a binding constant of $1.46 \pm 0.20 \times 10^7$ M^{-1} and a stoichiometry of 1. The high affinity of the interaction, coupled with the stoichiometry of 1, indicates that the measured binding interaction is attributed to the canonical Ca^{2+} binding loop of Net-1. These ITC results also show that the high affinity Net-1 - Ca^{2+} interaction is driven by favourable enthalpic and entropic contributions, assuming buffer effects and proton coupling interactions are negligible or absent (Figure 3.2C). The discrepancy in affinity of the Net-1 - Ca^{2+} interaction measured by ITC and TSA is expected because the concentration of protein used in the TSA is much higher than the K_D of the interaction. TSA experiments are known to produce inaccurate results when the K_D of the interaction is much lower than the concentration of protein used in the TSA (48, 49). We therefore take the affinity measured by ITC to be closest to the true value for the Net-1 - Ca^{2+} interaction. Net-1 titrated with excess Ca^{2+} did not produce a measurable heat signal, further indicating that the three other Sm^{3+} binding sites in the Net-1 - Sm^{3+} crystal structure are crystallographic artifacts rather than low affinity Ca^{2+} binding sites.

Molecular features of the Net-1 - Ca^{2+} interaction

In order to elaborate on the biophysical results and gain insights into the molecular basis and dynamic aspects of the Net-1 - Ca^{2+} interaction, MD simulations of holo-Net-1 were performed on the energy minimized Net-1 crystal structure. For this, all Sm^{3+} ions were removed and a Ca^{2+} was placed at the location of Sm^{3+} in the LN domain Ca^{2+} binding loop. Over the course of a 100 ns MD simulation, the RMSD and R_g of the Net-1 - Ca^{2+} complex is stable and retains an overall similar structure to the crystal structure (Figure S3.2). Throughout the MD simulation, Net-1 coordinates Ca^{2+} in a pentagonal bipyramidal geometry with seven coordinating ligands including a water molecule which forms hydrogen bonds with D278 and the backbone carbonyl of D110 (Figure 3.3A). The decomposed interaction energies for residues near the LN domain Ca^{2+} binding loop were determined using the molecular mechanics Poisson-Boltzmann surface area (MMPBSA) approach (Figure 3.3B). D110, which directly coordinates Ca^{2+} via its side chain and indirectly through the water hydrogen bonding with its backbone carbonyl, is found to make the greatest energetic contribution to Ca^{2+} binding. In addition, T118 forms a bidentate coordination with Ca^{2+} via its side chain and backbone carbonyl to make the second largest energy contribution. The backbone carbonyl of F107, S277 and the side chain of N112 also make considerable energetic contributions to Ca^{2+} binding. Finally, the side chain of D278 and the carbonyl backbone of L108 make a measurable energetic contribution by hydrogen bonding with the coordinating water ligand.

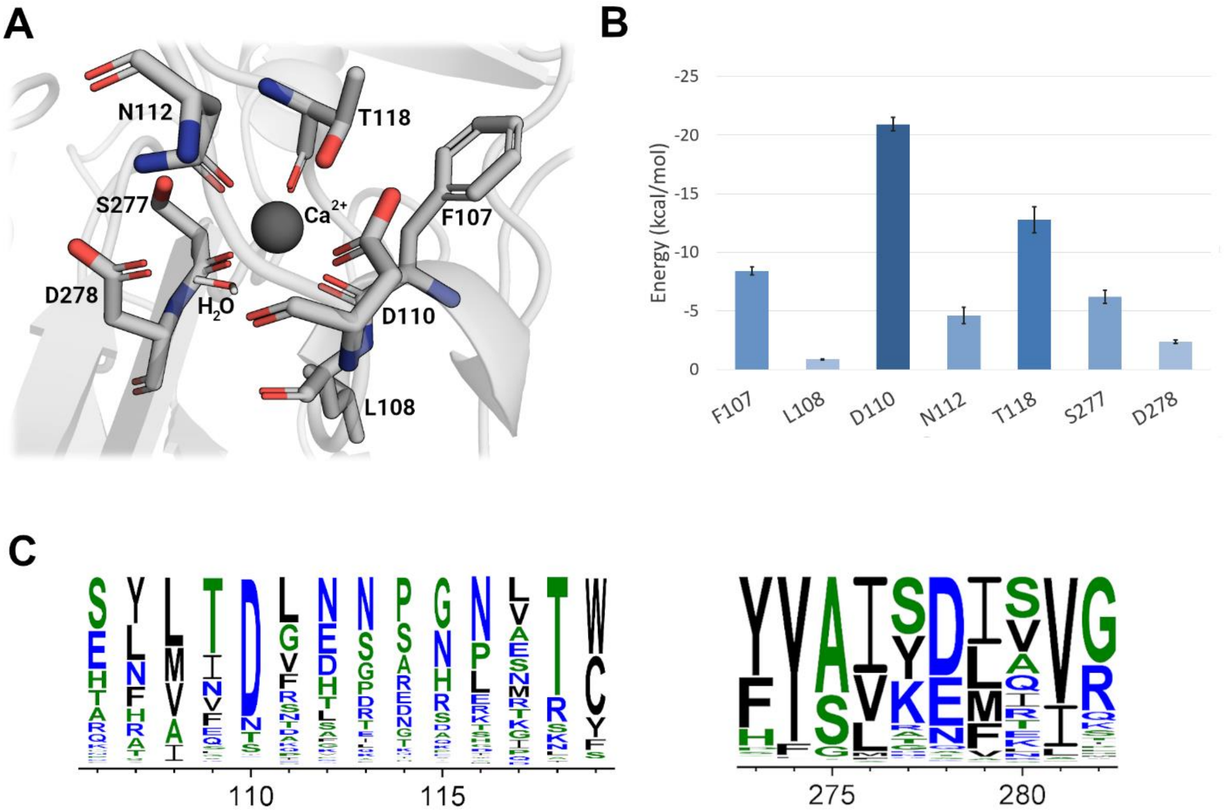


Figure 3.3. Molecular dynamics simulation analysis of the Ca^{2+} binding loop of Net-1. (a) Structural model of the Net-1 Ca^{2+} binding loop with bound to Ca^{+2}. Interacting residues and water over the 100 ns molecular dynamics trajectory are shown. (b) MMPBSA residue decomposition analysis of the ligand interaction. D110 is the major contributor to the interaction while L108 is only minimally involved via its carbonyl backbone group. Error bars are based on the standard deviation of the mean energy for each residue. (c) Sequence conservation of the Ca^{2+} binding interface for 831 Net-1 sequences from various species. The residue positions that contribute to the Ca^{2+} binding through their side chain functional groups are highly conserved.

We then used a bioinformatics approach to explore the importance of individual residues involved in Ca^{2+} coordination to explore this binding region in more detail. Generation and analysis of a multiple sequence alignment of 831 Net-1 related sequences from various species with a minimum sequence coverage of 75% of the target sequence indicates a clear pattern of residue conservation (Figure 3.3C). The multiple sequence alignment reveals that the D110 and T118 residues are also the most highly conserved as Ca^{2+} coordinating residues. N112 despite coordinating Ca^{2+} via the carbonyl of its amide side chain is poorly conserved (Figure 3.3C) but is substituted for amino acids containing carboxyl side chains like glutamate and aspartate, which collectively constitute the majority of residues found at position 112. Amino acids such as F107 and S277, which coordinate through their backbone carbonyl groups, are also found to be poorly conserved. This observation is not unexpected, considering that any amino acid can coordinate Ca^{2+} through its backbone carbonyl, as long as it doesn't interfere with protein folding. Finally, D278 is reasonably well conserved and is commonly substituted for other amino acids with carboxyl or amide containing side chains like glutamate or asparagine which could also serve to hydrogen bond with the coordinating water. The degree of sequence conservation shown in the multiple sequence alignment correlates strongly with the MMPBSA calculated per residue energy contributions, with highly conserved residues like D110 and T118 also making the largest energetic contribution to Ca^{2+} binding.

To expand this analysis of the Net-1 LN domain to other Netrin and Laminin LN domains, 150 sequences of the LN domain Ca^{2+} binding loops for each of the NTN1, NTN3, NTN4, NTNG1, NTNG2, LAMA5, LAMB1 and LAMC1 genes were aligned (Figure 3.4). These genes were chosen since, apart from NTN3 (Netrin-3), crystal structures for the expressed proteins have previously been solved. Most of these proteins are known to bind Ca^{2+} in their LN domain Ca^{2+} binding loop except for Laminin subunit α-5 and Laminin subunit β-1 which do not bind Ca^{2+} (28–31, 55, 60), and Netrin-3 which has not been crystallized or tested for Ca^{2+} binding ability.

From these sequence alignments it is apparent that each of the known Ca^{2+} binding LN domains contain conserved aspartate and threonine residues (equivalent to Net-1 D110 and T118) separated by 7 amino acids, while Laminin subunit β-1 and α-5 lack these conserved residues (Figure 3.4). Examination of the crystal structures of each of these Ca^{2+} binding LN domains confirms the importance of these conserved residues in coordinating Ca^{2+} (28–31, 55). All Ca^{2+} binding LN domains also have a conserved asparagine or glutamate in the helix-loop segment (equivalent to Net-1 N112) except for Netrin-4 which has a conserved serine which does not coordinate Ca^{2+} (28). MMPBSA analysis of the Net-1 Ca^{2+} binding residues shows N112 makes the smallest energy contribution of all directly coordinating amino acids. Together this demonstrates the asparagine or glutamate of the helix-loop segment is less important for Ca^{2+} binding compared to the conserved aspartate and threonine. All LN domains have a conserved aspartate or asparagine on the β-strand involved in Ca^{2+} coordination (equivalent to Net-1 D278).

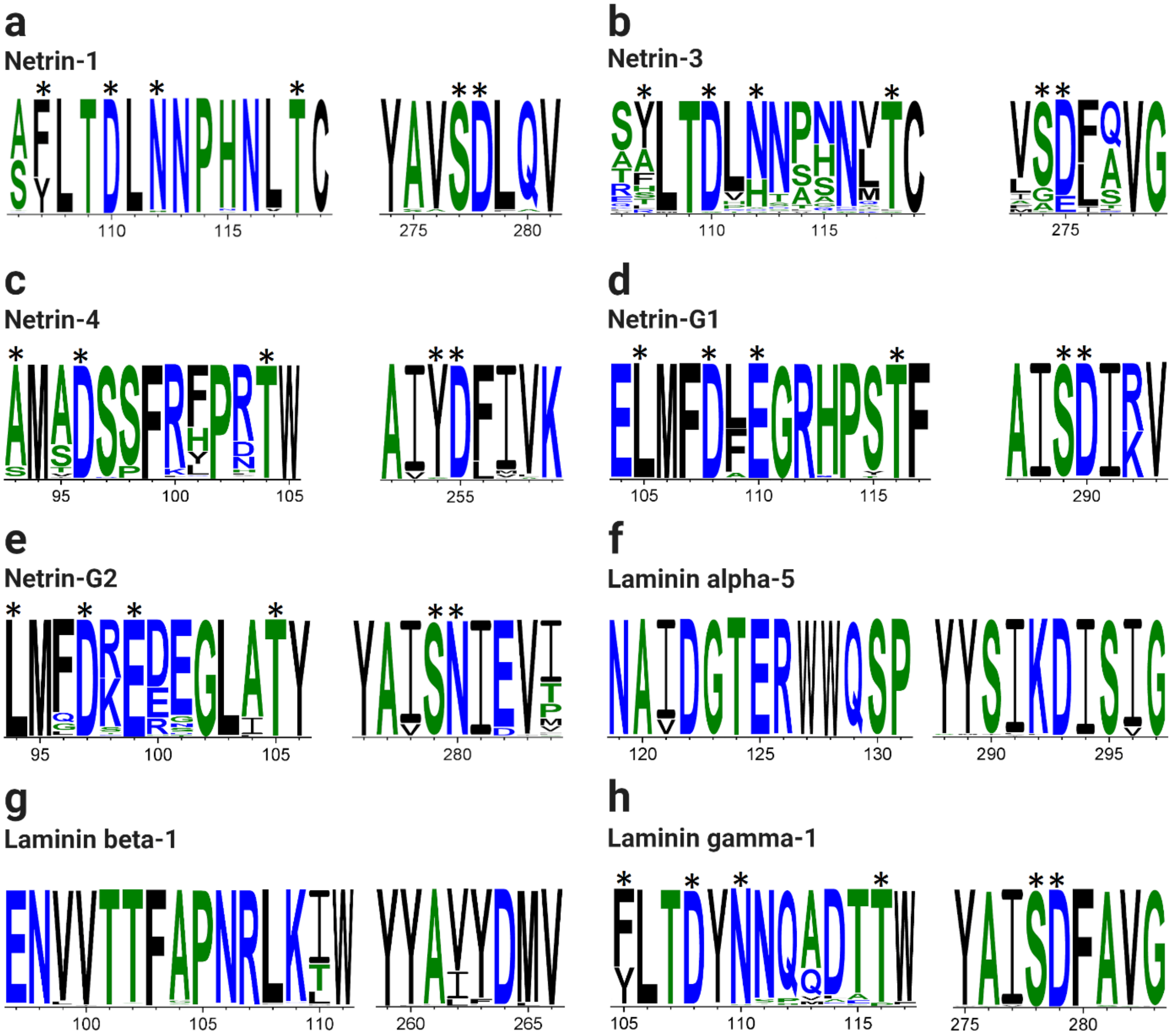


Figure 3.4. Multiple sequence alignment of the Ca2+ binding region of NTN1, NTN3, NTN4, NTNG1, NTNG2, LAMA5, LAMB1 and LAMC1 genes from 150 different species. Conserved residues involved in Ca^{2+} are marked with an asterisk (*).

3.2.5 Discussion

In this work we combine with computational, biophysical and bioinformatic techniques to determine the structural and functional basis of LN domain Ca^{2+} binding using Net-1 as a model. The Net-1 LN domain Ca^{2+} binding site is composed of a short helix-loop segment that provides five Ca^{2+} coordinating ligands. The last β-strand of the LN domain provides the remaining two coordinating ligands which link the helix-loop segment to the of the LN domain β-sandwich through the coordinated Ca^{2+}. In this manner the Ca^{2+} binding site of the Net-1 LN domain is a non-continuous Ca^{2+} binding site, dissimilar to previously characterized Ca^{2+} binding motifs and domains (61–63). Sequence alignment of Net-1 related proteins, as well as other LN domain containing proteins reveals amino acids involved in Ca^{2+} coordination are highly conserved in Net-1 and other Ca^{2+} binding LN domains. In particular the conserved aspartate and threonine of the helix-loop segment are present in all known Ca^{2+} binding LN domains and absent in LN domains which do not bind Ca^{2+} (28–31, 55, 60). Based on the conserved LN domain Ca^{2+} binding sequence identified here, we propose that Netrin-3 which also has this conserved sequence, including the aspartate and threonine, would be capable of binding Ca^{2+}.

Bioinformatic sequence alignments and MMPBSA analysis demonstrate that aspartate and threonine residues of the helix-loop segment are critical for LN domain Ca^{2+} binding. This bidentate coordination of Ca^{2+} by threonine is relatively uncommon but is known to occur in other Ca^{2+} binding proteins like thermolysin (62, 64). Interestingly, several members of the F5/F8 type C domain family which share the β-sandwich jelly roll fold topology with the LN domain also contain a Ca^{2+} binding site formed by a helix-loop segment adjacent to the β-sandwich core (65–69). Comparison of several crystal structures of these Ca^{2+} binding F5/F8 type C domains to the LN domain reveals a remarkably similar method of Ca^{2+} ion coordination. Like the LN domain, Ca^{2+} is coordinated by an aspartate and bidentate threonine in the helix-loop segment. The pentagonal bipyramidal coordination sphere of the Ca^{2+} is completed by backbone carbonyl groups of the helix-loop segment, and a backbone carbonyl and glutamate side chain of a β-strand of the β-sandwich.

To explore the biophysical properties and function of the Net-1 LN domain Ca^{2+} binding site we utilized several biophysical techniques. ITC and TSA results reveal that the Ca^{2+} binding loop of Net-1 and Lam-γ LN domains bind a single Ca^{2+} ion with a low micromolar to nanomolar affinity. Because of this high affinity and the high concentration of free Ca^{2+} in the extracellular environment, these LN domains Ca^{2+} binding sites would be fully saturated under physiological conditions(56).Further biophysical analysis of the Net-1 and Lam-γ - Ca^{2+} interactions by TSA reveals a considerable increase in the T_M of these proteins in the presence of physiological Ca^{2+} concentrations compared to the apo – proteins. This result demonstrates the role of Ca^{2+} in the structural stabilization of the LN domain.

Previous studies exploring the effects of Ca^{2+} on Lam-γ and Net-1 ligand binding have shown that Ca^{2+} is required for structural stabilization of the Ca^{2+} binding loop, rather than facilitating direct contacts between Ca^{2+} and the ligand (18, 27). Ca^{2+} is used for structural stabilization purposes by many other extracellular proteins including E-cadherins and fibrillin-1, which like the LN domains studied here, also bind Ca^{2+} with high affinity (10, 11). The role of E-cadherin Ca^{2+} binding is particularly well studied and has been shown to stabilize the structural conformation of E-cadherins and reduce its susceptibility to proteases (10), while mutations in the Ca^{2+} binding sites of E-cadherins cause disease (4). Given the substantial increase in the T_M of both Net-1 and Lam-γ, it would seem likely that Ca^{2+} binding to the LN domain would similarly stabilize the protein structure and reduce susceptibility to proteolytic degradation. Interestingly though, not all LN domains bind Ca^{2+} and therefore is not strictly required to stabilize all LN folds.

The coordination of Ca^{2+} among the Laminin family thus seems to be important for structural stability and in specific cases, for functional purposes. The data generated by this study presents the unique Ca^{2+} coordination in a highly conserved loop region of known Ca^{2+} binding Laminin proteins. The structural and biophysical approach combined with computational techniques allowed for the comprehensive characterization of the molecular basis for protein Ca^{2+} interactions among the Laminin protein family.

3.2.6 Supplemental Information

Table S3.1. Crystallographic data table of Net-1 in complex with Sm^{3+} (PDB ID: 8SNP).

Parameter	*Net-1 in complex with Sm^{3+}*
Wavelength (Å)	1.0349
Resolution range (Å)	56.62 - 3.4 (3.522 - 3.4)
Space group	P 32 2 1
Unit cell (Å)	69.5 69.5 333.6 90 90 120
Total reflections	111382
Unique reflections	13728 (1347)
Multiplicity	8.3 (8.5)
Completeness (%)	99.35 (99.18)
Mean I/sigma(I)	5.8 (1.9)
Wilson B-factor (Å^2)	90.49
R_{merge}	0.227 (0.930)
R_{meas}	0.257 (1.054)
$CC_{1/2}$	0.994 (0.923)
Reflections used in refinement	13655 (1336)
Reflections used for R-free	615 (60)
R_{work}	0.2514 (0.3784)
R_{free}	0.2719 (0.4399)
Number of non-hydrogen atoms	3238
macromolecules	3140
ligands	183
solvent	0
Protein residues	409
RMS (bonds)	0.006
RMS (angles)	0.91
Ramachandran favored (%)	95.51
Ramachandran allowed (%)	4.49
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.59
Clashscore	14.55
Average B-factor (Å^2)	124.6
macromolecules	122.66
ligands	186.62
Number of TLS groups	3

Table S3.2. Run parameters for the Net-1 and Net-1 + Ca^{+2} MD simulations [a].

Parameter	**Models**	
System	**Net-1**	**Net-1 + Ca^{+2} [b]**
Time	100 ns, 2 fs steps	100 ns, 2 fs steps
Temperature	310 K	310 K
Pressure	1.0 Bar	1.0 Bar
Force field	Amber14 (ff14SB), GLYCAM06j	Amber14 (ff14SB), GLYCAM06j
Electrostatics	Particle-Mesh Ewald	Particle-Mesh Ewald
Coulomb cut-off	1.2 nm	1.2 nm
Temperature coupling	V-rescale	V-rescale
Pressure coupling	Parrinello-Rahman	Parrinello-Rahman
Ions (neutral charge equilibration)	Na^+, Cl^-	Ca^{+2}, Na^+, Cl^-
Total Atoms	420465	420446
Net-1	6401	6401

[a] Additional supporting data and models are available on the Open Science Framework (OSF.IO/C9VQX).

[b] Net-1 D110A and Net-1 T118A mutants were simulated with the same experimental parameters.

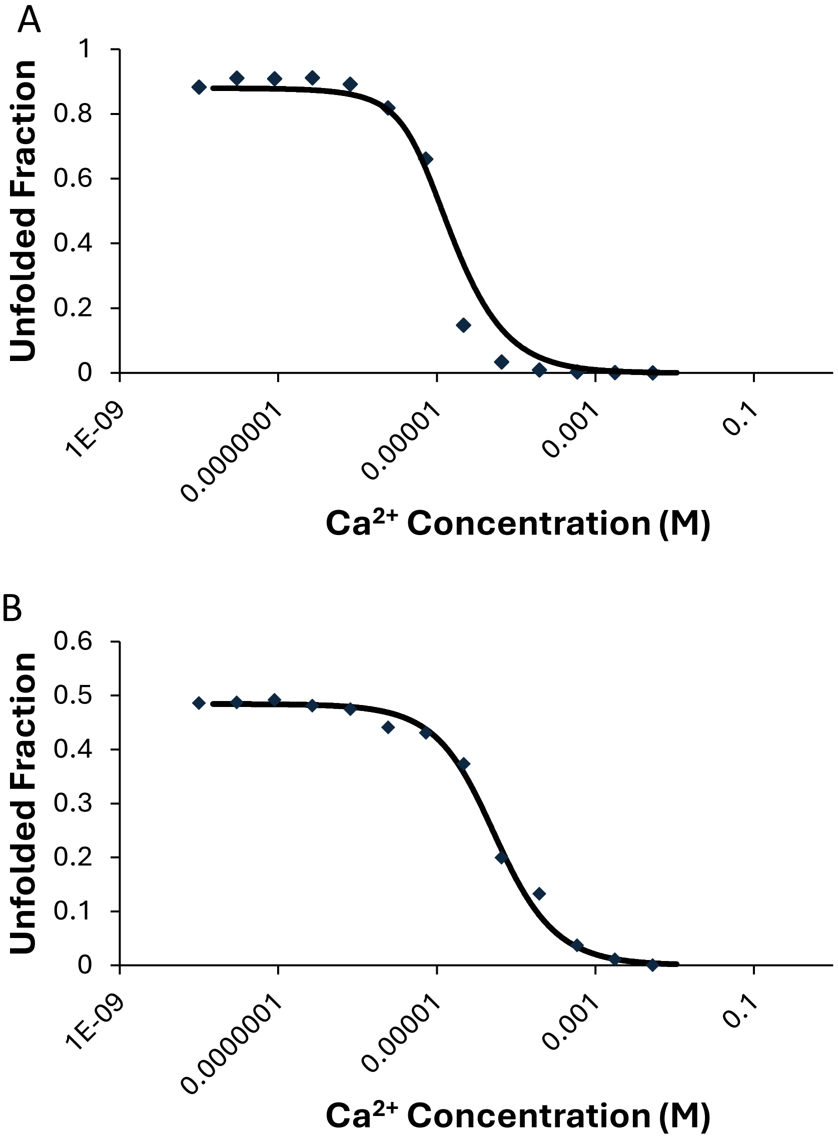


Figure S3.1. Thermal shift assay binding isotherm for protein – Ca^{2+} interaction. (a) Thermal shift assay binding isotherm for Net-1 – Ca^{2+} exhibiting a K_D 1.36 μM asymmetric 95% CI [0.925, 1.94] at 42 °C. (b) Thermal shift assay binding isotherm for Lam-γ – Ca^{2+} exhibiting a K_D of 22.4 uM asymmetric 95% CI [20.4, 24.6] at 47 °C. All experiments are performed in Ca^{2+} free 20 mM HEPES pH 7.3, 150 mM NaCl. Each data point represents the mean of 3 experimental replicates.

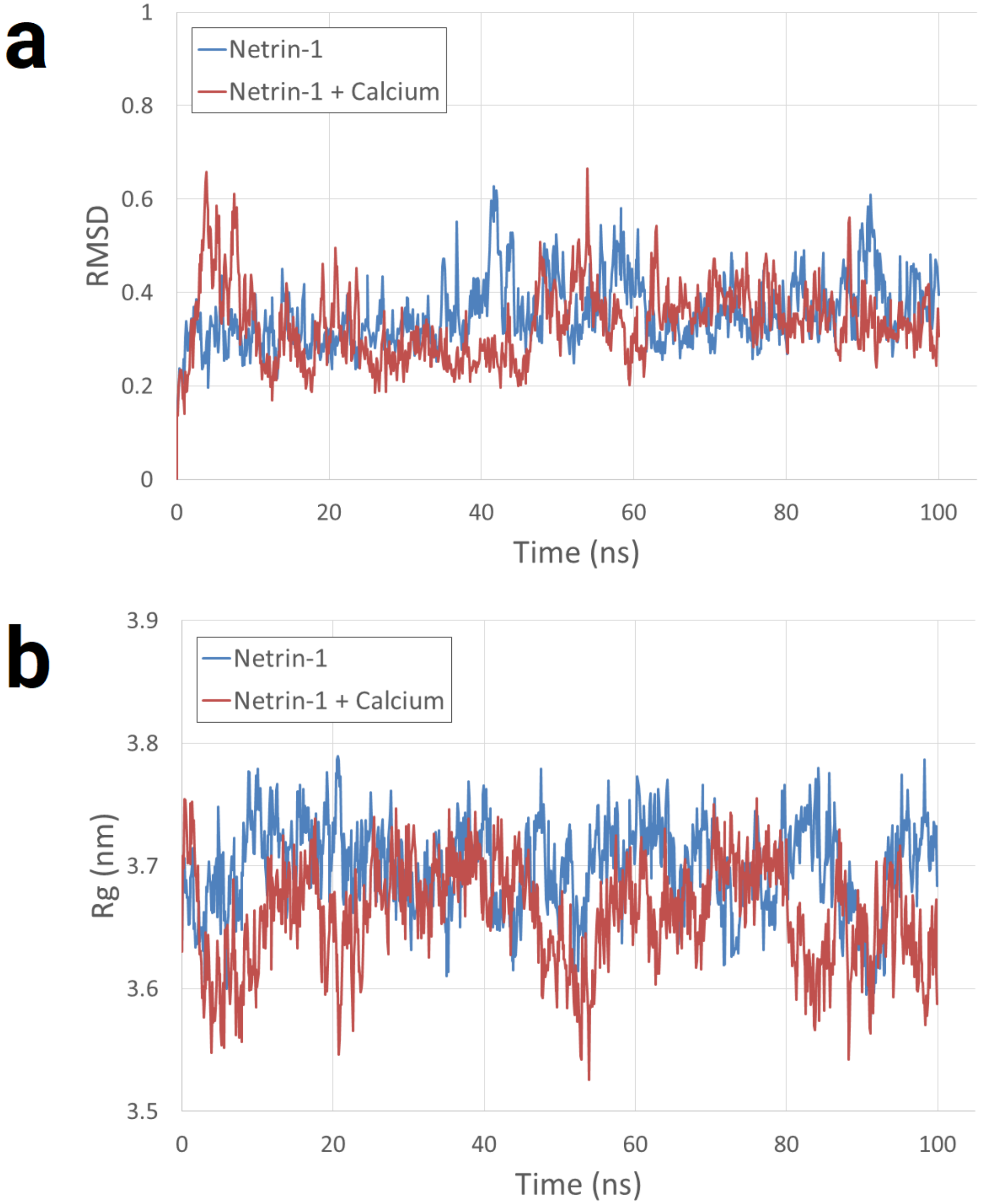


Figure S3.2. Net-1 dynamic properties over a 100 ns molecular dynamics simulation trajectory.

(A) Root-mean squared deviation (RMSD) of Net-1 in the absence and presence of bound Ca^{+2}.

(B) Radius of gyration (R_g) of Net-1 in the absence and presence of bound Ca^{+2}.

3.2.7 References

1. Carafoli, E., Santella, L., Branca, D., and Brini, M. (2001) Generation, control, and processing of cellular calcium signals. *Crit. Rev. Biochem. Mol. Biol.* **36**, 107–260
2. Clapham, D. E. (2007) Calcium Signaling. *Cell*. **131**, 1047–1058
3. Brown, E. M., and MacLeod, R. J. (2001) Extracellular calcium sensing and extracellular calcium signaling. *Physiol. Rev.* **81**, 239–297
4. Handschuh, G., Candidus, S., Luber, B., Reich, U., Schott, C., Oswald, S., Becke, H., Hutzler, P., Birchmeier, W., Höfler, H., and Becker, K.-F. (1999) Tumour-associated E-cadherin mutations alter cellular morphology, decrease cellular adhesion and increase cellular motility. *Oncogene*. **18**, 4301–4312
5. Zhang, K., and Chen, J. (2012) The regulation of integrin function by divalent cations. *Cell Adhes. Migr.* **6**, 20–29
6. Shikamoto, Y., Morita, T., Fujimoto, Z., and Mizuno, H. (2003) Crystal structure of Mg2+- and Ca2+-bound Gla domain of factor IX complexed with binding protein. *J. Biol. Chem.* **278**, 24090–24094
7. Freedman, S. J., Blostein, M. D., Baleja, J. D., Jacobs, M., Furie, B. C., and Furie, B. (1996) Identification of the phospholipid binding site in the vitamin K-dependent blood coagulation protein factor IX. *J. Biol. Chem.* **271**, 16227–16236
8. McGettrick, A. J., Knott, V., Willis, A., and Handford, P. A. (2000) Molecular effects of calcium binding mutations in Marfan syndrome depend on domain context. *Hum. Mol. Genet.* **9**, 1987–1994
9. Maurer, P., and Hohenester, E. (1997) Structural and functional aspects of calcium binding in extracellular matrix proteins. *Matrix Biol.* **15**, 569–580
10. Pokutta, S., Herrenknecht, K., Kemler, R., and Engel, J. (1994) Conformational changes of the recombinant extracellular domain of E-cadherin upon calcium binding. *Eur. J. Biochem.* **223**, 1019–1026
11. Smallridge, R. S., Whiteman, P., Doering, K., Handford, P. A., and Downing, A. K. (1999) EGF-like domain calcium affinity modulated by N-terminal domain linkage in human fibrillin-1. *J. Mol. Biol.* **286**, 661–668
12. Scott, D. L., White, S. P., Otwinowski, Z., Yuan, W., Gelb, M. H., and Sigler, P. B. (1990) Interfacial catalysis: the mechanism of phospholipase A2. *Science*. **250**, 1541–1546
13. Ng, K. K.-S., Drickamer, K., and Weis, W. I. (1996) Structural Analysis of Monosaccharide Recognition by Rat Liver Mannose-binding Protein. *J. Biol. Chem.* **271**, 663–674
14. Hohenester, E., and Yurchenco, P. D. (2013) Laminins in basement membrane assembly. *Cell Adhes. Migr.* **7**, 56–63
15. Cirulli, V., and Yebra, M. (2007) Netrins: beyond the brain. *Nat. Rev. Mol. Cell Biol.* **8**, 296–306
16. Rajasekharan, S., and Kennedy, T. E. (2009) The netrin protein family. *Genome Biol.* **10**, 239
17. McKee, K. K., Hohenester, E., Aleksandrova, M., and Yurchenco, P. D. (2021) Organization of the laminin polymer node. *Matrix Biol.* **98**, 49–63
18. Kulczyk, A. W., McKee, K. K., Zhang, X., Bizukojc, I., Yu, Y. Q., and Yurchenco, P. D. (2023) Cryo-EM reveals the molecular basis of laminin polymerization and LN-lamininopathies. *Nat. Commun.* **14**, 317
19. Fahey, B., and Degnan, B. M. (2012) Origin and Evolution of Laminin Gene Family Diversity. *Mol. Biol. Evol.* **29**, 1823–1836
20. Leclère, L., and Rentzsch, F. (2012) Repeated Evolution of Identical Domain Architecture in Metazoan Netrin Domain-Containing Proteins. *Genome Biol. Evol.* **4**, 883–899

21. Serafini, T., Kennedy, T. E., Gaiko, M. J., Mirzayan, C., Jessell, T. M., and Tessier-Lavigne, M. (1994) The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. *Cell*. **78**, 409–424
22. Kennedy, T. E., Serafini, T., de la Torre, J. R., and Tessier-Lavigne, M. (1994) Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell*. **78**, 425–435
23. Wang, H., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Tessier-Lavigne, M. (1999) Netrin-3, a Mouse Homolog of Human NTN2L, Is Highly Expressed in Sensory Ganglia and Shows Differential Binding to Netrin Receptors. *J. Neurosci.* **19**, 4938–4947
24. Nakashiba, T., Nishimura, S., Ikeda, T., and Itohara, S. (2002) Complementary expression and neurite outgrowth activity of netrin-G subfamily members. *Mech. Dev.* **111**, 47–60
25. Nakashiba, T., Ikeda, T., Nishimura, S., Tashiro, K., Honjo, T., Culotti, J. G., and Itohara, S. (2000) Netrin-G1: a Novel Glycosyl Phosphatidylinositol-Linked Mammalian Netrin That Is Functionally Divergent from Classical Netrins. *J. Neurosci.* **20**, 6540–6550
26. Koch, M., Murrell, J. R., Hunter, D. D., Olson, P. F., Jin, W., Keene, D. R., Brunken, W. J., and Burgeson, R. E. (2000) A novel member of the netrin family, beta-netrin, shares homology with the beta chain of laminin: identification, expression, and functional characterization. *J. Cell Biol.* **151**, 221–234
27. Xu, K., Wu, Z., Renier, N., Antipenko, A., Tzvetkova-Robev, D., Xu, Y., Minchenka, M., Nardi-Dei, V., Rajashankar, K. R., Himanen, J., Tessier-Lavigne, M., and Nikolov, D. B. (2014) Structures of netrin-1 bound to two receptors provide insight into its axon guidance mechanism. *Science*. **344**, 1275–1279
28. Reuten, R., Patel, T. R., McDougall, M., Rama, N., Nikodemus, D., Gibert, B., Delcros, J.-G., Prein, C., Meier, M., Metzger, S., Zhou, Z., Kaltenberg, J., McKee, K. K., Bald, T., Tüting, T., Zigrino, P., Djonov, V., Bloch, W., Clausen-Schaumann, H., Poschl, E., Yurchenco, P. D., Ehrbar, M., Mehlen, P., Stetefeld, J., and Koch, M. (2016) Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nat. Commun.* **7**, 13515
29. Brasch, J., Harrison, O. J., Ahlsen, G., Liu, Q., and Shapiro, L. (2011) Crystal Structure of the Ligand Binding Domain of Netrin G2. *J. Mol. Biol.* **414**, 723–734
30. Seiradake, E., Coles, C. H., Perestenko, P. V., Harlos, K., McIlhinney, R. A. J., Aricescu, A. R., and Jones, E. Y. (2011) Structural basis for cell surface patterning through NetrinG–NGL interactions. *EMBO J.* **30**, 4479–4488
31. Grandin, M., Meier, M., Delcros, J. G., Nikodemus, D., Reuten, R., Patel, T. R., Goldschneider, D., Orriss, G., Krahn, N., Boussouar, A., Abes, R., Dean, Y., Neves, D., Bernet, A., Depil, S., Schneiders, F., Poole, K., Dante, R., Koch, M., Mehlen, P., and Stetefeld, J. (2016) Structural decoding of the Netrin-1/UNC5 interaction and its therapeutical implications in cancers. *Cancer Cell*. **29**, 173–185
32. Moya-Torres, A., Gupta, M., Heide, F., Krahn, N., Legare, S., Nikodemus, D., Imhof, T., Meier, M., Koch, M., and Stetefeld, J. (2021) Homogenous overexpression of the extracellular matrix protein Netrin-1 in a hollow fiber bioreactor. *Appl. Microbiol. Biotechnol.* **105**, 6047–6057
33. Heide, F., Legare, S., To, V., Gupta, M., Gabir, H., Imhof, T., Moya-Torres, A., McDougall, M., Meier, M., Koch, M., and Stetefeld, J. (2022) Heparins mediate the multimer assembly of secreted Noggin. *Protein Sci.* **31**, e4419
34. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
35. Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley, P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L., Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K.,

Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Malý, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray, J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito, M., Sánchez Rodríguez, F., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T., Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M., and Yamashita, K. (2023) The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. Sect. Struct. Biol.* **79**, 449–461

36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
37. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
38. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877
39. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25
40. Huang, J., and MacKerell, A. D. (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145
41. Mark, P., and Nilsson, L. (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A*. **105**, 9954–9960
42. Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., and Moreno, E. (2021) gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J. Chem. Theory Comput.* **17**, 6281–6291
43. Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., and Zaretskaya, I. (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* **41**, W29–W33
44. Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., and Madden, T. L. (2012) Domain enhanced lookup time accelerated BLAST. *Biol. Direct*. **7**, 1–14
45. Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014) Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife*. 10.7554/eLife.02030
46. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
47. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190
48. Bai, N., Roder, H., Dickson, A., and Karanicolas, J. (2019) Isothermal analysis of ThermoFluor data can readily provide quantitative binding affinities. *Sci. Rep.* **9**, 2650

49. Niebling, S., Burastero, O., Bürgi, J., Günther, C., Defelipe, L. A., Sander, S., Gattkowski, E., Anjanappa, R., Wilmanns, M., Springer, S., Tidow, H., and García-Alai, M. (2021) FoldAffinity: binding affinities from nDSF experiments. *Sci. Rep.* **11**, 9572
50. Burastero, O., Niebling, S., Defelipe, L. A., Günther, C., Struve, A., and Garcia Alai, M. M. (2021) eSPC: an online data-analysis platform for molecular biophysics. *Acta Crystallogr. Sect. Struct. Biol.* **77**, 1241–1250
51. Robinson, R. A., Griffiths, S. C., Haar, L. L. van de, Malinauskas, T., Battum, E. Y. van, Zelina, P., Schwab, R. A., Karia, D., Malinauskaite, L., Brignani, S., Munkhof, M. H. van den, Düdükcü, Ö., Ruiter, A. A. D., Heuvel, D. M. A. V. den, Bishop, B., Elegheert, J., Aricescu, A. R., Pasterkamp, R. J., and Siebold, C. (2021) Simultaneous binding of guidance cues NET1 and RGM blocks extracellular NEO1 signaling. *Cell*. **184**, 2103-2120.e31
52. Carvin, D., S.A. Islam, M.J.E. Sternberg, and T.L. Blundell. 2012. The preparation of heavy-atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering. In: International Tables for Crystallography Volume F: Crystallography of biological macromolecules. Arnold E, DM Himmel, MG Rossmann, editors. Wiley/Chichester
53. Meier, M., Gupta, M., Akgül, S., McDougall, M., Imhof, T., Nikodemus, D., Reuten, R., Moya-Torres, A., To, V., Ferens, F., Heide, F., Padilla-Meier, G. P., Kukura, P., Huang, W., Gerisch, B., Mörgelin, M., Poole, K., Antebi, A., Koch, M., and Stetefeld, J. (2023) The dynamic nature of netrin-1 and the structural basis for glycosaminoglycan fragment-induced filament formation. *Nat. Commun.* **14**, 1226
54. Priest, J.M., and E. Ozkan. 2023. Structure of C. elegans UNC-6 LamN and EGF domains. *RCSB Protein Data Bank*. 17 October 2023. https://www.rcsb.org/
55. Carafoli, F., Hussain, S.-A., and Hohenester, E. (2012) Crystal Structures of the Network-Forming Short-Arm Tips of the Laminin β1 and γ1 Chains. *PLOS ONE*. **7**, e42473
56. Schaafsma, G. 1988. Calcium in extracellular fluid: Homeostasis. In: Calcium in human biology. ILSI hman nutrition reviews. Nordin B.E.C, editor. Springer/ London. pp 241-259
57. Paulsson, M. (1988) The role of Ca2+ binding in the self-aggregation of laminin-nidogen complexes. *J. Biol. Chem.* **263**, 5425–5430
58. Choong, G., Liu, Y., and Templeton, D. M. (2014) Interplay of calcium and cadmium in mediating cadmium toxicity. *Chem. Biol. Interact.* **211**, 54–65
59. Goyer, R. A. (1995) Nutrition and metal toxicity. *Am. J. Clin. Nutr.* **61**, 646S-650S
60. Hussain, S.-A., Carafoli, F., and Hohenester, E. (2011) Determinants of laminin polymerization revealed by the structure of the α5 chain amino-terminal region. *EMBO Rep.* **12**, 276–282
61. McPhalen, C.A., N.C.J. Strynadka, and M.N.G. James. 1991. Calcium-binding sites in proteins: A structural perspective. In: Advances in protein chemistry. Anfinsen CB, JT Edsall, FM Richards, DS Eisenberg, editors. Springer/ London. pp. 77-144
62. Kirberger, M., and J.J. Yang. 2013. Calcium-Binding Protein Site Types. In: Encyclopedia of Metalloproteins. Kretsinger RH, VN Uversky, EA Permyakov, editors. Springer/ New York. pp. 511–521.
63. Elíes, J., M. Yáñez, T.M.C. Pereira, J. Gil-Longo, D.A. MacDougall, and M. Campos-Toimil. 2020. An update to calcium binding proteins. In: Calcium Signaling. Islam M, editor. Springer/ London pp. 183–213.
64. Holden, H. M., and Matthews, B. W. (1988) The binding of L-valyl-L-tryptophan to crystalline thermolysin illustrates the mode of interaction of a product of peptide hydrolysis. *J. Biol. Chem.* **263**, 3256–3260
65. Ficko-Blean, E., Stuart, C. P., Suits, M. D., Cid, M., Tessier, M., Woods, R. J., and Boraston, A. B. (2012) Carbohydrate Recognition by an Architecturally Complex α-N-Acetylglucosaminidase from Clostridium perfringens. *PLOS ONE*. **7**, e33524

66. Ficko-Blean, E., and Boraston, A. B. (2006) The Interaction of a Carbohydrate-binding Module from a Clostridium perfringens N-Acetyl-β-hexosaminidase with Its Carbohydrate Receptor. *J. Biol. Chem.* **281**, 37748–37757
67. Feil, S. C., Lawrence, S., Mulhern, T. D., Holien, J. K., Hotze, E. M., Farrand, S., Tweten, R. K., and Parker, M. W. (2012) Structure of the Lectin Regulatory Domain of the Cholesterol-Dependent Cytolysin Lectinolysin Reveals the Basis for Its Lewis Antigen Specificity. *Structure*. **20**, 248–258
68. Ramelot, T. A., Raman, S., Kuzin, A. P., Xiao, R., Ma, L.-C., Acton, T. B., Hunt, J. F., Montelione, G. T., Baker, D., and Kennedy, M. A. (2009) Improving NMR protein structure quality by Rosetta refinement: A molecular replacement study. *Proteins Struct. Funct. Bioinforma.* **75**, 147–167
69. Rogers, M. S., Hurtado-Guerrero, R., Firbank, S. J., Halcrow, M. A., Dooley, D. M., Phillips, S. E. V., Knowles, P. F., and McPherson, M. J. (2008) Cross-Link Formation of the Cysteine 228−Tyrosine 272 Catalytic Cofactor of Galactose Oxidase Does Not Require Dioxygen. *Biochemistry*. **47**, 10428–10439

Chapter 4

Method development for studying enzyme-ligand interactions

Adapted from: Legare, S., Heide, F., Bailey-Elkin, B.A., and Stetefeld, J. (2022). Improved SARS-CoV-2 main protease high-throughput screening assay using a 5-carboxyfluorescein substrate. *Journal of Biological Chemistry* **298**, 101739.

Contribution statement

All authors were involved in study design. S.L. produced M^{pro}, produced and analysed DSF, DLS and enzymatic data. F.H. performed CD measurements and data analysis. S.L. generated figures, tables and wrote the manuscript. S.L., F.H., B.A.B.E. and J.S. contributed to the editorial process of the final manuscript.

4.1

The significance of methodological advancements in enzyme-ligand interaction analysis

Because protein – ligand interactions can often be desirable drug targets a great deal of research is aimed at identifying inhibitors of protein- ligand interactions. Biophysical assays are often employed to identify inhibitors of protein – ligand interactions because of their versatility and ease of implementation to high throughput screening. They can also easily be modified to study the molecular mechanism of identified inhibitors.

To ensure a robust and repeatable method is developed for identifying inhibitors and characterizing inhibitors, it is useful to perform a basic yet through biophysical and biochemical characterization of the protein – ligand interaction being studied. In particular it is important to characterize the quality of the proteins used in the assays. Small changes in protein structure and composition can have a substantial effect on protein function and it is therefore important to characterize these properties of proteins. Other variables such as temperature, pH, ions and small molecules can also impact protein function and should also be characterized.

In this chapter we perform the biochemical, biophysical and enzymatic characterization of the SARS-CoV-2 Main protease and use this characterization to guide the development of an improved high throughput screening assay. In doing so we demonstrate the process of developing and biophysical assay for identification and characterization of protein – ligand interaction inhibitors, which is a critical step in any project seeking to develop therapeutics targeting protein – ligand interactions.

4.2

Improved SARS-CoV-2 main protease high-throughput screening assay using a 5-carboxyfluorescein substrate.

4.2.1 Abstract

The emergence of SARS-CoV-2 as a global threat to human health has highlighted the need for the development of novel therapies targeting current and emerging coronaviruses with pandemic potential. The coronavirus main protease (M^{pro}, also called $3CL^{pro}$) is a validated drug target against coronaviruses and has been heavily studied since the emergence of SARS-CoV-2 in late 2019. We report the biophysical and enzymatic characterization of native M^{pro}, then characterize the steady-state kinetics of several commonly used fluorescence resonance energy transfer substrates, fluorogenic substrates, and 6 of the 11 SARS-CoV-2 polyprotein cleavage sequences. We assess the suitability of these substrates for high throughput screening. Guided by our assessment of these substrates, we develop an improved 5-carboxyfluorescein based fluorescence resonance energy transfer substrate which is better suited for high throughput screening and is less susceptible to interference and false positives than existing substrates. This study provides a useful framework for the design of coronavirus M^{pro} enzyme assays to facilitate the discovery and development of therapies targeting M^{pro}.

4.2.2 Introduction

Coronaviruses are a family of viruses commonly found in wildlife, companion animals, livestock, and humans. Human coronaviruses include HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1 continuously circulate in the population and mostly cause mild symptoms associated with the common cold. In contrast, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and SARS-CoV-2 are highly pathogenic coronaviruses causing the SARS epidemic, MERS epidemic and most recently the coronavirus disease 2019 (COVID-19) pandemic(1). To date there have been over 220 million reported cases of COVID-19 and 4.6 million reported deaths. Despite the development of efficacious vaccines against COVID-19, SARS-CoV-2 transmission continues at high levels and case numbers continue to increase(2, 3). As a result, there is an urgent need for effective antiviral drugs targeting SARS-CoV-2 that can be used to treat COVID-19.

SARS-CoV-2 is an enveloped positive-stranded RNA virus which infects cells of the upper and lower respiratory tract. Upon entry into the host cell, the viral RNA genome is translated into two polyproteins, pp1a and pp1ab. These polyproteins are proteolytically processed by two viral proteases, a papain-like protease (PL^{pro}) and a chymotrypsin-like main protease (M^{pro}, also called $3CL^{pro}$), releasing 16 non-structural proteins(1). M^{pro} inhibitors can effectively block SARS-CoV-2 replication in cell culture, demonstrating M^{pro} is a valid drug target(4–8). SARS-CoV-2 M^{pro} has been shown to preferentially recognize the A-X-L-Q-(A/S) consensus sequence (where X is any amino acid), with cleavage occurring after the glutamine(9, 10). Interestingly, other coronaviruses including the related SARS-CoV and MERS-CoV share similar substrate specificity with SARS-CoV-2 M^{pro}, suggesting that inhibitors of SARS-CoV-2 M^{pro} could serve as broad spectrum antiviral drugs against future epidemic or pandemic causing coronaviruses(4, 11).

Discovery of M^{pro} inhibitors has relied heavily on the use of high throughput screens (HTS) using a fluorescence resonance energy transfer (FRET) based peptide substrate to monitor protease activity(5, 7, 12–17). Fluorogenic substrates which cleave an amino-coumarin based fluorophore attached to the carboxyl terminus of a peptide have also been used(14, 15, 18). A number of M^{pro} enzyme assays have been developed using different substrates, M^{pro} constructs, and buffer conditions(14, 15, 19–22). As a result, there have been inconsistent findings regarding the identification of potential M^{pro} inhibitors(20, 21, 23). This has highlighted a clear need for an improved SARS-CoV-2 M^{pro} assay that delivers better performance and improved consistency.

Here we provide the detailed biophysical and enzymatic characterization of SARS-CoV-2 M^{pro} with native N- and C-termini, and asses the steady-state kinetic parameters of three commonly used SARS-CoV-2 M^{pro} fluorescent substrates(4, 7, 18). We measure the kinetic efficiency of 6 SARS-CoV2 M^{pro} polyprotein cleavage sequences to determine the optimal substrate amino acid sequence. Guided by these results, an improved 5-carboxyfluorescein based FRET substrate was developed that is better suited for HTS and is less susceptible to interference and false positives than previously reported substrates. This study provides a useful framework for the design assays aimed at discovering and developing new coronavirus M^{pro} inhibitors.

4.2.3 Results

Both fluorogenic and FRET based substrates were used in this work (Figure 4.1). The previously reported VKLQ - AMC fluorogenic substrate consists of a short peptide with a cleavable 7-amino-4-methylcoumarin (AMC) fluorophore at the P1’ position (Figure 4.1*A*) (18). The FRET substrates consist of a fluorophore and quencher pair separated by a SARS-CoV-2 polyprotein cleavage sequence. The previously reported FRET nsp4-5-EDANS substrate uses an EDANS fluorophore and a DABCYL quencher (Figure 4.1*B*), while the nsp4-5-MCA substrate uses a MCA fluorophore with a lysine-2,4-dinitrophenyl quencher (Figure 4.1*C*)(4, 7). The 5-carboxyfluorescein (FAM) based FRET substrate developed in this work uses a DABCYL quencher (Figure 4.1*D*). Table 4.2 lists the amino acid sequence, cleavage site location, fluorophore, and quencher used for each substrate tested in this work.

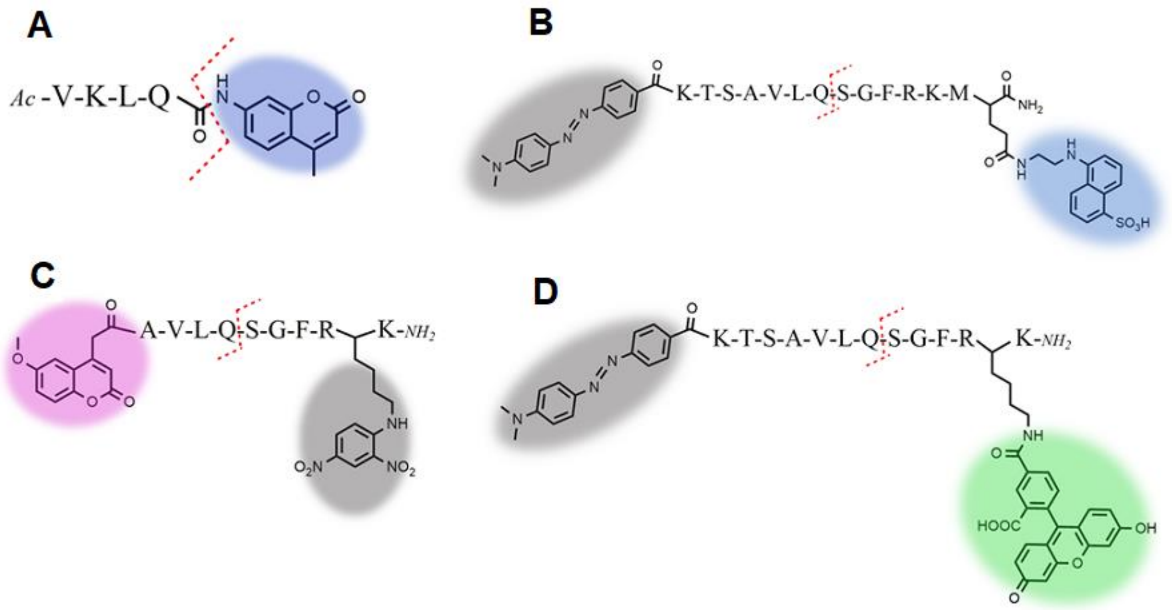


Figure 4.1. Structure and names of fluorescent substrates used in this study. *A*, Fluorogenic VKLQ – AMC substrate, with AMC fluorophore shown in blue substrate. *B*, nsp4-5-EDANS FRET substrate uses a 5-(2-Aminoethylamino)-1-naphthalenesulfonic acid (EDANS) fluorophore shown in blue, and a 4-(4-dimethylaminophenylazo)benzoic acid (DABCYL) quencher shown in gray. *C*, nsp4-5-MCA FRET substrate uses a 7-methoxycoumarin-4-acetic acid (MCA) fluorophore shown in purple with a lysine-2,4-dinitrophenyl (lys(DNP)) quencher shown in grey. *D*, nsp4-5-FAM FRET substrate uses a 5-carboxyfluorescein fluorophore shown in green with a DABCYL quencher shown in grey. The red dashed line indicates the cleavage site on the substrates.

Biophysical characterization demonstrates excellent protein quality

Following the method developed by Xue et al. SARS-CoV-2 M^{pro} with native N- and C-termini was expressed and purified to apparent homogeneity (Figure S4.1) for further biophysical and enzymatic characterization(24). Differential scanning fluorimetry (DSF), dynamic light scattering (DLS), and circular dichroism (CD) spectroscopy were used to establish the quality of M^{pro} used in this work. CD measurements confirmed that the M^{pro} secondary structure is in agreement with what is expected based on the crystal structure (PDB ID: 6Y2E) (Figure 4.2*A* and Table 4.1). DLS was used to measure the hydrodynamic radius (R_H) of M^{pro}, and to measure the state of M^{pro} self oligomerization (Figure 4.2*B*). Using the size distribution analysis model, the major intensity peak had an R_H of 3.76 ± 0.14 nm with a polydispersity index of 0.18 ± 0.03. The measured R_H of 3.76 ± 0.14 nm agrees with the expected R_H for the M^{pro} dimer based on the crystal structure (PDB ID: 6Y2E). The DSF melt curve measured by the 350/330 nm fluorescence ratio showed a melting onset beginning at 51.1°C and an inflection point or melting point (T_m) of 56.8°C (Figure 4.2*C*). The baseline turbidity measurement is stable from 20°C until the onset in turbidity increased beginning at 47.9°C, indicating M^{pro} was stable against aggregation until 3.2°C before the onset of melting begins (Figure 4.2*D*). The inflection point of the turbidity measurement was 56.8°C corresponding to the T_m of M^{pro}. The measured T_m of 56.8°C was slightly higher than previously reported values of between 48.5 to 54.2°C(25, 26). Taken together these results show M^{pro} is highly pure, properly folded, thermodynamically stable, and monodisperse in solution with very little aggregation or higher order oligomerization present.

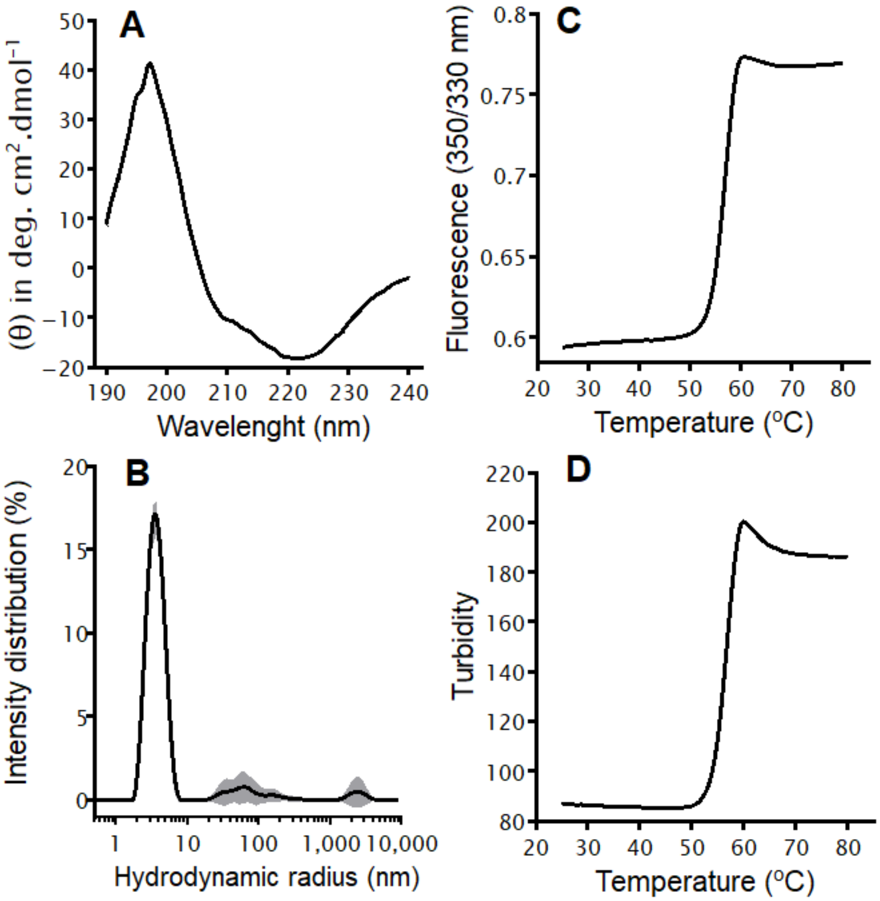


Figure 4.2. Biophysical characterization of SARS-CoV-2 M^{pro}**.** *A*, CD measurement showing secondary structure content of M^{pro}. *B*, DLS intensity distribution with major peak showing a hydrodynamic radius (R_H) of 3.76 ± 0.14 nm corresponding to M^{pro}, black line shows mean intensity distribution with ± 1 standard deviation shown in grey shaded area, n = 10. *C*, 350/330nm fluorescence ratio shows a melting onset at 51.1°C and a T_m of 56.8°C. *D*, Turbidity shows an onset in aggregation beginning at 47.9°C with an inflection point at 56.8°C.

Table 4.1. SARS-CoV-2 M^{pro} circular dichroism spectra analysis.

	Helix (%)	Sheet (%)	Turn (%)	Random coil (%)	Other[c] (%)
CDSSTR[a]	29	26	19	28	
PDB ID: 6Y2E[b]	25	27			48

[a] secondary structure content of CD spectra estimated by CDSSTR algorithm. [b] secondary structure of crystal structure estimated using PDBsum. [c] includes turns and random coils.

Optimization of assay buffer conditions

The effects of different buffer conditions on M^{pro} activity were assessed to determine optimal assays conditions. The AKLQ - AMC substrate was chosen for buffer optimization because it displayed good solubility up to a concentration of around 2.5 mM in all buffers tested. The optimal pH for M^{pro} was found to be pH 7.0 (Figure 4.3*A*). M^{pro} had a strong preference for $NaPO_4$ as a buffering agent, followed by BIS-TRIS, HEPES and TRIS, when tested at 20 mM buffering agent, 150 mM NaCl, pH 7.0 (Figure 4.3*B*), however this preference was lessened when tested at 20 mM buffering agent pH 7.0 without NaCl (Figure 4.3*C*). The highest enzyme activity was achieved when NaCl was omitted from the buffer while adding between 10 to 300 mM NaCl decreases the enzyme activity roughly the same degree (Figure 4.3*D*). Both glycerol and DMSO were found to have a negative effect on enzyme activity (Figures 4.3*E* and 4.3*F*). Based on these results the optimized buffer found in this study consists of 20 mM $NaPO_4$ at pH 7.0. It was found that the FRET substrates used in this work showed better solubility in buffers of low ionic strength, therefore 20 mM BIS-TRIS pH 7.0 was used instead of a $NaPO_4$ based buffer. These optimized buffer conditions closely agree with other work showing that M^{pro} is most thermodynamically stable at pH 7.0 in the absence of NaCl(26).

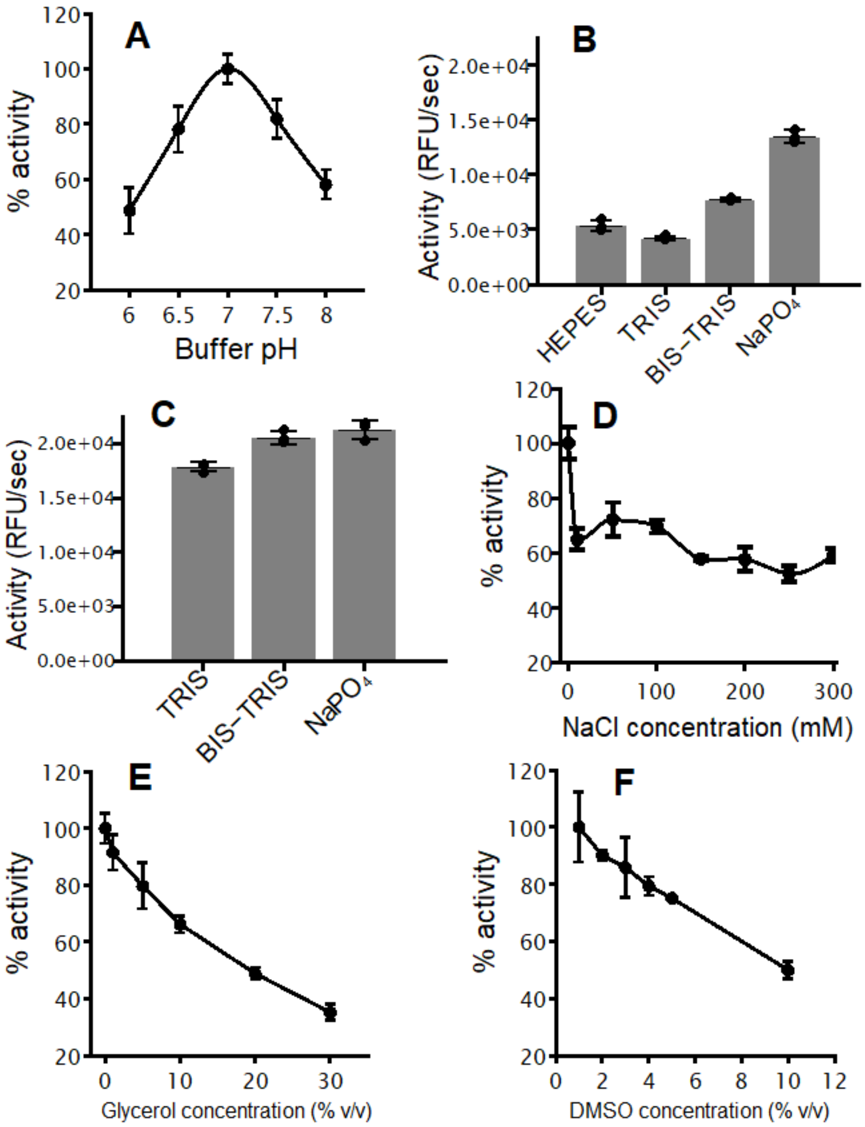


Figure 4.3. Assessing the effects of common buffer conditions on M^{pro} initial rate. All tests are done in 100 µL buffer containing 100 µM VKLQ - AMC substrate and 200 nM enzyme. *A*, pH optimization in 20 mM $NaPO_4$ pH 6.0 to 8.0, 150 mM NaCl. *B*, Preference of M^{pro} for HEPES, TRIS, BIS-TRIS or $NaPO_4$, in 20 mM buffering agent, pH 7.0 with 150 mM NaCl. *C*, Preference of M^{pro} for TRIS, BIS-TRIS, and $NaPO_4$ in 20 mM buffering agent, pH 7.0. *D*, Effect of 0 to 300 mM NaCl in 20 mM $NaPO_4$ pH 7.0. *E*, Effect of 0 to 30% v/v glycerol in 20 mM $NaPO_4$ pH 7.0, 150 mM NaCl. *F*, Effect of 1 to 10% v/v DMSO in 20 mM $NaPO_4$ pH 7.0, 150 mM NaCl. Each measurement is reported as the mean with error bars showing ± 1 standard deviation, n = 3.

Measuring the reaction progress curve for the complete hydrolysis of substrate can help assess and identify non-optimal buffer conditions and loss of enzyme activity due to inactivation and inhibition. It is also helpful to verify that the measured initial rate corresponds to the true linear steady-state portion of the reaction progress curve(27). When hydrolysing the FRET substrates, the M^{pro} began to lose activity after about 400 seconds of reaction time in 20 mM BIS-TRIS pH 7.0 (Figure 4.4). Adding 2 mM EDTA and 2mM DTT to the assay buffer prevented this inactivation for the nsp4-5-FAM and nsp4-5-EDANS substrates (Figure 4.4*B*-*C*), but not for the nsp4-5-MCA substrate (Figure 4.4*D*) which was inactive in 20 mM BIS-TRIS pH 7.0, 2 mM EDTA. As previously discussed, the FRET substrates showed reduced solubility in buffers with higher ionic strength. The inactivity of the nsp4-5-MCA substrate in the 20 mM BIS-TRIS pH 7.0, 2 mM EDTA buffer may be caused by the increase in ionic strength of the buffer from the addition of 2 mM EDTA, reducing the solubility of the nsp4-5-MCA substrate. In contrast to the FRET substrates, M^{pro} did not fully lose activity when hydrolysing the VKLQ - AMC substrate (Figure 4.4*A*), and addition of EDTA and DTT to the reaction buffer had a minimal effect on the enzyme's behaviour. The linear initial rate portion of the reaction progress curve could be measured for at least the first 600 seconds of reaction time with the VKLQ - AMC substrate, and about 200 seconds for the FRET substrates. This initial rate was unaffected by the addition of EDTA and DTT (Figure 4.4).

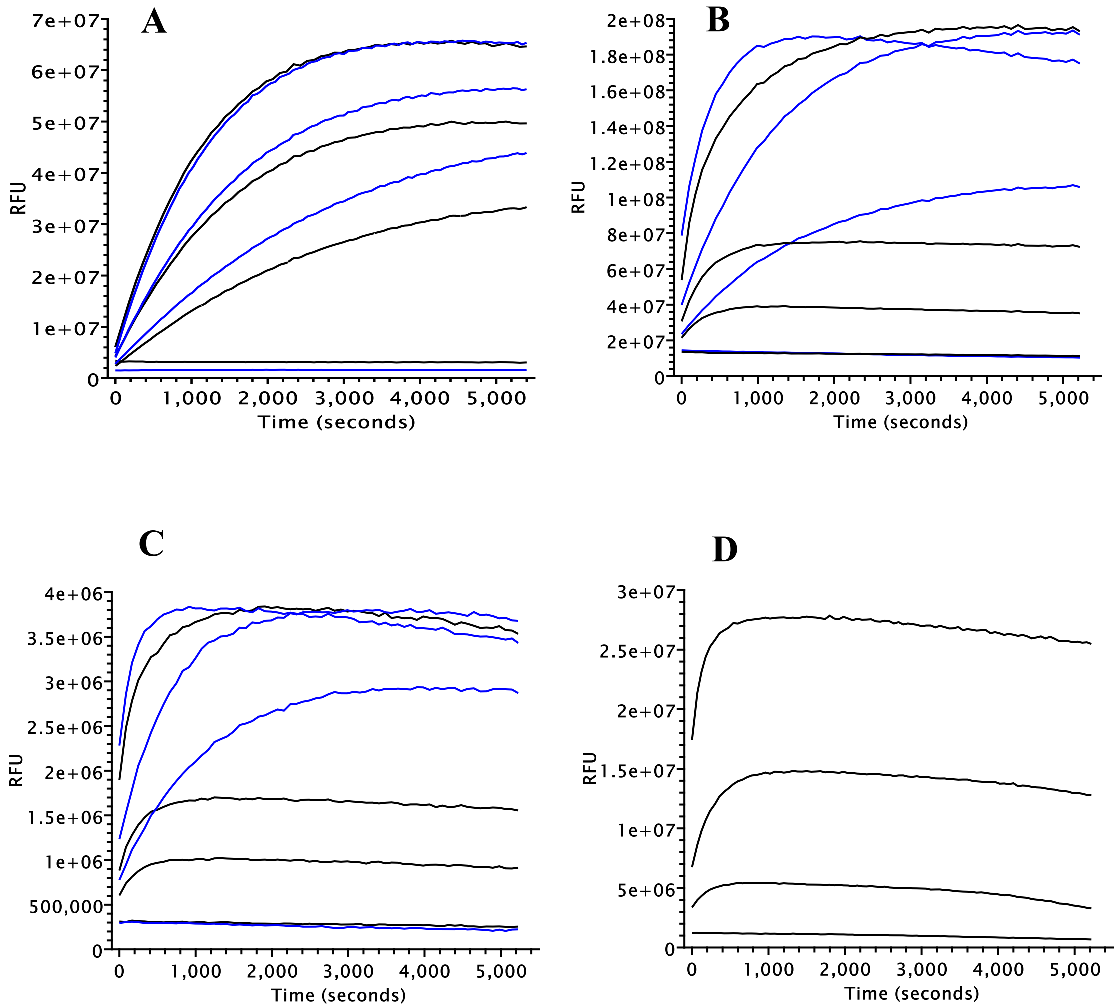


Figure 4.4. Reaction progress curve for the complete hydrolysis of substrate. 20 μM substrate in 20 mM BIS-TRIS pH 7.0 (black lines) and 20 mM BIS-TRIS pH 7.0, 2 mM EDTA, 2mM DTT (blue lines). Each condition was tested at four enzyme concentrations. *A*, For the VKLQ - AMC substrate 0, 2, 4 and 8 μM enzyme was used. *B-D*, For the nsp4-5-FAM, nsp4-5-EDANS, and nsp4-5-MCA substrates 0, 40, 80 and 160 nM enzyme was used. nsp4-5-MCA substrate was inactive in the presence of EDTA. Each curve represents the mean of 3 independent measurements.

Substrate steady-state kinetic parameters

Measurements were performed with each substrate to determine their utility for characterizing M^{pro}. k_{cat}/K_m was measured at low substrate concentrations where the initial rate increased linearly with substrate concentration. The results show the nsp4-5-MCA substrate had the highest k_{cat}/K_m, followed by the nsp4-5-FAM and nsp4-5-EDANS substrates, while the VKLQ - AMC substrate had the lowest k_{cat}/K_m value (Figure 4.5*A-D* and Table 4.2). The FRET substrates suffered from poor solubility and large inner filter effects when used at high concentrations needed to reach saturating substrate concentrations (V_{max}). These are commonly reported problems for FRET substrates in general, including FRET substrates used for SARS-CoV-2 M^{pro} (18, 28–30). As a result, a full Michaelis-Menten plot that reaches saturating substrate concentrations could only be measured using the VKLQ - AMC substrate (Figure 4.6 and Table 4.2).

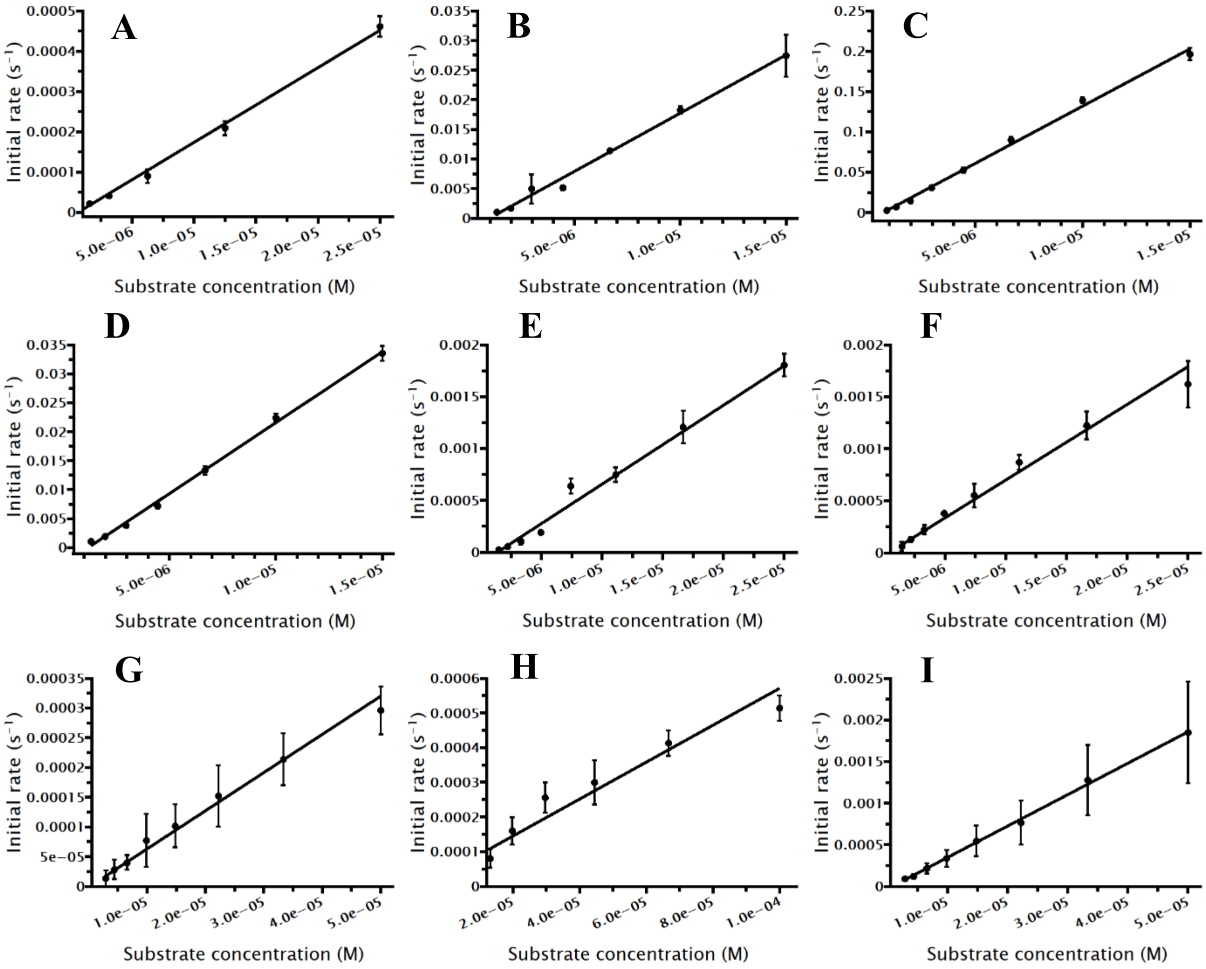


Figure 4.5. Measurement of K_{cat}/K_m at substrate concentrations below substrate K_m. Y axis shows the initial rate of product formation per second per enzyme active site in units of s^{-1}. *A*, VKLQ - AMC substrate. *B*, nsp4-5-EDANS substrate. *C*, nsp4-5-MCA substrate. *D*, nsp4-5-FAM substrate. *E*, nsp5-6-FAM substrate. *F*, nsp6-7-FAM substrate. *G*, nsp8-9-FAM substrate. *H*, nsp10-12-FAM substrate. *I*, nsp14-15-FAM substrate. Each data point is the mean ± 1 standard deviation, n = 3.

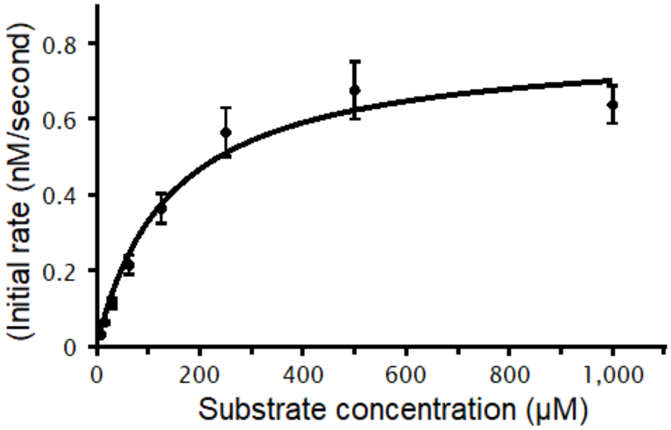


Figure 4.6. Michaelis-Menten plot for the VKLQ - AMC substrate. Values of 172±28 μM, 0.842±0.065 nM s^{-1} and 0.00421±0.00053 s^{-1} are measured for the K_m, V_{max} and k_{cat} respectively, k_{cat}/K_m was found to be 24.5±5.0 $M^{-1}s^{-1}$. Each data point is the mean with error bars showing ± 1 standard deviation, n = 3.

M^{pro} recognizes and cleaves 11 sites on the viral pp1a and pp1ab during viral replication. The nsp4-5 cleavage sequence at the N-terminus of M^{pro} is the sequence commonly used in M^{pro} FRET substrates and is the sequence used in each of the nsp4-5-MCA, nsp4-5-EDANS and nsp4-5-FAM substrates (Table 4.2). To test the kinetic efficiency of other SARS-CoV-2 polyprotein cleavage sequences, the k_{cat}/K_m of five additional FAM based FRET substrates utilizing the nps5-6, nsp6-7, nsp8-9, nsp10-12 and nsp14-15 cleavage sequences were tested and compared to the nsp4-5-FAM substrate (Table 4.2). The k_{cat}/K_m values for these substrates shows that nsp4-5 has by far the highest k_{cat}/K_m value followed nsp5-6, nsp6-7 and nsp14-15, while nsp8-9 and nsp10-12 cleavage sites have by far the lowest k_{cat}/K_m values (Figure 4.5*E-I* and Table 4.2). These results show that out of the substrates tested, the nps4-5 sequence has the most desirable kinetic properties for use in enzyme assays.

Table 4.2. Steady-state kinetic parameters for SARS-CoV-2 M^{pro} fluorescent substrates

Substrate	Sequence [a]	k_{cat}/K_m ($M^{-1}s^{-1}$)
AKLQ - AMC	*Ac*-VKLQ ↓ {AMC}	18.5 ± 1.0
		24.5 ± 5.0 [b]
nsp4-5-MCA	{MCA}-AVLQ ↓ SGFR{K(DNP)}K-NH_2	14,190 ± 420
nsp4-5-EDANS	{DABCYL}-KTSAVLQ ↓ SGFRKM{E(EDANS)}-NH_2	1960 ± 190
nsp4-5-FAM	{DABCYL}-KTSAVLQ ↓ SGFR{K(FAM)}K-NH_2	2448 ± 85
nsp5-6-FAM	{DABCYL}-KSGVTFQ ↓ SAVK{K(FAM)}K-NH_2	77.5 ± 4.2
nsp6-7-FAM	{DABCYL}-KKVATVQ ↓ SKMS{K(FAM)}K-NH_2	68 ± 10
nsp8-9-FAM	{DABCYL}-KSAVKLQ ↓ NNEL{K(FAM)}K-NH_2	6.01 ± 0.61
nsp10-12-FAM	{DABCYL}-KREPMLQ ↓ SADA{K(FAM)}K-NH_2	4.74 ± 0.48
nsp14-15-FAM	{DABCYL}-KTFTRLQ ↓ SLEN{K(FAM)}K-NH_2	38.0 ± 5.2

[a] Text in brackets denotes fluorophore or quencher, ↓ denotes cleavage site. [b] Kinetic parameters determined from Michaelis-Menten plot. Each value reported as the mean ± 1 standard deviation, n = 3.

Characterizing the improved nsp4-5-FAM substrate

To assess the suitability of each of the three FRET substrates for HTS, the Z'-factor for each substrate was determined. The Z'-factor is a statistical parameter used to evaluate the quality of a HTS(31). The Z'-factor reflects two critical criteria which a good quality HTS must have. The first criteria is the signal dynamic range, which describes the difference in signal produced by a positive and negative control. When the assay signal dynamic range is large, the signal produced by an active compound can more confidently be distinguished from an inactive compound. The second criteria reflected in the Z'-factor is the variability or standard deviation of the signal produced by the positive and negative controls. When the positive and negative controls produce a consistent signal with low variability, the signal produced by an active compound can be more confidently differentiated from signal variability. To assess the Z'-factor for a FRET substrate, the mean and standard deviation of the initial rate measured for 16 positive and 16 negative controls was calculated. The signal dynamic range and Z'-factor were calculated as described in the experimental procedures. Baicalein is a non-covalent inhibitor of SARS-CoV-2 (32) and was used as a positive control, the negative control contained DMSO instead of baicalein. This assay was repeated in triplicate for each FRET substrate, with results reported in Table 4.2. Of the FRET substrates tested in this work, the nsp4-5-EDANS substrate produces the smallest standard deviation in signal for the positive and negative controls, followed by the nsp4-5-FAM and nsp4-5-MCA substrates. The nsp4-5-FAM substrate produced the largest signal dynamic range followed by the nsp4-5-MCA and nsp4-5-EDANS substrates (Table 4.3). The large signal dynamic range produced by the nsp4-5-FAM substrate is largely attributed to the much higher brightness of the FAM fluorophore in comparison to the MCA and EDANS fluorophores (Figure S4.2). In this study, the nsp4-5-MCA and nsp4-5-EDANS substrates produced a Z'-factor of between 0.72 to 0.79, while the nsp4-5-FAM substrate produces a considerably higher Z'-factor of between 0.82 to 0.85 (Table 4.3).

Table 4.3. High throughput screening assay quality statistics for SARS-CoV-2 M^{pro} FRET substrates.

Substrate	Replicate	Signal mean (RFU/second)[a]		Signal standard deviation (RFU/second)[a]		Signal dynamic range (RFU/sec)	Z'-factor
		Positive control	Negative control	Positive control	Negative control		
nsp4-5-EDANS	1	287	6,182	141	315	5,895	0.768
	2	699	6,638	132	382	5,939	0.741
	3	551	6,022	179	297	5,471	0.739
nsp4-5-MCA	1	1,190	84,193	407	6,521	83,003	0.750
	2	2,389	98,870	646	6,115	96,482	0.790
	3	3,023	91,495	580	7,568	88,472	0.724
nsp4-5-FAM	1	5,004	162,872	575	8,923	157,868	0.820
	2	3,418	140,138	471	6,135	136,720	0.855
	3	4,194	156,189	1,086	7,013	151,955	0.840

[a] Signal mean and standard deviation, n = 16.

4.2.4 Discussion

The SARS-CoV-2 M^{pro} is a validated drug, and M^{pro} inhibitors have been shown to block viral replication in cell culture(4–8). Additionally, M^{pro} inhibitors could have broad spectrum antiviral activity against related coronaviruses because of the conserved features of the M^{pro} recognition sequence(4, 11). Florescent substrates are commonly used to study M^{pro} enzymatic activity, identify inhibitors through HTS, and test inhibitor efficacy. In this study we perform a biophysical characterization of SARS-CoV-2 M^{pro} and asses the steady-state kinetic parameters of three commonly used substrates, as well as 6 polyprotein cleavage sequences. We then develop the improved nsp4-5-FAM substrate that is better suited for HTS when compared to commonly used FRET substrates, resulting from the higher brightness of the FAM fluorophore. Additionally, the FAM fluorophore is less susceptible to interference and false positives due to the green-shifted absorption and emission spectra of the FAM fluorophore.

The protease used in this study was produced recombinantly in *E.coli* following a previously described method(24). M^{pro} produced by this method has been successfully used for structural and enzymatic studies (4, 7, 8, 18). The primary advantage of this method is that it generates M^{pro} with native N- and C-termini which is known to be structurally different, and more catalytically active than M^{pro} with non-native N- or C-termini(4, 24). In addition, conflicting M^{pro} enzymatic data has been published in the literature which has been in part attributed to the use of different M^{pro} constructs with non-native termini (20, 21, 23). Work on the closely related SARS-CoV M^{pro} has recommended the standard adoption of native termini M^{pro} for enzymatic and structural studies(30). For these reasons, native SARS-CoV-2 M^{pro} was used for the biophysical and enzymatic work done in this study.

Characterization of substrate kinetic parameters is critical for understanding the behaviour of both the substrate and enzyme, and can also help guide the development of a properly optimized enzyme assay. k_{cat}/K_m is an informative and useful parameter that gives a measure of substrate specificity and is the apparent second order rate constant for product formation. We found that the value of 14,190 ± 420 $M^{-1}s^{-1}$ for the nsp4-5-MCA substrate was 6 to 7 times higher than the value of 2448 ± 85 $M^{-1}s^{-1}$ and 1960 ± 190 $M^{-1}s^{-1}$ measured for the nsp4-5-FAM and nsp4-5-EDANS substrates respectively, which is consistent with values reported in the literature(4, 7). The k_{cat}/K_m value of 18.5 ± 1.0 $M^{-1}s^{-1}$ for the VKLQ - AMC substrate is far lower than the FRET substrates due to the shorter recognition sequence of the VKLQ - AMC substrate which lacks the residues C-terminal to the cleavage site(18).

The VKLQ - AMC substrate was the only substrate that could be used at concentrations needed reach V_{max}. By measuring the Michaelis-Menten kinetics of the VKLQ - AMC substrate, we report a k_{cat}/K_m of 24.5 ± 5.0 $M^{-1}s^{-1}$ which agrees with the value of 18.5 ± 1.0 $M^{-1}s^{-1}$ obtained using low substrate concentrations. This demonstrates that the behaviour of the VKLQ - AMC substrate is consistent at low and high concentrations, therefore the VKLQ - AMC substrate is relatively unaffected by the inner filter effect. Others have also found that these fluorogenic substrates are better suited for use at high concentrations than FRET substrates(18). A chromogenic substrate very similar to the fluorogenic VKLQ - AMC substrate was more useful for catalytic mechanism studies of SARS-CoV M^{pro} than the nsp4-5-EDANS FRET substrate(33).

Of the M^{pro} polyprotein cleavage sequences tested, we found that the nsp4-5 cleavage sequence has by far the highest k_{cat}/K_m and therefore is best suited for use in enzyme assays. This result is consistent with measurements done on the SARS-CoV M^{pro} which also show that the nsp4-5 cleavage sequence has the highest k_{cat}/K_m of the 11 polyprotein cleavage sequences (34). N-terminomics studies have identified the preferred cleavage sequence of SARS-CoV-2 M^{pro} to be A-X-L-Q↓(A/S)(9, 10). Of the cleavage sequences we tested, the nsp4-5 sequence is the only one which strictly represents this consensus sequence.

By characterizing the properties of the previously published nsp4-5-EDANS, nsp4-5-MCA, and VKLQ - AMC substrates, we recognized that an improved substrate for HTS could be developed.

Because the k_{cat}/K_m of the VKLQ - AMC substrate is low, high concentrations of substrate and enzyme are needed to generate a measurable fluorescent signal. This makes the VKLQ - AMC substrate undesirable for HTS due to the larger amounts of enzyme and substrate that would be consumed. We also recognized that both the nsp4-5-MCA and nsp4-5-EDANS substrates use fluorophores with relatively low brightness and undesirable excitation and emission spectra that makes them susceptible to interference from assay compounds(35). To develop an improved FRET substrate for HTS, we chose to use a FAM fluorophore because of its higher brightness, and spectral properties that are less prone to interference. Additionally, FAM is an inexpensive and readily available fluorophore that can easily be incorporated into the peptide substrate by most custom peptide synthesis companies.

To assess whether our new nsp4-5-FAM substrate is better for HTS than existing nsp4-5-MCA and nsp4-5-EDANS substrates, we characterized the Z'-factor for these substrates. An ideal assay produces a Z'-factor of 1, however an experimental assay could never achieve this value. A Z'-factor of greater than 0.5 is usually considered an excellent quality assay(31). We found that both the nsp4-5-MCA and nsp4-5-EDANS substrates produced a Z'-factor of 0.75, indicating our assay conditions are well optimized. However, because of the low brightness of the MCA and EDANS fluorophores, these substrates produce a low signal dynamic range which limits the Z'-factor. The most effective way to develop a substrate that is better suited for HTS is to use a brighter fluorophore. We found that by using the brighter FAM fluorophore, we were able to greatly increase the signal dynamic range of the assay and increase the Z'-factor of the assay to 0.84. This demonstrates that the improved nsp4-5-FAM substrate is better suited for HTS.

False positives are another common issue encountered in HTS assays and are especially problematic when screening large libraries of compounds(35). Fluorescent based assays are especially susceptible to false positives caused by compounds which interfere with the measured fluorescent signal. Additionally, the vast majority of compounds tested in HTS absorb and fluoresce at wavelengths in the ultraviolet and blue region of the spectrum (< 490 nm)(36). This makes the nsp4-5-MCA and nsp4-5-EDANS substrates (ex/em 320/405 nm, and 350/480 nm respectively) especially susceptible to interference and false positives. In contrast, the nsp4-5-FAM substrate absorbs and emits green light (ex/em 490/530 nm) and is therefore largely unaffected by this source of interference, reducing the potential for false positives(35).

The biophysical and enzymatic characterization of the native SARS-CoV-2 M^{pro} described in this work will serve as a valuable reference for future studies investigating the activity of SARS-CoV-2 M^{pro}. Using optimized assay conditions, we were able to compare properties of commonly used M^{pro} substrates and develop an improved nsp4-5-FAM substrate that is better suited for HTS. When compared to commonly used M^{pro} FRET substrates, this substrate generates a better-quality HTS assay because of the higher brightness of the FAM fluorophore, and is less susceptible to interference from assay compounds due to its green-shifted absorbance and emission spectra. This substrate will thus serve as a valuable tool in the development and design of future HTS assays aimed at identifying and characterizing novel direct-acting antivirals targeting the SARS-CoV-2 M^{pro}.

4.2.5 Methods

Construct design, enzyme expression and storage

Design of the expression vector followed previously reported methods(4, 24). The codon optimized SARS-CoV-2 M^{pro} open reading frame was inserted at the BamHI and XhoI restriction sites of a PGEX-6p-1 expression vector. The M^{pro} open reading frame contained the N-terminal autocleavage site AVLQ↓SGFRK (↓ denotes cleavage site) and a modified version of the C-terminal autocleavage site VTFQ↓GP followed by a His_6-tag. Auto-cleavage occurs during protein expression to produce a native N-terminus. The modified C-terminal autocleavage site is not cleaved by M^{pro} but can be cleaved by human rhinovirus 3C protease (HRV-3C) to produce the native M^{pro} C-terminus during protein purification. This SARS-CoV-2 M^{pro} expression vector was synthesized by GenScript.

The SARS-CoV-2 M^{pro} expression vector was transformed into *E.coli* strain BL21-Gold (DE3)(37). A single colony was used to inoculate a 50 mL culture of Miller's LB broth containing 100 µg/mL ampicillin overnight at 30^{o}C with shaking. 10 mL of overnight culture was used to inoculate 600 mL of LB broth containing 100 µg/mL ampicillin. This culture was grown at 37^{o}C until an OD$_{600}$ of around 0.6 was reached, then induced with 0.5 mM isopropyl-D-thiogalactoside (IPTG) for 14 hours at 20^{o}C. Cells were harvested at 4^{o}C by centrifugation 4000 x g for 20 minutes. The cell pellet was resuspended in a minimal volume of 20 mM TRIS pH 8.0, 300 mM NaCl and stored at -20^{o}C. Cells were thawed then lysed by sonication on ice. Lysate was clarified at 4^{o}C by centrifugation at 48 000 x g for 20 minutes. The supernatant was passed onto a 5 mL HisTrap HP (Cytiva) equilibrated with buffer A (20 mM TRIS pH 8.0, 20 mM Imidazole, 500 mM NaCl, 1 mM DTT). The column was washed with 25 mL of buffer A, then M^{pro} was eluted with a linear gradient from 0 to 100 % buffer B (20 mM TRIS pH 8.0, 500 mM NaCl, 500 mM NaCl, 1 mM DTT) over 75 mL, 4 mL fractions were collected. Fractions were analysed by SDS-PAGE, those containing M^{pro} were pooled and mixed with HRV-3C protease (Sigma-Aldrich) in a 40:1 ratio and dialyzed into 50 mM TRIS pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT overnight at room temperature. Next the protein was exchanged into buffer C (50 mM TRIS pH 8.0, 1 mM DTT) and concentrated to 10 mg/mL using a 10 kDa nominal molecular weight centrifugal filter. The protein was loaded onto a Mono Q 4.6/100 PE anion exchange column (cytiva) pre-equilibrated with buffer C. The column was washed with 20 mL of buffer C, then M^{pro} was eluted with a linear gradient from 0 to 30% buffer D (50 mM TRIS pH 8.0, 500 mM NaCl, 1 mM DTT) over 20 mL, 0.5 mL fractions were collected. Fractions were analysed by SDS-PAGE, those containing pure M^{pro} free of detectable contamination were pooled and buffer exchanged into 20 mM TRIS pH 8.0, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP) and concentrated using a centrifugal filter.

Protein concentration was measured spectrophotometrically using an extinction coefficient of 32890 $M^{-1}cm^{-1}$ and a molecular weight of 33796 Da, calculated by ProtParam(38). M^{pro} was stored at a concentration of 45.3 µM in 50 % v/v ethylene glycol at -80°C for long term storage and -20°C for short terms storage. There was no substantial loss of enzyme activity under these storage conditions.

Differential Scanning Fluormetry (DSF) and Dynamic Light Scattering (DLS)

Thermodynamic stability and particle size distribution of M^{pro} was measured using a Prometheus Panta (NanoTemper Technologies GmbH). M^{pro} at a concentration of 1.1 mg/mL in 20 mM TRIS pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT was filtered through a 0.1 µm Ultrafree centrifugal filter (Merck Millipore). Standard Prometheus capillaries (PR-C002) were used. DSF measurements were done from 25 to 80°C with a temperature gradient of 1°C/minute, intrinsic tryptophan fluorescence was measured at 350 and 330 nm. For DLS measurements 10 acquisitions of 5000 ms each were done at a temperature of 20°C with 100 % DLS power. Buffer viscosity was calculated to be 1.02139 mPa•s using the buffer builder incorporated in the Panta Control software v1.2.1. The autocorrelation function was fit to a size distribution analysis model. Data analysis was done with the Panta Analysis software v1.2.

Circular dichroism (CD) spectroscopy

CD measurements of M^{pro} were performed at a concentration of 0.5 mg/mL in 10mM Na_2HPO_4, pH 8.0. CD measurements were taken with a Jasco J-810 (JASCO corporation) spectropolarimeter at 20°C in a 0.05 cm path length quartz cuvette. Raw data were converted to mean residue ellipticity and secondary structure deconvolution was done using the CDSSTR algorithm and the SMP180 reference set on the DichroWeb server(39, 40). Experimental secondary structure fractions were compared to the protease crystal structure (PDB ID: 6Y2E) using PDBsum(41).

Fluorescent substrates

Amino acid sequences of the substrates used in this study can be found in Table 1 and Figure 1. The nsp4-5-MCA substrate was purchased from CanPeptide Inc, all other substrates were purchased from GenScript. All substrates had a purity greater than 95% confirmed by HPLC and the molecular weight confirmed by mass spectrometry (testing done by supplier). The excitation and emission wavelengths used for each substrate are as follows: VKLQ - AMC 360/460 nm ex/em; nsp4-5-EDANS 350/480 nm ex/em; nsp4-5-MCA 320/405 nm ex/em; FAM based substrates 490/530 nm ex/em. All substrates came lyophilized as the trifluoroacetic acid salt, stock solutions were prepared in DMSO and stored protected form light at -20°C.

Enzyme assay general methods

All measurements were taken on a SpectraMax iD5 microplate reader controlled by Softmax pro 7.1 software (Molecular Devices). All reading were done in a black 96 well flat-bottom polypropylene microplate (Greiner Bio-One, ref 655209). Readings were taken every 20 seconds for 600 seconds to measure initial reaction rates, and up to 1.5 hours to measure complete hydrolysis. Measurements were done at ambient temperature. Initial rates were fit to the linear portion of the reaction progress, usually the first 200 seconds corresponding to less than 10% substrate hydrolysis. Fluorescence units were converted to concentration using a standard curve generated using a fluorophore standard in 20mM BIS-TRIS pH 7.0. 7-Methoxycoumarin-4-acetic acid (MCA) and 5-(2-Aminoethylamino)-1-naphthalenesulfonic acid (EDANS) were purchased from Fisher Scientific, 5-carboxyfluorescein (FAM) was purchased from Cayman Chemical Company and 7-amino-4-methylcoumarin (AMC) was purchased from Sigma-Aldrich. All parameter fitting by linear and non-linear regression was done in QtiPlot. All measurements were performed in triplicate, final values are expressed as the mean ± 1 standard deviation of the three measurements.

Steady-state enzyme kinetics

k_{cat}/K_m measurements were done in 20 mM BIS-TRIS pH 7.0 with 100uL well volume. For the nsp4-5-FAM, nsp4-5-EDANS and nsp4-5-MCA substrates 80nM enzyme was used with substrate ranging from 15 to 1.3μM for nsp4-5-FAM and nsp4-5-EDANS or 15 to 0.88 μM for nsp4-5-MCA. For the VKLQ - AMC, nsp5-6-FAM and nsp6-7-FAM substrates 200 nM enzyme was used, with substrate concentrations ranging from 25 to 0.78 μM for the VKLQ - AMC substrate and 25 to 1.46 μM for the nsp5-6-FAM and nsp6-7-FAM substrates. For the nsp8-9-FAM and nsp14-15-FAM substrates, 200 nM enzyme was used with substrate concentrations from 50 to 2.93 μM. For the nsp14-15-FAM substrate, the baseline rate of substrate hydrolysis in the absence of enzyme was subtracted from the rate of hydrolysis with enzyme. For the nsp10-12-FAM substrate, 400 nM enzyme was used with substrate concentrations ranging from 100 to 13.2 μM. Initial rate measured in RFU/second was converted to M/second. This rate was divided by the molar concentration of enzyme used in the assay, to give the rate of product formation per second per enzyme active site in units of s^{-1}. A plot of this rate against molar substrate concentration was made, the slope of this plot gives the value of k_{cat}/K_m. To measure the full Michaelis-Menten plot for the VKLQ - AMC substrate 100 μL of 200nM enzyme and between 1000 to 7.8 μM substrate in 20 mM BIS-TRIS pH 7.0 was used. Initial rate in RFU/second was converted to M/second. A plot of reaction rate in M/second vs the molar substrate concentration was fit to the Michaelis-Menten equation to obtain values of K_m and V_{max}. To calculate k_{cat}, V_{max} was divided by the molar concentration of enzyme used in the assay. With these values of k_{cat} and K_m the value of k_{cat}/K_m can be calculated independently from the method described above.

High throughput screening assessment

The Z'-factor was assessed by measuring enzyme activity of 16 positive and 16 negative controls, and was repeated in triplicate for each FRET substrate. Baicalein (CAS # 491-67-8, Sigma-Aldrich) a non-covalent inhibitor of SARS-CoV-2 M^{pro} was used as a positive control, the negative control contained DMSO instead of baicalein. The reaction contained 100µL of 10µM substrate, 40nM enzyme and either 50µM baicalein or DMSO as the positive and negative controls respectively. The final buffer composition was 20 mM BIS-TRIS pH 7.0, 1.2 % v/v DSMO. For each assay the mean and standard deviation of the initial rate for positive and negative controls was calculated. The signal dynamic range was calculated according to the following, where $\bar{\mu}_n$ and $\bar{\mu}_p$ are the mean of the negative and positive controls respectively.

$$Signal\ dynamic\ range\ = \bar{\mu}_n - \bar{\mu}_p$$

The Z'-factor was calculated according to Zhang et al.(31) where σ_p and σ_n are the standard deviation of the positive and negative controls respectively.

$$Z' = 1 - \frac{(3\sigma_p + 3\sigma_n)}{|\bar{\mu}_p - \bar{\mu}_n|}$$

4.2.6 Supplemental Information

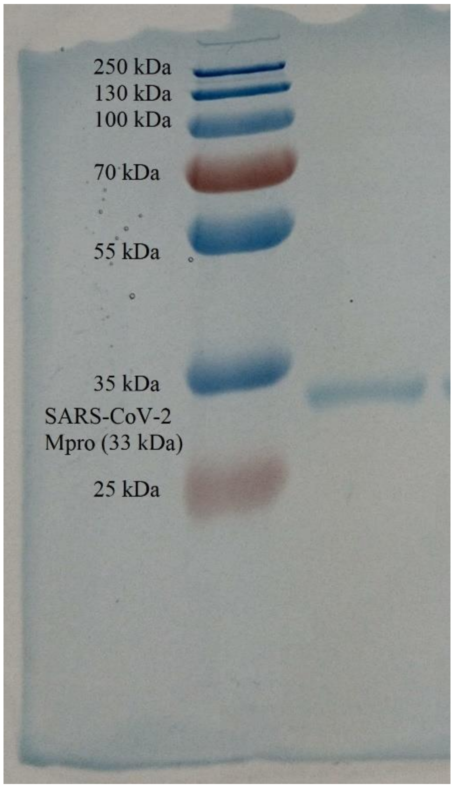


Figure S4.1. SDS-PAGE of SARS-CoV-2 M^{pro} used for biophysical and enzymatic studies.

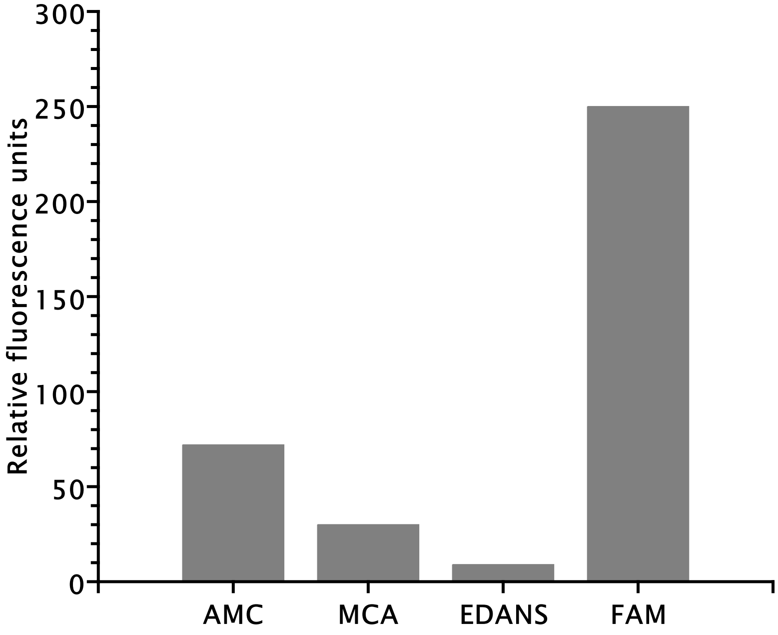


Figure S4.2. Relative fluorescence units produced by 100 µL of 2 µM fluorophore in 20mM BIS-TRIS pH 7.0. This shows the FAM fluorophore is much brighter in comparison to other fluorophores used in the fluorescent substrates. Values are reported as the mean of 3 measurements.

4.2.7 References

1. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol*. **19**, 155–170
2. Dong, E., Du, H., and Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534
3. CDCMMWR (2021) COVID-19 Vaccine Breakthrough Infections Reported to CDC — United States, January 1–April 30, 2021. *MMWR Morb. Mortal. Wkly. Rep*. 10.15585/mmwr.mm7021e3
4. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*. **368**, 409–412
5. Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M. T., Chen, Y., and Wang, J. (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Research*. **30**, 678–692
6. Vuong, W., Khan, M. B., Fischer, C., Arutyunova, E., Lamer, T., Shields, J., Saffran, H. A., McKay, R. T., van Belkum, M. J., Joyce, M. A., Young, H. S., Tyrrell, D. L., Vederas, J. C., and Lemieux, M. J. (2020) Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun*. **11**, 4282
7. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., and Yang, H. (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*. **582**, 289–293
8. Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L.-K., Xu, Y., Yang, H., and Liu, H. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science*. **368**, 1331–1335
9. Pablos, I., Machado, Y., de Jesus, H. C. R., Mohamud, Y., Kappelhoff, R., Lindskog, C., Vlok, M., Bell, P. A., Butler, G. S., Grin, P. M., Cao, Q. T., Nguyen, J. P., Solis, N., Abbina, S., Rut, W., Vederas, J. C., Szekely, L., Szakos, A., Drag, M., Kizhakkedathu, J. N., Mossman, K., Hirota, J. A., Jan, E., Luo, H., Banerjee, A., and Overall, C. M. (2021) Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CLpro substrate degradome. *Cell Reports*. **37**, 109892
10. Koudelka, T., Boger, J., Henkel, A., Schönherr, R., Krantz, S., Fuchs, S., Rodríguez, E., Redecke, L., and Tholey, A. (2021) N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *PROTEOMICS*. **21**, 2000246
11. Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., von Brunn, A., Leyssen, P., Lanko, K., Neyts, J., de Wilde, A., Snijder, E. J., Liu, H., and Hilgenfeld, R. (2020) α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **63**, 4562–4578
12. Coelho, C., Gallo, G., Campos, C. B., Hardy, L., and Würtele, M. (2020) Biochemical screening for SARS-CoV-2 main protease inhibitors. *PLOS ONE*. **15**, e0240079
13. Jo, S., Kim, S., Kim, D. Y., Kim, M.-S., and Shin, D. H. (2020) Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro. *J. Enzyme Inhib. Med. Chem.* **35**, 1539–1544

14. Breidenbach, J., Lemke, C., Pillaiyar, T., Schäkel, L., Hamwi, G. A., Diett, M., Gedschold, R., Geiger, N., Lopez, V., Mirza, S., Namasivayam, V., Schiedel, A. C., Sylvester, K., Thimm, D., Vielmuth, C., Vu, L. P., Zyulina, M., Bodem, J., Gütschow, M., and Müller, C. E. (2021) Targeting the Main Protease of SARS-CoV-2: From the Establishment of High Throughput Screening to the Design of Tailored Inhibitors. *Angew. Chem. Int. Ed*. **60**, 10423–10429
15. Baker, J. D., Uhrich, R. L., Kraemer, G. C., Love, J. E., and Kraemer, B. C. (2021) A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. *PLOS ONE*. **16**, e0245962
16. Kuzikov, M., Costanzi, E., Reinshagen, J., Esposito, F., Vangeel, L., Wolf, M., Ellinger, B., Claussen, C., Geisslinger, G., Corona, A., Iaconis, D., Talarico, C., Manelfi, C., Cannalire, R., Rossetti, G., Gossen, J., Albani, S., Musiani, F., Herzog, K., Ye, Y., Giabbai, B., Demitri, N., Jochmans, D., Jonghe, S. D., Rymenants, J., Summa, V., Tramontano, E., Beccari, A. R., Leyssen, P., Storici, P., Neyts, J., Gribbon, P., and Zaliani, A. (2021) Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen. *ACS Pharmacol. Transl. Sci.* **4**, 1096–1110
17. Zhu, W., Xu, M., Chen, C. Z., Guo, H., Shen, M., Hu, X., Shinn, P., Klumpp-Thomas, C., Michael, S. G., and Zheng, W. (2020) Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative High-Throughput Screening. *ACS Pharmacol. Transl. Sci.* **3**, 1008–1016
18. Rut, W., Groborz, K., Zhang, L., Sun, X., Zmudzinski, M., Pawlik, B., Wang, X., Jochmans, D., Neyts, J., Młynarski, W., Hilgenfeld, R., and Drag, M. (2021) SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol*. **17**, 222–228
19. Dražić, T., Kühl, N., Leuthold, M. M., Behnam, M. A. M., and Klein, C. D. (2021) Efficiency Improvements and Discovery of New Substrates for a SARS-CoV-2 Main Protease FRET Assay. SLAS *Discov*. **26**, 1189–1199
20. Li, Z., Li, X., Huang, Y.-Y., Wu, Y., Liu, R., Zhou, L., Lin, Y., Wu, D., Zhang, L., Liu, H., Xu, X., Yu, K., Zhang, Y., Cui, J., Zhan, C.-G., Wang, X., and Luo, H.-B. (2020) Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc. Natl. Acad. Sci. USA*. **117**, 27381–27387
21. Ma, C., and Wang, J. (2021) Dipyridamole, chloroquine, montelukast sodium, candesartan, oxytetracycline, and atazanavir are not SARS-CoV-2 main protease inhibitors. *Proc. Natl. Acad. Sci. USA*. **118**, e2024420118
22. Hattori, S., Higashi-Kuwata, N., Hayashi, H., Allu, S. R., Raghavaiah, J., Bulut, H., Das, D., Anson, B. J., Lendy, E. K., Takamatsu, Y., Takamune, N., Kishimoto, N., Murayama, K., Hasegawa, K., Li, M., Davis, D. A., Kodama, E. N., Yarchoan, R., Wlodawer, A., Misumi, S., Mesecar, A. D., Ghosh, A. K., and Mitsuya, H. (2021) A small molecule compound with an indole moiety inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun*. **12**, 688
23. Li, Z., Lin, Y., Huang, Y.-Y., Liu, R., Zhan, C.-G., Wang, X., and Luo, H.-B. (2021) Reply to Ma and Wang: Reliability of various in vitro activity assays on SARS-CoV-2 main protease inhibitors. *Proc. Natl. Acad. Sci. USA*. **118**, e2024937118
24. Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., Bartlam, M., and Rao, Z. (2007) Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *J. Mol. Bio.* **366**, 965–975
25. Abian, O., Ortega-Alarcon, D., Jimenez-Alesanco, A., Ceballos-Laita, L., Vega, S., Reyburn, H. T., Rizzuti, B., and Velazquez-Campoy, A. (2020) Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. *Int. J. Biol. Macromol.* **164**, 1693–1703

26. Ferreira, J. C., and Rabeh, W. M. (2020) Biochemical and biophysical characterization of the main protease, 3-chymotrypsin-like protease (3CLpro) from the novel coronavirus SARS-CoV 2. *Sci. Rep*. **10**, 22200
27. Copeland, R. A. (2000) *Enzymes: A Practical Introduction to Structure Mechanism and Data Analysis*, 2nd Ed., John Wiley & Sons Inc, New York, NY
28. Ferreira, J. C., Fadl, S., Ilter, M., Pekel, H., Rezgui, R., Sensoy, O., and Rabeh, W. M. (2021) Dimethyl sulfoxide reduces the stability but enhances catalytic activity of the main SARS-CoV-2 protease 3CLpro. *The FASEB Journal*. **35**, e21774
29. Liu, Y., Kati, W., Chen, C.-M., Tripathi, R., Molla, A., and Kohlbrenner, W. (1999) Use of a Fluorescence Plate Reader for Measuring Kinetic Parameters with Inner Filter Effect Correction. *Anal. Biochem*. **267**, 331–335
30. Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S. C., and Mesecar, A. D. (2008) Evaluating the 3C-like protease activity of SARS-Coronavirus: Recommendations for standardized assays for drug discovery. *Virus Res.* **133**, 63–73
31. Zhang, J. H., Chung, T. D. Y., Oldenburg, K. R. (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* **4**, 67-73
32. Su, H., Yao, S., Zhao, W., Zhang, Y., Liu, J., Shao, Q., Wang, Q., Li, M., Xie, H., Shang, W., Ke, C., Feng, L., Jiang, X., Shen, J., Xiao, G., Jiang, H., Zhang, L., Ye, Y., and Xu, Y. (2021) Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease. *Nat. Commun*. **12**, 3623
33. Solowiej, J., Thomson, J. A., Ryan, K., Luo, C., He, M., Lou, J., and Murray, B. W. (2008) Steady-State and Pre-Steady-State Kinetic Evaluation of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CLpro Cysteine Protease: Development of an Ion-Pair Model for Catalysis. *Biochemistry*. **47**, 2617–2630
34. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J., and Lai, L. (2004) Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J. Biol. Chem.* **279**, 1637–1642
35. Simeonov, A., Jadhav, A., Thomas, C. J., Wang, Y., Huang, R., Southall, N. T., Shinn, P., Smith, J., Austin, C. P., Auld, D. S., and Inglese, J. (2008) Fluorescence Spectroscopic Profiling of Compound Libraries. *J. Med. Chem.* **51**, 2363–2371
36. Simeonov, A., and Davis, M. I. (2004) *Assay Guidance Manual*, Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda, MD
37. Inoue, H., Nojima, H., and Okayama, H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene*. **96**, 23–28
38. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) *The Proteomics Protocols Handbook*, Humana Press, Totowa, NJ
39. Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers*. **89**, 392–400
40. Abdul-Gader, A., Miles, A. J., and Wallace, B. A. (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics*. **27**, 1630–1636
41. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., and Thornton, J. M. (2018) PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134

Chapter 5

Characterization of protein-ligand interactions for rational inhibitor design

Legare, S., Peters, O., Gupta, M., Rafiei, F., Padilla-Meier, G.P., Koch, M., Cabrele, C., Stetefeld, J. (2024). Targeting the heparin co-receptor for the disruption of Netrin-1 UNC5B interaction.

Contribution statement

Work presented in this chapter was performed in collaboration with the research group of Dr. Chiara Cabrele from Paris Lodron University of Salzburg. The research group of Dr. Cabrele designed and synthesized the peptide library used for initial inhibitor testing. Dr. Joerg Stetefeld and Scott Legare were involved in the peptide design process.

Gay Pauline Padilla-Meier, Manuel Koch, Oliver Peters and Scott Legare produced and purified proteins used in these studies. Monika Gupta, Faride Rafiei and Scott Legare designed and performed binding experiments to characterize the Netrin-1 UNC5B binding interaction and the requirement for heparin. Scott Legare designed, validated, and performed peptide IC_{50} assays. Scott Legare analyzed the results of these binding and IC_{50} assays. ITC assays characterizing the binding of peptides to heparin were designed, performed, and analyzed by Oliver Peters and Scott Legare. Scott Legare prepared figures and wrote the manuscript.

5.1

Consolidation of system characterization and ligand development for exploring protein – ligand interactions

Chapter 3 demonstrates the use of a multidisciplinary approach to develop a detailed understanding of the molecular basis and functional properties of protein – ligand interactions. Chapter 4 demonstrates the development of biophysical assay used to identify inhibitors of protein – ligand interactions for use in therapeutics or biotechnology applications. Together these chapters demonstrate the use of techniques and an experimental approach for understanding protein – ligand interactions and identifying valuable inhibitors of these interactions. This chapter demonstrates the combined application of techniques and experimental approaches used in chapters 3 and 4 for the rational development of inhibitors of a protein – ligand interaction.

In this chapter a detailed biophysical analysis of the Netrin-1 UNC5B interaction is performed and used to demonstrate that this protein – protein interaction requires a heparin co-receptor. This finding substantially expands the understanding of the Netrin-1 UNC5B interaction and demonstrates the need for consideration of heparins when studying this cell signaling pathway in a biological system. This newfound understanding is then combined with the existing structural understanding of the Netrin-1 UNC5B and Netrin-1 heparin interactions, in order to rationally design peptide inhibitors of the Netrin-1 UNC5B interaction.

5.2

Targeting the heparin co-receptor for the disruption of Netrin-1 UNC5B interaction

5.2.1 Introduction

Netrin-1 is secreted laminin-related glycoprotein first identified for its role in axon outgrowth and guidance during development of the nervous system (1, 2). Netrin-1 is also involved in processes such as leukocyte migration and inflammation, angiogenesis, cell survival and tumorigenesis (3–8). Netrin-1 interacts with a number of cell surface receptors including Deleted in Colon Cancer (DCC), Neogenin, uncoordinated-5 (UNC5A, UNC5B, UNC5C, UNC5D) and heparan sulfate proteoglycans (9–11). DCC and UNC5 are dependance receptors that are suggested to be viable drug targets for anticancer therapies (12). Upregulation of the Netrin-1 ligand in cancers is known to occur and has been shown to increase tumor cell survival, progression, and metastasis (7, 13, 14). Recently, an anti-Netrin-1 antibody blocking of the Netrin-1 – UNC5B interaction has shown success in a phase I clinical trial for use in endometrial cancer treatment and has demonstrated promise for melanoma treatment by blocking the Netrin-1 – DCC interaction (15, 16). This demonstrates that modulation of the interactions between Netrin-1 and its dependance receptors has great scientific and therapeutic value.

In addition to its receptors DCC and UNC5, Netrin-1 is known to bind sulfated glycosaminoglycans like heparin sulfate and heparan sulfate proteoglycans (11). These glycosaminoglycans often play a role in extracellular matrix cell signaling processes by serving as co-receptors for ligand-receptor binding, inducing ligand or receptor multimerization, and facilitating ligand immobilization and ligand gradient formation in many biological processes

(17). Because of their role in cell signaling processes, glycosaminoglycans have been explored as therapeutic targets and several strategies to modulate these glycosaminoglycan mediated processes have been devised (18, 19). The use of heparin binding peptides capable of blocking protein – heparin interactions have been developed for blocking of viral cell entry and infectivity, reversing the effects of heparin anticoagulants, reducing neutrophile adhesion and migration during inflammation and disrupting growth factor signaling (20–23).

In this work we demonstrate that the Netrin-1 UNC5B interaction requires a heparin co-receptor and that high affinity Netrin-1 UNC5B binding cannot occur in the absence of this heparin co-receptor. We then develop synthetic peptide fragments designed to mimic the structure of the Netrin-1 LE2 domain involved in heparin and UNC5B binding. We demonstrate these peptides can block the Netrin-1 UNC5B interaction in vitro. We further demonstrate that the inhibitory mechanism of these peptides involves binding of the heparin co-receptor in order to block the Netrin-1 UNC5B interaction. This study advances the biochemical understanding of the Netrin-1 UNC5B interaction and demonstrates the effective use of heparin binding peptides for inhibiting this heparin mediated protein-protein interaction.

5.2.2 Results

Netrin-1 - UNC5B - heparin binding

Netrin-1 and UNC5B proteins were prepared by methods previously used to produce high quality proteins suitable for biophysical and crystallography studies (11, 24, 25). Netrin-1 purification includes a chromatography step for the removal of bound heparin (11). Biolayer interferometry (BLI) was used to measure Netrin-1 UNC5B binding. In this assay the binding of Netrin-1 in solution to the UNC5B immobilized on the biosensor could not be detected at a Netrin-1 concentration of 1.8 μM (Figure 5.1A). Given that Netrin-1 is a known heparin binding protein, and a specific purification step was performed to remove heparin from the recombinantly produced Netrin-1, we reasoned that heparin may be necessary for the Netrin-1 UNC5B interaction to occur. Therefore, we then tested the binding between Netrin-1 and UNC5B using Netrin-1 that had not undergone the purification step required to remove co-purified heparin. We found that Netrin-1 prepared in this manner is capable of binding UNC5B (Figure 5.1B). This Netrin-1 UNC5B binding interaction clearly shows biphasic binding behavior in the association and dissociation phases, which was fit to a heterogeneous ligand binding model with a K_{D1} of 176 nM and a K_{D2} of 303 nM. Given that heparin is a naturally heterogeneous molecule with variable sulfation patterns we rationalize that two different types of heparin molecules could be present giving rise to the heterogeneous binding. To verify the hypothesis that the binding in Figure 5.1B is due to the presence of co-purified heparin in the Netrin-1 preparation and that the lack of Netrin-1 UNC5B binding shown in Figure 5.1A is the result of the removal of this heparin, we added commercially available porcine intestinal mucosa heparin to the Netrin-1 preparation used in Figure 5.1A. Addition of heparin to the assay buffer at a concentration of 500ng/mL enables this preparation of Netrin-1 to bind UNC5B with a K_D of 4.6 nM (Figure 5.1C). These results clearly demonstrate that heparin co-purified with Netrin-1, enables Netrin-1 to bind UNC5B, while removal of this co-purified heparin results in loss of binding ability which can be reversed by adding external heparin.

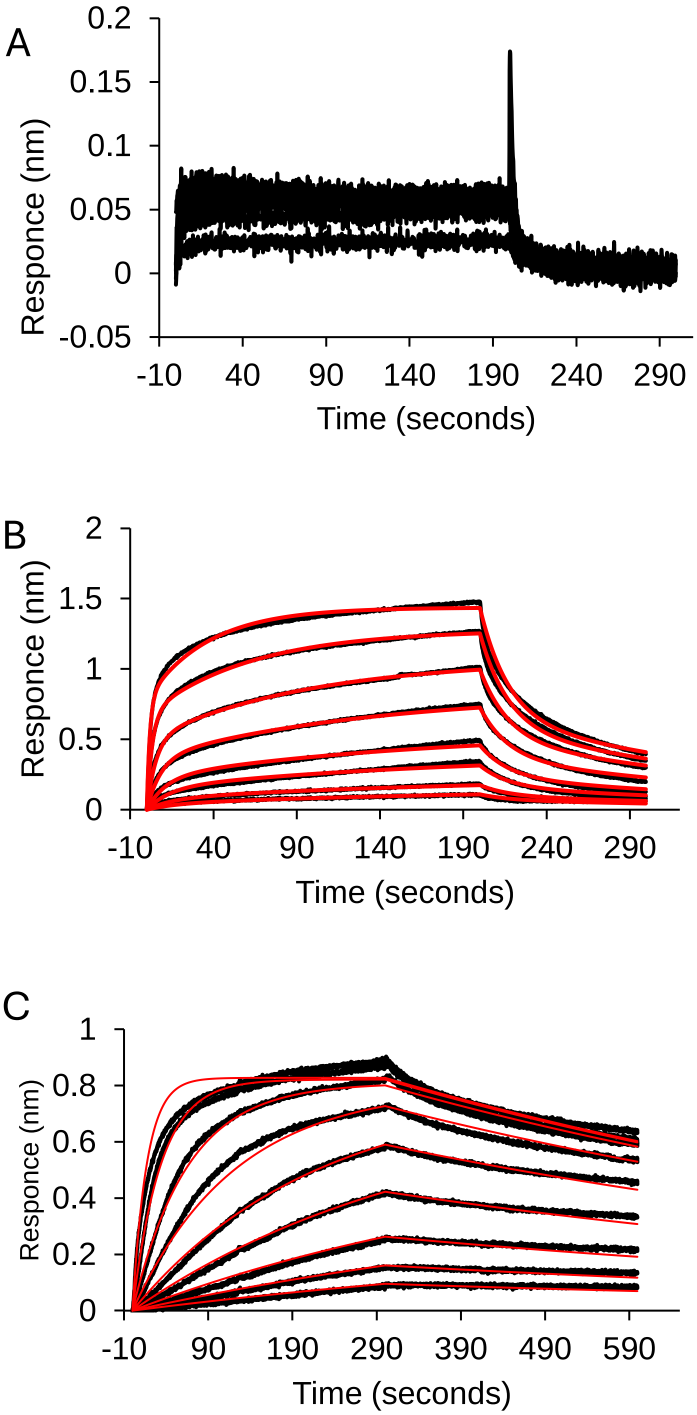


Figure 5.1. Biolayer interferometry binding assays detecting the binding of Netrin-1 in solution to surface immobilized UNC5B. A) Absence of measurable binding interaction between Netrin-1 and UNC5B when Netrin-1 is prepared to remove co-purified heparin. B) Binding between

Netrin-1 and UNC5B when co-purified heparin has not been removed from Netrin-1. Binding is fit to a heterogeneous ligand model with a K_{D1} of 176 nM, K_{D2} of 303 nM, k_{a1} of 3.16 x 10^5 $M^{-1}s^{-1}$, k_{a2} of 1.56 x 10^4 $M^{-1}s^{-1}$, k_{d1} of 5.60 x 10^{-3} s^{-1}, k_{d2} of 4.71 x 10^{-4} s^{-1}. C) Binding between Netrin-1 and UNC5B in the presence of 500 ng/mL exogenous porcine intestinal mucosa heparin. K_D of 4.6 nM, k_a 2.3 x 10^5 $M^{-1}s^{-1}$, and k_d of 1.1 x 10^{-3} s^{-1}.

We next explored Netrin-1 heparin binding affinity and kinetics using BLI, where heparin is immobilized on the biosensor surface and the binding of Netrin-1 in solution is detected. We found that Netrin-1 binds to heparin with an affinity of around 1.1 nM (Figure 5.2A). Comparing the Netrin-1 UNC5B binding in the presence of heparin (Figure 5.1C) to the Netrin-1 – heparin binding (Figure 5.2A), it is clear that the Netrin-1 heparin binding interaction is higher affinity with a faster association rate constant and similar dissociation rate constant. The binding of UNC5B to heparin was also tested and found to be very weak and non-saturating with an estimated affinity of around 11 μM (Figure 5.2B). This very weak interaction between heparin and UNC5B suggest that the high affinity heparin mediated interaction between Netrin-1 and UNC5B is largely stabilized by protein – protein contacts between Netrin-1 and UNC5B rather than protein – heparin contacts between UNC5B and heparin bound to Netrin-1.

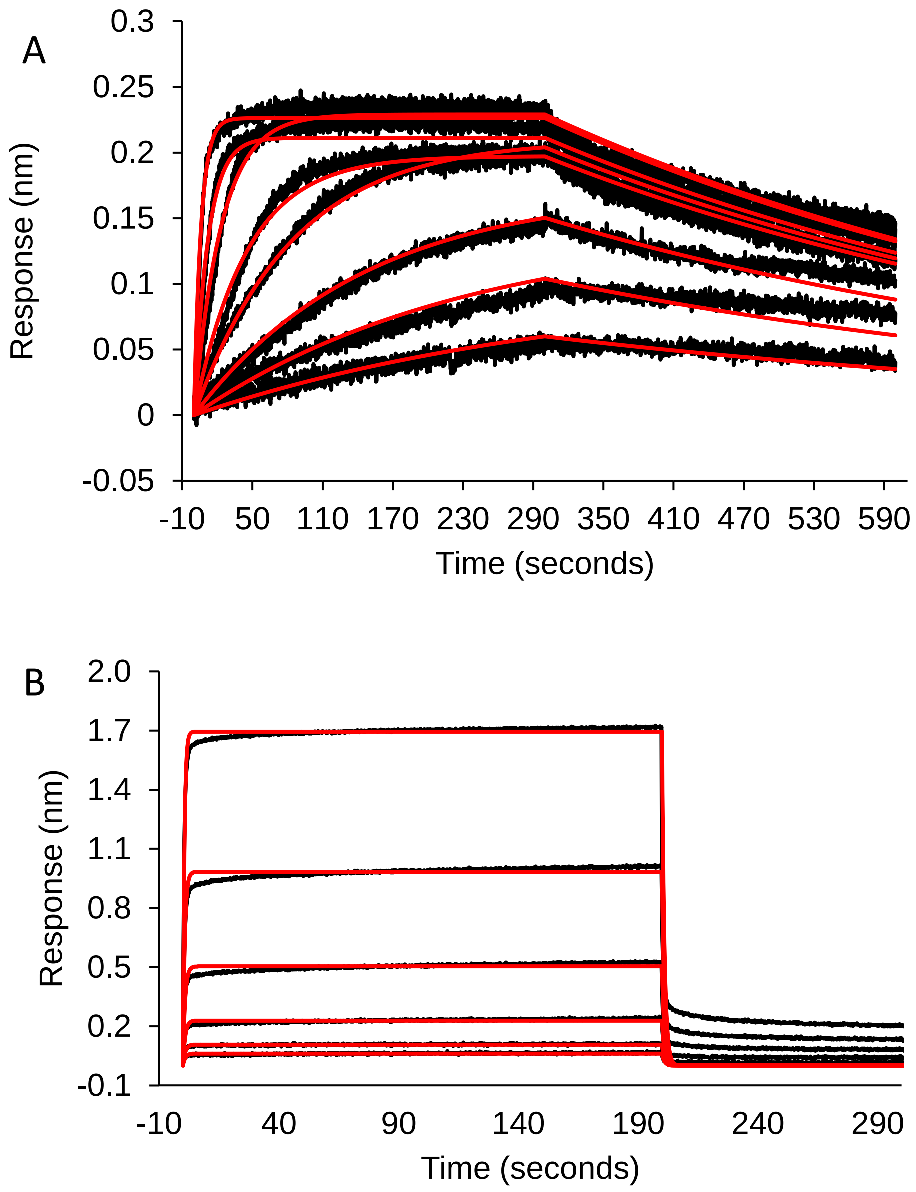


Figure 5.2. Biolayer interferometry binding assays detecting the binding of Netrin-1 and UNC5B in solution to surface immobilized porcine intestinal mucosa heparin. A) Netrin-1 binding with porcine intestinal mucosa heparin. K_D of 1.1 nM, k_a of 1.6 x 10^6 $M^{-1}s^{-1}$, and k_d of 1.8 x 10^{-3} s^{-1}. B) UNC5B binding with porcine intestinal mucosa heparin. Estimated K_D of 11 μM, k_a of 9.0 x 10^4 $M^{-1}s^{-1}$, k_d of 1.0 s^{-1}.

From the knowledge that Netrin-1 and UNC5B binding requires heparin and that Netrin-1 binds heparin with much higher affinity than the interaction between UNC5B and heparin, we can propose the following as a possible reaction mechanism.

$$\text{Netrin} + \text{Heparin} \underset{k_{d1}}{\overset{k_{a1}}{\rightleftharpoons}} \text{Netrin-Heparin} + \text{UNC5B} \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} \text{Netrin-Heparin-UNC5B} \qquad \text{Eq. 1}$$

To further explore the effects of heparin on the Netrin-1 UNC5B binding interaction affinity and kinetics, we measured the binding between Netrin-1 and UNC5B at different heparin concentrations. We first measured the binding rate of 250 nM Netrin-1 to UNC5B immobilized on the biosensor with varying concentrations of heparin in the buffer from 19.5 ng/mL to 5000 ng/mL (Figure 5.3A). This experiment shows that the heparin concentration affects both the rate of Netrin-1 UNC5B complex formation and the equilibrium binding position. This demonstrates that when heparin concentrations are low, heparin is involved in a rate limiting step in the formation of the Netrin-1 UNC5B complex. Measuring the initial rate of the reaction as a function of heparin concentration, shows that the reaction order with respect to heparin is initially first order when heparin concentrations are low, and changes to zero order when concentrations of heparin are high (Figure 5.3B).

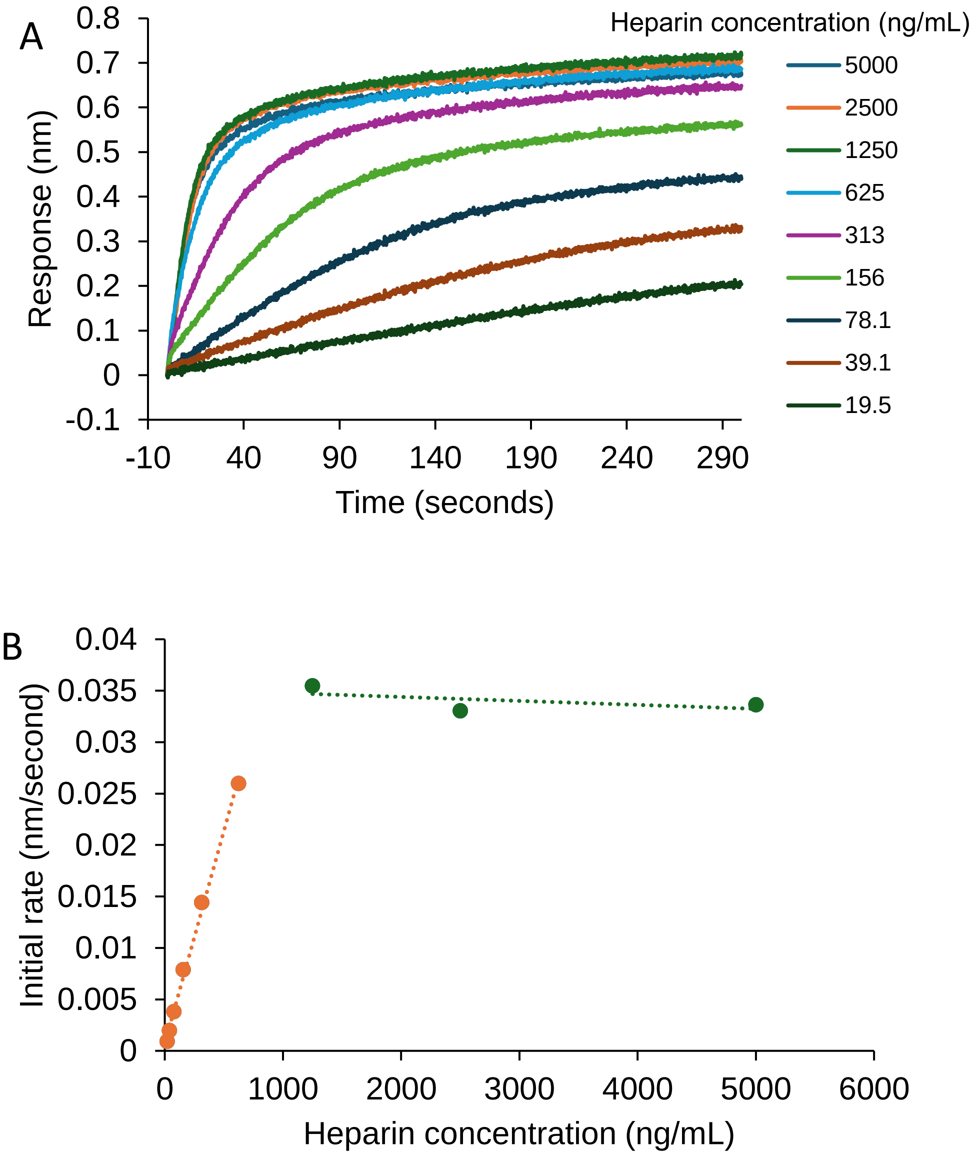


Figure 5.3. Effect of heparin concentration on binding rate of 250 nM Netrin-1 to surface immobilized UNC5B. A) Binding of Netrin-1 to UNC5B in buffer containing heparin between 19.5 to 5000 ng/mL. B) Initial binding rate of data shown in A, as a function of heparin concentration. Trendlines indicate the regions where binding is first order and zero order with respect to heparin.

The kinetic behavior shown in Figure 5.3 is consistent with mechanism 1, where Netrin-1 and heparin form a complex, which can then bind to UNC5B. In Figure 5.3B when heparin concentrations are low, the reaction rate is low because the concentration of the Netrin-1 heparin complex which binds to UNC5B is also low. As the concentration of heparin increases, the concentration of the Netrin-1 heparin complex increases and so the rate of Netrin-1 UNC5B complex formation also increases. It is under these conditions that the reaction rate is first order with respect to heparin. When there is a large excess of heparin, all available Netrin-1 will exist in the complex form and increasing the concentration of heparin further will no longer increase the rate of Netrin-1 UNC5B binding, resulting in a reaction rate that is zero order with respect to heparin. Taking the results of Figure 5.1 and Figure 5.3 together, we demonstrate that heparin is required for Netrin-1 UNC5B complex formation and under certain conditions can limit the rate of Netrin-1 UNC5B complex formation.

Competition assay development and validation

To screen and identify peptides capable of blocking the Netrin-1 UNC5B interaction we first developed a BLI based competition assay. For this assay UNC5B is immobilized on the biosensor and binding of Netrin-1 in solution to the biosensor is detected. Netrin-1 is held in solution at a constant concentration, while the inhibitor concentration is varied. To validate this competition assay we measured the inhibition of the Netrin-1 UNC5B interaction by an anti-Netrin-1 FAB previously shown to block the Netrin-1 – UNC5B interaction in vitro (24). The FAB blocked the netrin-1 UNC5B interaction in the competition assay with an IC_{50} of 44.8 ± 4.9 nM, and a Hill coefficient of -1.02 ± 0.11 (Figure 5.4A). Using the Cheng-Prusoff equation and the measured affinity between Netrin-1 and UNC5B under these competition assay conditions (Figure S5.1), the in-solution affinity of the netrin-1 FAB interaction is calculated to be 22.2 ± 4.1 nM, which is consistent with the 33.9 ± 5.2 nM Netrin-1 FAB interaction affinity measured by BLI in a direct binding experiment (Figure S5.2).

We also tested an $(ARRSR)_3$ peptide which has previously been shown to block the netrin-1 UNC5B interaction in vitro (24). In the competition assay, this peptide was found to be capable of fully blocking the netrin-1 UNC5B interaction with an IC_{50} of 1.55 ± 0.11 μM and a Hill coefficient of -1.08 ± 0.04 (Figure 5.4B).

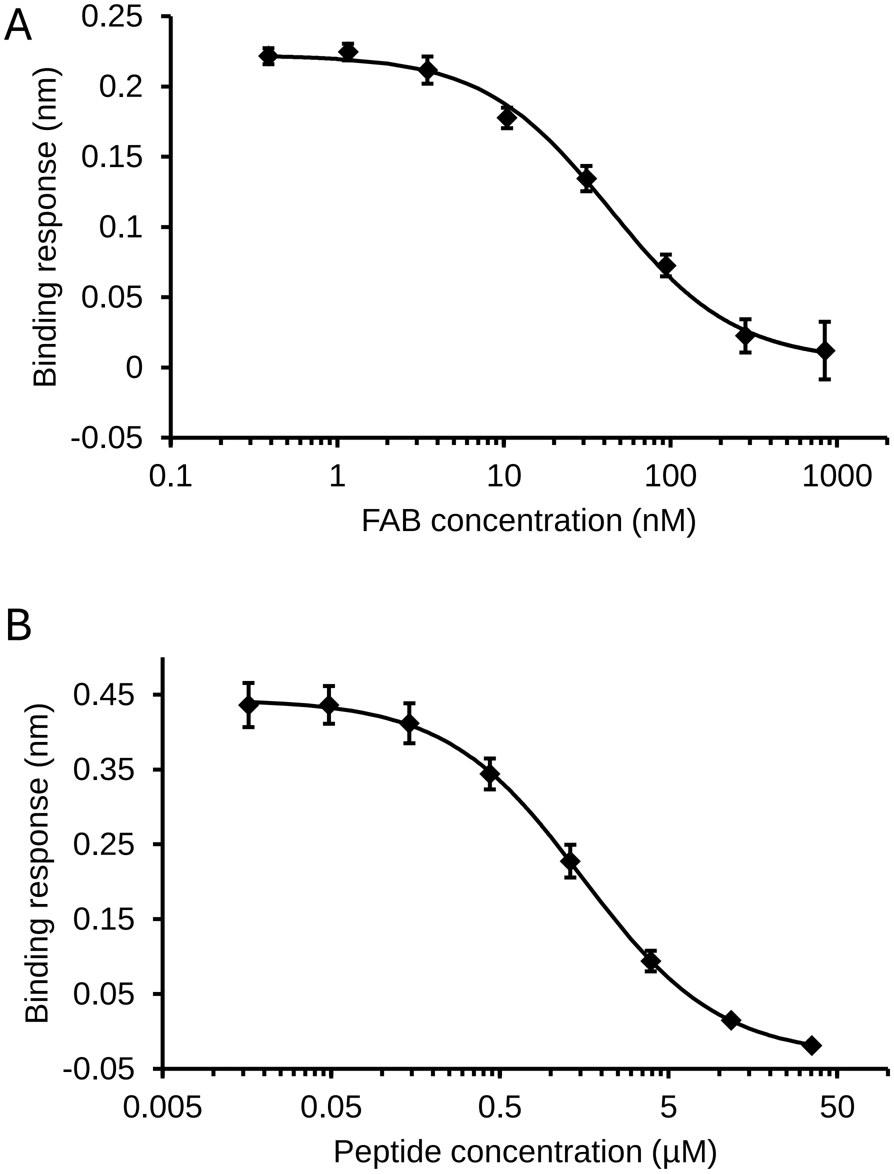


Figure 5.4: BLI competition assay showing the inhibition of the Netrin-1 UNC5B interaction by A) anti-Netrin-1 FAB (IC_{50} = 44.8 ± 4.9 nM) and B) an $(ARRSR)_3$ peptide (IC_{50} = 1.55 ± 0.11 µM)

To quantify the assay quality in terms of binding response magnitude and binding response variability, the Z’-value of the assay was measured (26). To calculate the Z’-value for this assay, the binding response of 21 Netrin-1 positive controls and 21 Netrin-1 with FAB negative controls was measured. Figure 5.5 shows that the competition assay produces an adequately large binding response with a suitably low signal variation and a Z’-value of 0.58, indicating an excellent quality assay.

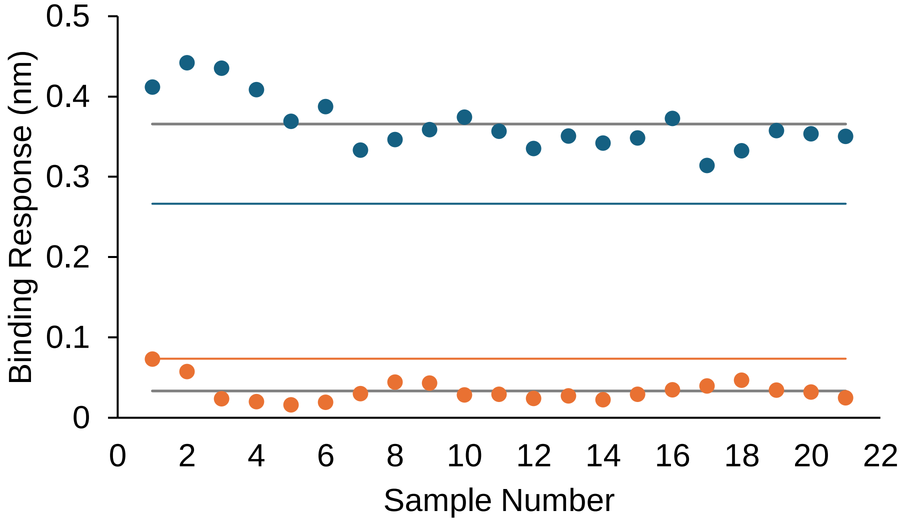


Figure 5.5. Measurement of assay quality using Z'-value. Blue dots are positive controls consisting of 12 nM netrin-1 only, orange dots are negative controls consisting of 12 nM netrin-1 with 845 nM FAB. Grey lines indicate mean of positive and negative controls. Blue line is the mean of positive control – 3* standard deviation of positive control. Orange line is mean of negative control + 3* standard deviation of negative control.

Peptide inhibitor design

The Netrin-1 LE2 domain consists of irregularly structured loop segments (loops a-d) stabilized by disulfide bridges (Figure 5.6A). A previous study has located the binding epitope of both UNC5B and anit-Netrin-1 FAB to be located on loops a and b of the Netrin-1 LE2 domain. Mutations to residues between R348 to S369 located in loops a and b reduced Netrin-1 UNC5B binding and Netrin-1 FAB binding in vitro. It was further shown that an $(ARRSR)_3$ peptide designed to mimic the R348-R349-C350-R351 motif of the LE2 domain loop a effectively blocked the Netrin-1 UNC5B interaction in vitro (24). This motif has been identified as a Cardin-Weintraub motif important for heparin binding (11). We reasoned that by designing a peptide which mimics the structure of the Netrin-1 LE2 domain loops a and b involved in Netrin-1 UNC5B binding and heparin binding, this peptide would be able to inhibit the Netrin-1 UNC5B interaction by either binding UNC5B to block Netrin-1 binding, or by binding heparin and blocking the heparin binding step required for Netrin-1 UNC5B complex formation.

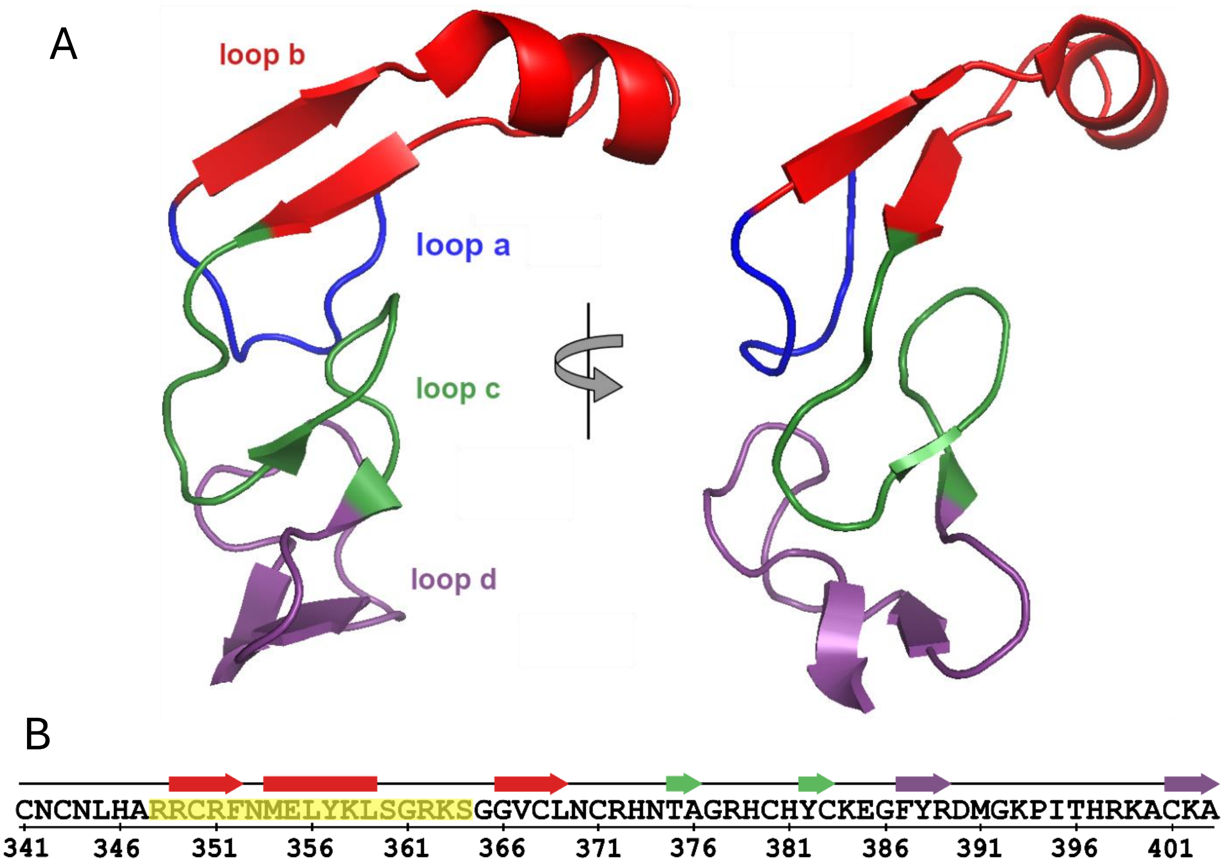


Figure 5.6. *Mus musculus* Netrin-1 LE2 domain structure and amino acid sequence. A) Netrin-1 LE2 domain structure (PDB ID: 4OVE)(27) with loops a-b indicated by color. Loop a at the N-terminus is colored blue, loop b at the top of the LE2 domain is colored red, loop c in the middle of the domain is colored green and loop d at the C-terminus of the domain is colored purple. B) Netrin-1 LE2 domain amino acid sequence with arrows indicate β-strands, rectangle indicates α-helix, yellow highlighted sequence indicates the UNC5B, anti-Netrin-1 FAB and heparin binding epitopes. Peptides have been designed to mimic the sequence highlighted in yellow. (Numbering from UniPortKB entry O09118. https://www.uniprot.org/uniprotkb/O09118/entry)

To develop peptide inhibitors of the Netrin-1 UNC5B interaction we designed peptides intended to mimic the structure and function of the heparin and UNC5B binding epitope on the Netirn-1 LE2 domain (residues 348 to 364 highlighted in yellow figure 5.6B). In total 20 unique peptides were prepared and tested in an initial library screen (Table 5.1). To effectively mimic the helix-loop structure of the LE2 domain loop b, the use of the non-proteinogenic amino acid 2-aminoisobutyrate (Aib) and peptide cyclization by lactam bond formation between lysine and aspartate side chains was tested.

By analyzing the IC_{50} values listed in Table 5.1, several general yet important conclusions can be made. First, peptides which incorporated Aib or cyclization showed increased potency compared to peptide 20, a linear sequence of residues 353 to 364 which shows no inhibition in the IC_{50} assay. Second, cyclized peptides were consistently more potent than peptides which used Aib. Third, addition of RRSRF sequence to the peptide N-terminus consistently improved the potency of the peptides by a substantial amount. To verify these results, peptides which showed the highest affinity in this initial screen were then ordered from a commercial supplier with purity > 95%. To verify that the RRSRF sequence is important for peptide binding affinity, two versions of each peptide were ordered, one version with this RRSRF sequence and one without. To obtain more accurate IC_{50} values, these commercially prepared peptides were tested in the competition assay. While some IC_{50} values were considerably different from the values obtained in the initial screen, peptides 1, 2, 3 and 10 remain potent inhibitors of the Netrin-1 UNC5B interaction, and the three general conclusions made in the initial screen hold true.

Table 5.1. Peptides inhibitor of the Netrin-1 UNC5B interaction.

Number ID	Amino acid sequence	IC_{50} (μM)
1	*Ac*-RRSRFN{Nle}E**K**YR**D**TGRKS-NH_2	7.95 [b] 0.866
2	*Ac*-RRSRFNME**K**YRL**D**LRKS-NH_2	48.6 [b] 0.866
3	*Ac*-RRSRFN{Aib}EL{Aib}KL**K**GRR**D**-NH_2	6.03 [b] 1.74
4	*Ac*-NME**K**YRL**D**LRKS-NH_2	493 [b] 5.16
5	*Ac*-RRSRF{Aib}{Nle}E{Aib}LKL{Aib}ARKS-NH_2	10.4
6	*Ac*-RRSRFN{Aib}EL{Aib}KL{Aib}ARRS-NH_2	430 [b] 14.0
7	*Ac*-N{Aib}EL{Aib}KL**K**GRR**D**-NH_2	248 [b] 42.4
8	*Ac*-RRSRFN{Aib}EL{Aib}KLSGRKS-NH_2	43.8
9	*Ac*-NMELYKL**K**GRR**D**-NH_2	49.3
10	*Ac*-NME**K**YR**D**TGRKS-NH_2	12.9 [b] 52.4
11	*Ac*-N{Aib}EL{Aib}KL{Aib}ARRS-NH_2	3410 [b] 124
12	*Ac*-LKLAARKS-NH_2	137
13	*Ac*-{Aib}{Nle}E{Aib}LKL{Aib}ARKS-NH_2	165
14	*Ac*-LKL{Aib}ARKS-NH_2	198
15	*Ac*-{Aib}{Nle}E{Aib}YKLSGRKS-NH_2	205
16	*Ac*-{Aib}EE{Aib}LKL{Aib}ARKS-NH_2	333
17	*Ac*-{Aib}{Nle}E{Aib}YKL{Aib}ARKS-NH_2	370
18	*Ac*-N{Aib}EL{Aib}KLSGRKS-NH_2	635
19	*Ac*-NME{Aib}YK{Aib}TGRKS-NH_2	No inhibition at 413 μM
20	*Ac*-NMELYKLSGRKS-NH_2	No inhibition at 525 μM

[a] {Aib} is 2-aminoisobutyrate, {Nle} is norleucine, K and D amino acids in bold connected with “ └┘ ” symbol are connected by lactam bond via their side chain functional groups.

[b] These IC_{50} values were measured using commercially prepared peptides with purity > 95%.

Peptide mechanism of action

The peptides identified as inhibitor of the Netirn-1 UNC5B interaction are designed to mimic the structure of the heparin and UNC5B binding epitope of the Netrin-1 LE2 domain. Therefore, these peptides could act as competitive inhibitors binding to UNC5B and out-competing Netrin-1, or by inhibiting the heparin binding steps required for Netrin-1 UNC5B complex formation. To test which mechanism the peptides use to block the Netrin-1 UNC5B interaction we performed ITC binding experiments with the three most potent peptide versions identified in the competition assay (peptides 1, 2 and 3 in Table 5.1). ITC binding experiments between heparin and the peptides show a strong exothermic interaction with micromolar affinity (Figure 5.7A-C). In contrast binding experiments between UNC5B and each of the peptides tested produced a heat signal that is too small to justify the existence of an interaction between the peptide and UNC5B of sufficient affinity block the Netrin-1 UNC5B interaction (Figure 5.8A-C). Taken together this demonstrates that the peptides inhibit the Netrin-1 UNC5B interaction by disrupting the binding steps involving the heparin co-receptor.

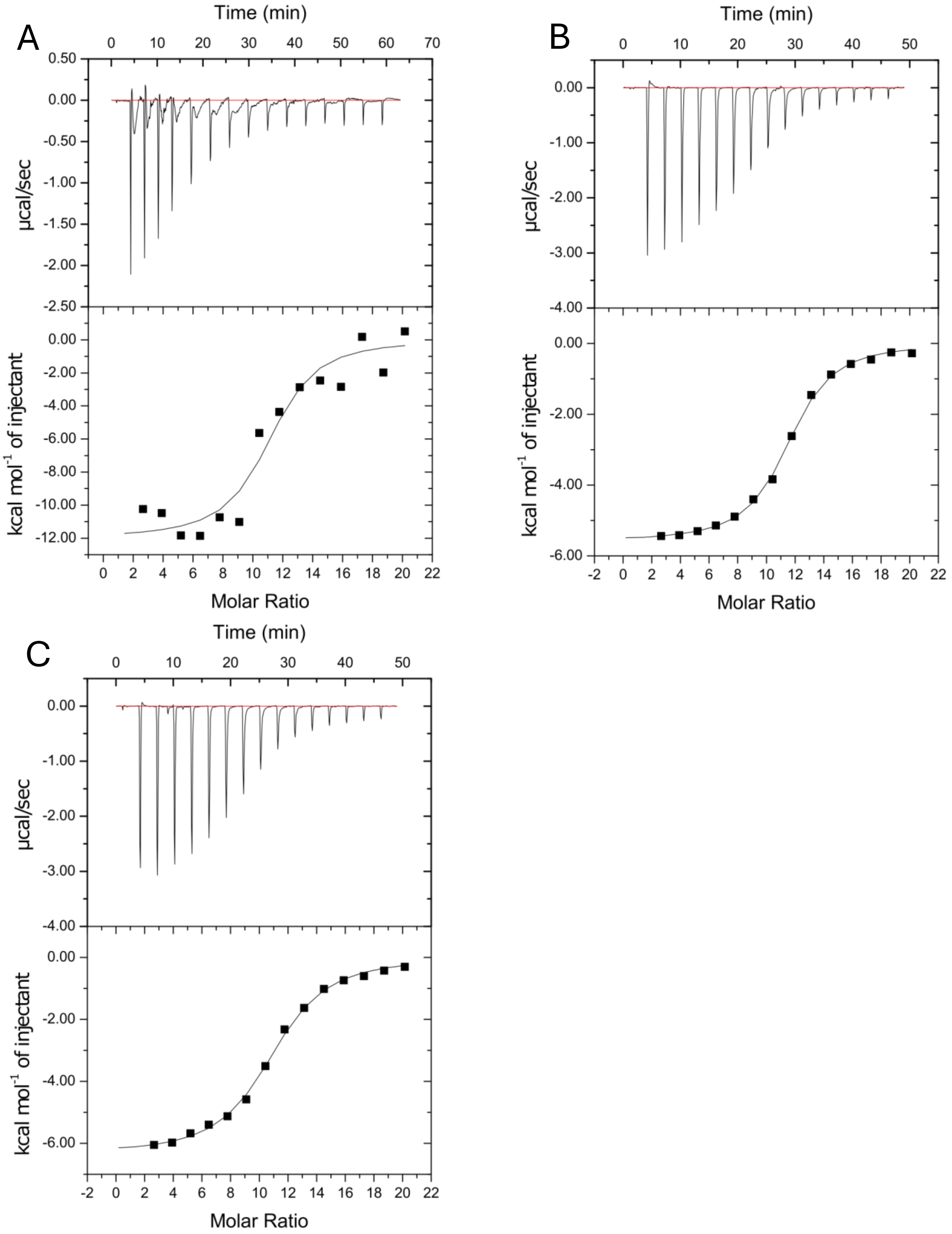
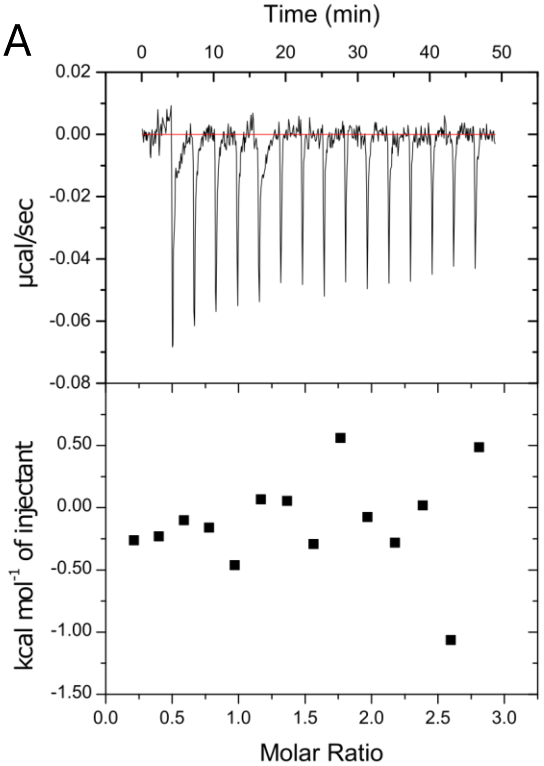
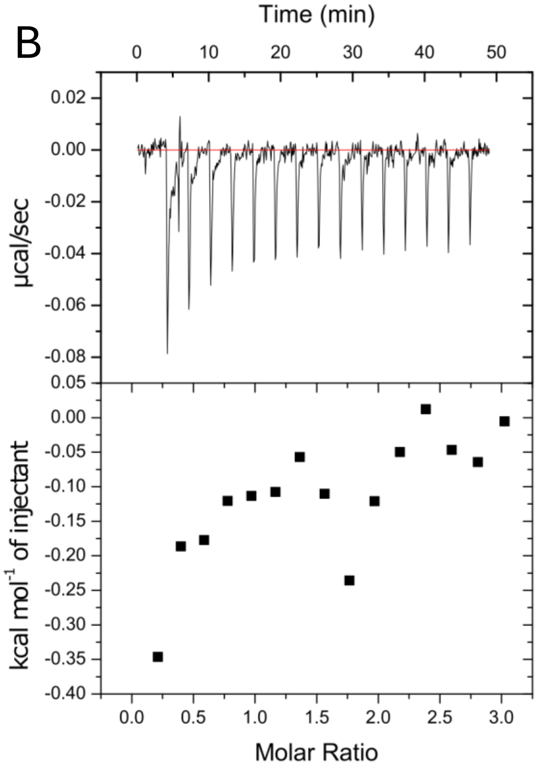

A
Time (min)
0
10
20
30
40
50
60
70
µcal/sec
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
kcal mol^{-1} of injectant
0.00
-2.00
-4.00
-6.00
-8.00
-10.00
-12.00
0
2
4
6
8
10
12
14
16
18
20
22
Molar Ratio
B
Time (min)
0
10
20
30
40
50
µcal/sec
0.00
-1.00
-2.00
-3.00
-4.00
kcal mol^{-1} of injectant
0.00
-2.00
-4.00
-6.00
-2
0
2
4
6
8
10
12
14
16
18
20
22
Molar Ratio
C
Time (min)
0
10
20
30
40
50
µcal/sec
0.00
-1.00
-2.00
-3.00
-4.00
kcal mol^{-1} of injectant
0.00
-2.00
-4.00
-6.00
-2
0
2
4
6
8
10
12
14
16
18
20
22
Molar Ratio

Figure 5.7. ITC measurement of the titration of porcine intestinal mucosa heparin with peptide. A) Peptide 1, K_D 2.2 x 10^{-6} ± 1.4 x 10^{-6} M, N 10.8 ± 0.5, ΔH -1.2 x 10^{4} ± 900 cal mol^{-1}, ΔS -14.4 cal mol^{-1} deg^{-1}. B) Peptide 2, K_D 5.1 x 10^{-6} ± 0.4 x 10^{-6} M, N 11.2 ± 0.1, ΔH -5590 ± 50 cal mol^{-1}, ΔS 5.46 cal mol^{-1} deg^{-1}. C) Peptide 3, K_D 8.7 x 10^{-6} ± 0.7 x 10^{-6} M, N 10.7 ± 0.1, ΔH -6350 ± 75 cal mol^{-1}, ΔS 1.85 cal mol^{-1} deg^{-1}.


A
Time (min)
0
10
20
30
40
50
0.02
0.00
-0.02
-0.04
-0.06
-0.08
µcal/sec
0.50
0.00
-0.50
-1.00
-1.50
kcal mol^{-1} of injectant
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Molar Ratio


B
Time (min)
0
10
20
30
40
50
0.02
0.00
-0.02
-0.04
-0.06
-0.08
µcal/sec
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
-0.35
-0.40
kcal mol^{-1} of injectant
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Molar Ratio

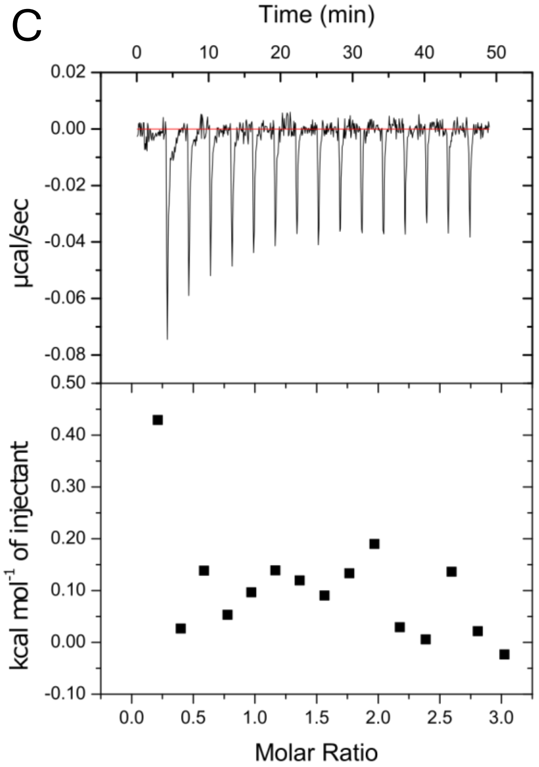

C
Time (min)
0
10
20
30
40
50
0.02
0.00
-0.02
-0.04
-0.06
-0.08
µcal/sec
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
kcal mol^{-1} of injectant
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Molar Ratio

Figure 5.8. ITC measurement of the titration of UNC5B with peptide. A) Peptide 1 B) Peptide 2 and C) Peptide 3. Each peptide produces a near zero heat signal after subtraction of the peptide dilution heat signal, indicating the absence of a binding interaction.

5.2.3 Discussion

Protein – heparin interactions are known to play a critical role in numerous biological processes with well-studied examples being the effects of heparin on anticoagulation, fibroblast growth factor signaling, and the recognition of cell surface heparins by viral pathogens. Heparin effects cell signaling through multiple mechanisms, such as serving as a co-receptor for ligand-receptor binding, inducing ligand or receptor multimerization, and facilitating ligand immobilization and gradient formation (17). Multiple studies have shown that heparins are involved in Netrin-1 signaling, and Netrin-1 is a well-known to have multiple high affinity heparin binding sites on both the C-terminal domain and the LE2 domain (11, 28–30). It has previously been shown that mutations to the heparin binding Cardin-Weintraub motif on the Netrin-1 LE2 domain reduce UNC5B binding and interfere with Netrin-1 signaling in *C.elegans* (11, 24). In addition, the anti-Netrin-1 antibody that blocks the Netrin-1 UNC5B interaction binds directly adjacent to the Cardin-Weintraub motif of the LE2 domain (24). From this available evidence it has been proposed that heparin and UNC5B binding epitopes are in close spatial proximity or that the UNC5B requires a heparin co-factor (11, 31). By assessing the Netrin-1 UNC5B binding affinity in the presence and absence of heparin, we demonstrate heparin is required for the Netrin-1 UNC5B interaction to occur. Through a kinetic analysis of Netrin-1 UNC5B binding in the presence of varying heparin concentrations, we further demonstrate that low heparin concentrations limit the rate of Netrin-1 UNC5B complex formation. From this kinetic information we proposed a possible binding mechanism where Netrin-1 first binds heparin, then this Netrin-1 heparin complex binds to UNC5B. Based on this mechanism we rationalized that the Netrin-1 UNC5B interaction could be inhibited by blocking the first Netrin-1 heparin binding step or the second Netrin-1 UNC5B binding step. We then developed peptide inhibitors of the Netrin-1 UNC5 interaction and demonstrated through ITC binding experiments that the inhibitory mechanism of these peptides involves binding to heparin and not UNC5B. Taken together the results of these binding experiments conclusively demonstrate the function of heparin as a co-receptor for the Netrin-1 UNC5B binding interaction, and further demonstrate that the Netrin-1 UNC5B binding can be inhibited by disrupting the Netrin-1 heparin interaction.

Protein- heparin interactions have emerged as promising therapeutic targets, given their involvement in numerous critical biological processes (18). Several studies have used heparin mimetics to modulate protein- heparin interactions, however these heparin mimetics often have off target effects such as anticoagulant activity (19). In this work we use an alternate approach that relies on a peptide mimicking the heparin binding region of the protein. We have identified several peptides designed to mimic a segment of the Netrin-1 LE2 domain comprising the UNC5B and heparin binding epitope. Our biophysical assays demonstrate these peptides are effective inhibitors of the Netrin-1 UNC5B interaction, with their mechanism of action involving the binding of the heparin co-factor. These peptides may therefore serve as leads in the development of therapeutics targeting the heparin mediated Netrin-1 UNC5B signaling. The peptides designed in this work may similarly serve as valuable research tools to better understand the role heparin plays in Netirn-1 function.

The ability to selectively modulate the protein-protein interactions of the biological interactome would enable the regulation of biological processes for the treatment of disease. However, development of small molecule drugs that target protein-protein interactions is a challenging task due to the physical properties of protein – protein interaction interfaces (32). Because many extracellular protein – protein interactions are modulated by heparins, targeting heparin binding may be an effective means of disrupting protein – protein interactions with small molecules. Heparin binding sites may be more suitable targets for small molecule drugs like heparin mimetics or peptides similar to the ones described in this work. Further research would be needed to determine whether these peptides can block the Netrin-1 UNC5B interaction in cell assays and animal models. Additional studies aimed at improving the potency and selectivity of these peptides for disrupting the Netrin-1 UNC5B interaction will likely also be required. Because disrupting the Netrin-1 UNC5B interaction is a promising drug target for cancer treatment, it would also be worthwhile exploring the development of heparin mimetics capable of disrupting the Netrin-1 UNC5B interaction.

5.2.4 Methods

Protein expression and purification

Gallus gallus Netrin-1 domains LN to LE3 (UniProtKB accession number Q90922, amino acids 26 – 458), *Mus musculus* UNC5B domains IG1 to TSP1 with a C-terminal His tag (UniProtKB accession number Q8K1S3, amino acids 48 – 300) and a previously described anti-Netrin-1 FAB were cloned expressed and purified following a previously published protocols (24, 33). Removal of heparin bound to Netrin-1 was carried out by dialysis and size exclusion chromatography in a buffer of 50 mM TRIS pH 7.5, 500 mM NaCl, 500 mM $(NH_4)_2SO_4$ as previously described (11). All proteins were quantified using their absorbance at 280 nm and an extinction coefficient and molecular weight calculated from their amino acid sequence with the ExPASy ProtParam tool (34).

Biolayer interferometry assays

All BLI binding experiments were carried out on an Octet K2 instrument (Sartorius AG) at 25^{o}C with 1000 RPM shaking. UNC5B was immobilized on Penta-His specific antibody biosensors (HIS1K) (Sartorius AG). Netrin-1 and porcine intestinal mucosa heparin (EMD Millipore Corp. USA) were biotinylated using EZ-Link NHS-LC-LC-biotin (Thermo Fisher Scientific Inc.) and immobilized on Streptavidin biosensor (Sartorius AG). Netrin-1 UNC5B binding experiments were conducted in 100 mM HEPES pH 7.3, 150 mM NaCl, 1 mg/mL BSA with porcine intestinal mucosa heparin (EMD Millipore Corp. USA) added to the buffer at a concentration specific to the experiment being conducted. Binding of Netrin-1 or UNC5B to heparin was performed in

100 mM HEPES pH 7.3, 150 mM NaCl, 1 mg/mL BSA. In all binding experiments were processed using the Data Analysis HT version 11.1.3.50 program (Sartorius AG). A reference sensor in buffer was used to subtract baseline sensor drift. Data was aligned to a binding response of zero at the end of the baseline step and fit to a 1:1 or 2:1 binding model, by globally fitting the association and dissociation steps with sensor R_{max} values unlinked and dissociation assumed to reach pre-association baseline. To calculate the initial rate of Netrin-1 UNC5B binding as a function of heparin concentration, data was collected and processed as described for the Netrin-1 UNC5B interaction. The initial rate was then calculated from the slope of the initial linear portion of the reaction progress curve.

Biolayer interferometry competition assays

Competition experiments were preformed in the same manner as the Netrin-1 UNC5B binding experiments with the following modifications. The competition assay buffer was composed of 100 mM MES pH 6.5, 150 mM NaCl, 1 mg/mL BSA. Netrin-1 was used in solution at a concentration of 12 nM and the binding between Netrin-1 and UNC5B was measured for 200s in the presence of varying concentrations of inhibitor. Each experiment included a buffer only reference sensor and a positive control of 12 nM Netrin-1 with no inhibitor. Data was processed by subtracting the reference sensor to account for baseline drift and aligning binding response to zero at the end of the baseline step. The binding response at the end of the 200s association step was plotted as a function of inhibitor concentration and fit to a 4-parameter logistic function in QtiPlot (IONDEV SRL) to determine IC_{50} values and the Hill coefficient.

To calculate the in-solution affinity of the anti-Netrin-1 FAB for Netrin-1 from the IC_{50} value, the binding affinity of Netrin-1 to UNC5B in the competition assay buffer was measured following the same method used for the Netrin-1 UNC5B binding described above. Using this Netrin-1 UNC5B binding affinity and the IC_{50} value of the FAB, the in-solution affinity for interaction between Netrin-1 and the anti-Netrin-1 FAB could be calculated using the Cheng-Prusoff equation. The binding affinity of the anti-Netrin-1 FAB to Netrin-1 interaction was measured using biotinylated Netrin-1 immobilized on Streptavidin biosensors in competition assay buffer. Data was collection, processing and analysis was performed following the same method used for the Netrin-1 UNC5B interaction described above.

Measurement of Z'-value for this assay was performed under the same conditions used for the competition assay. The binding response of 21 Netrin-1 positive controls and 21 Netrin-1 with FAB negative controls was measured. The Netrin-1 positive controls contained 12 nM Netrin-1 and the Negative controls contained 12 nM Netrin-1 with 845 nM anti-Netrin-1 FAB. Calculation of the Z'-value was performed according to Zhang 1999.

Peptide library screening

The initial library of peptides used to identify peptides capable of inhibiting the Netrin-1 UNC5B interaction was prepared by conventional solid phase peptide synthesis techniques and existed as the trifluoracetic acid salt with purity ranging from 20 to 95%. Purity was assessed by HPLC with UV-Vis detection at 220 nm. Peptide composition was measured by mass-spectrometry. From screening these peptides, several leads were identified and ordered from Genscript with a purity of > 95 % and came as the trifluoracetic acid salt. The ARRSRARRSRARRSR peptide referred to as $(ARRSR)_3$ was synthesized by CanPeptide Inc to a purity > 95% and used trifluoracetic acid as a counter ion. Commercially prepared peptide purity was assessed by HPLC with UV-Vis detection at 220 nm and peptide composition was assessed by mass-spectrometry by the peptide supplier.

Isothermal titration calorimetry

ITC was performed on a MicroCal iTC 200 titration calorimeter (Malvern Panalytical Ltd.) in 20 mM HEPES pH 7.3, 150 mM NaCl buffer. The experiments were performed at 25 °C with a mixing speed of 500 rpm, a reference power of 11 µcal/sec in the high gain mode, and the reference cell was filled with water. The titration of porcine intestinal mucosa heparin (EMD Millipore Corp. USA) with peptide consisted of an initial 0.4 µL injection over 0.8 seconds, followed by 15 injections of 2.49 µL over 4.98 seconds and 180 seconds between injections. For peptide 1 the concentration of heparin was estimated to be 15 µM and the concentration of peptide 1 was 1 mM. For peptide 2 and 3, the estimated concentration of heparin was 25 µM and the concentration of heparin was 2.5 mM. The titration of 50 µM UNC5B with 750 µM peptide consisted of an initial 0.4 µL injection over 0.8 seconds followed by 15 injections of 2.49 µL over 4.98 seconds and 180 seconds between injections. The concentration of heparin was determined assuming a molecule weight of 15 kDa, the concentration of UNC5B was determined using the absorbance at 280 nm and an extinction coefficient of 58160 $M^{-1}cm^{-1}$ determined using Expasy ProtParam tool. Peptide concentration was determined from the absorbance of the peptide solution at 205 nm and an extinction coefficient of 31 mL mg^{-1} cm^{-1} on a NanoDrop One spectrophotometer (ThermoFisher Scientific). A reference experiment to account for the heat of dilution of peptide was performed under the same conditions, except the measurement cell was filled with buffer rather than heparin or UNC5B. Data analysis, reference subtraction and fitting to a one set of sites binding model was performed in the MicroCal supplied Origin 7.0 software package (OriginLab Corp.).

5.2.5 Supplemental Information

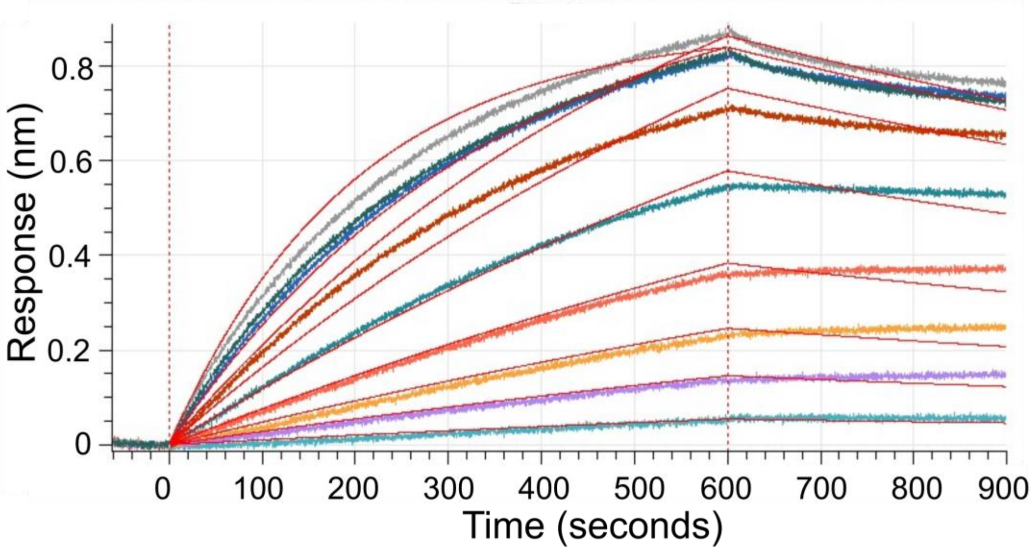


Figure S5.1. Biolayer interferometry binding assay detecting the binding of Netrin-1 in solution to surface immobilized UNC5B under the conditions used in the competition assay. K_D of 11.8 ± 1.3 nM.

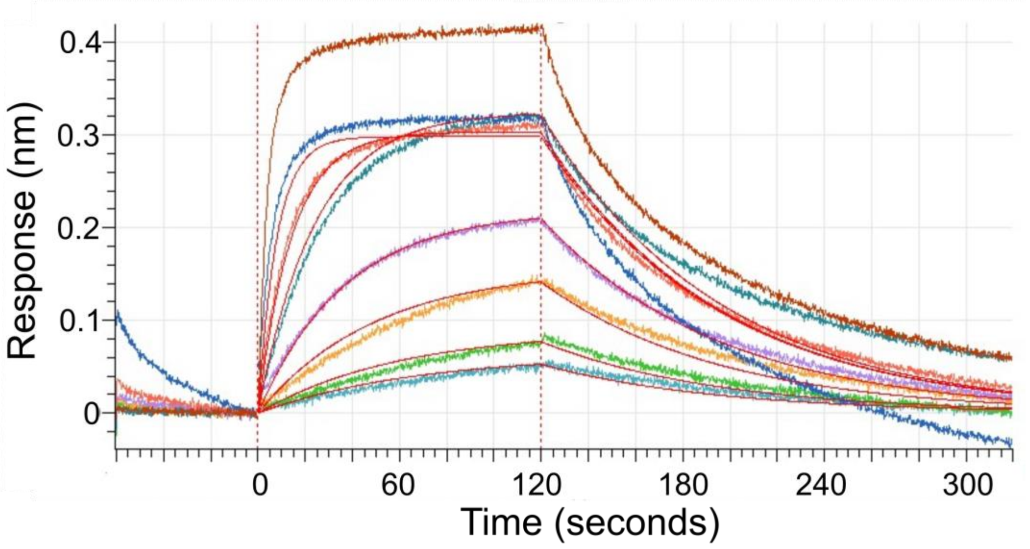


Figure S5.2: Biolayer interferometry binding assay detecting the binding of anti-Netrin-1 FAB in solution to surface immobilized Netrin-1. K_D of 33.9 ± 5.2 nM.

5.2.6 References

1. Serafini, T., Kennedy, T. E., Gaiko, M. J., Mirzayan, C., Jessell, T. M., and Tessier-Lavigne, M. (1994) The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. *Cell*. **78**, 409–424
2. Kennedy, T. E., Serafini, T., de la Torre, J. R., and Tessier-Lavigne, M. (1994) Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell*. **78**, 425–435
3. Ly, N. P., Komatsuzaki, K., Fraser, I. P., Tseng, A. A., Prodhan, P., Moore, K. J., and Kinane, T. B. (2005) Netrin-1 inhibits leukocyte migration in vitro and in vivo. *Proc. Natl. Acad. Sci.* **102**, 14729–14734
4. Ramkhelawon, B., Hennessy, E. J., Ménager, M., Ray, T. D., Sheedy, F. J., Hutchison, S., Wanschel, A., Oldebeken, S., Geoffrion, M., Spiro, W., Miller, G., McPherson, R., Rayner, K. J., and Moore, K. J. (2014) Netrin-1 promotes adipose tissue macrophage retention and insulin resistance in obesity. *Nat. Med.* **20**, 377–384
5. Rosenberger, P., Schwab, J. M., Mirakaj, V., Masekowsky, E., Mager, A., Morote-Garcia, J. C., Unertl, K., and Eltzschig, H. K. (2009) Hypoxia-inducible factor–dependent induction of netrin-1 dampens inflammation caused by hypoxia. *Nat. Immunol.* **10**, 195–202
6. Park, K. W., Crouse, D., Lee, M., Karnik, S. K., Sorensen, L. K., Murphy, K. J., Kuo, C. J., and Li, D. Y. (2004) The axonal attractant Netrin-1 is an angiogenic factor. *Proc. Natl. Acad. Sci.* **101**, 16210–16215
7. Fitamant, J., Guenebeaud, C., Coissieux, M.-M., Guix, C., Treilleux, I., Scoazec, J.-Y., Bachelot, T., Bernet, A., and Mehlen, P. (2008) Netrin-1 expression confers a selective advantage for tumor cell survival in metastatic breast cancer. *Proc. Natl. Acad. Sci.* **105**, 4850–4855
8. Arakawa, H. (2004) Netrin-1 and its receptors in tumorigenesis. *Nat. Rev. Cancer*. **4**, 978–987
9. Keino-Masu, K., Masu, M., Hinck, L., Leonardo, E. D., Chan, S. S.-Y., Culotti, J. G., and Tessier-Lavigne, M. (1996) Deleted in Colorectal Cancer (DCC) Encodes a Netrin Receptor. *Cell*. **87**, 175–185
10. Leonardo, E. D., Hinck, L., Masu, M., Keino-Masu, K., Ackerman, S. L., and Tessier-Lavigne, M. (1997) Vertebrate homologues of C. elegans UNC-5 are candidate netrin receptors. *Nature*. **386**, 833–838
11. Meier, M., Gupta, M., Akgül, S., McDougall, M., Imhof, T., Nikodemus, D., Reuten, R., Moya-Torres, A., To, V., Ferens, F., Heide, F., Padilla-Meier, G. P., Kukura, P., Huang, W., Gerisch, B., Mörgelin, M., Poole, K., Antebi, A., Koch, M., and Stetefeld, J. (2023) The dynamic nature of netrin-1 and the structural basis for glycosaminoglycan fragment-induced filament formation. *Nat. Commun.* **14**, 1226
12. Mehlen, P., Delloye-Bourgeois, C., and Chédotal, A. (2011) Novel roles for Slits and netrins: axon guidance cues as anticancer targets? *Nat. Rev. Cancer*. **11**, 188–197
13. Paradisi, A., Maisse, C., Coissieux, M.-M., Gadot, N., Lépinasse, F., Delloye-Bourgeois, C., Delcros, J.-G., Svrcek, M., Neufert, C., Fléjou, J.-F., Scoazec, J.-Y., and Mehlen, P. (2009) Netrin-1 up-regulation in inflammatory bowel diseases is required for colorectal cancer progression. *Proc. Natl. Acad. Sci.* **106**, 17146–17151
14. Sung, P.-J., Rama, N., Imbach, J., Fiore, S., Ducarouge, B., Neves, D., Chen, H.-W., Bernard, D., Yang, P.-C., Bernet, A., Depil, S., and Mehlen, P. (2019) Cancer-Associated Fibroblasts Produce Netrin-1 to Control Cancer Cell Plasticity. *Cancer Res.* **79**, 3651–3661
15. Cassier, P. A., Navaridas, R., Bellina, M., Rama, N., Ducarouge, B., Hernandez-Vargas, H., Delord, J.-P., Lengrand, J., Paradisi, A., Fattet, L., Garin, G., Gheit, H., Dalban, C., Pastushenko, I., Neves, D., Jelin, R., Gadot, N., Braissand, N., Léon, S., Degletagne, C., Matias-Guiu, X., Devouassoux-Shisheboran, M., Mery-Lamarche, E., Allard, J., Zindy, E.,

Decaestecker, C., Salmon, I., Perol, D., Dolcet, X., Ray-Coquard, I., Blanpain, C., Bernet, A., and Mehlen, P. (2023) Netrin-1 blockade inhibits tumour growth and EMT features in endometrial cancer. *Nature*. **620**, 409–416

16. Boussouar, A., Tortereau, A., Manceau, A., Paradisi, A., Gadot, N., Vial, J., Neves, D., Larue, L., Battistella, M., Leboeuf, C., Lebbé, C., Janin, A., and Mehlen, P. (2020) Netrin-1 and Its Receptor DCC Are Causally Implicated in Melanoma Progression. *Cancer Res.* **80**, 747–756
17. Linhardt, R. J., and Toida, T. (2004) Role of Glycosaminoglycans in Cellular Communication. *Acc. Chem. Res.* **37**, 431–438
18. Sasisekharan, R., Shriver, Z., Venkataraman, G., and Narayanasami, U. (2002) Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat. Rev. Cancer*. **2**, 521–528
19. Ghiselli, G. (2019) Heparin Binding Proteins as Therapeutic Target: An Historical Account and Current Trends. *Medicines*. **6**, 80
20. Dogra, P., Martin, E. B., Williams, A., Richardson, R. L., Foster, J. S., Hackenback, N., Kennel, S. J., Sparer, T. E., and Wall, J. S. (2015) Novel Heparan Sulfate-Binding Peptides for Blocking Herpesvirus Entry. *PLOS ONE*. **10**, e0126239
21. Schick, B. P., Maslow, D., Moshinski, A., and Antonio, J. D. S. (2004) Novel concatameric heparin-binding peptides reverse heparin and low-molecular-weight heparin anticoagulant activities in patient plasma in vitro and in rats in vivo. *Blood*. **103**, 1356–1363
22. Martínez-Burgo, B., Cobb, S. L., Pohl, E., Kashanin, D., Paul, T., Kirby, J. A., Sheerin, N. S., and Ali, S. (2019) A C-terminal CXCL8 peptide based on chemokine–glycosaminoglycan interactions reduces neutrophil adhesion and migration during inflammation. *Immunology*. **157**, 173–184
23. Mader, J. S., Smyth, D., Marshall, J., and Hoskin, D. W. (2006) Bovine Lactoferricin Inhibits Basic Fibroblast Growth Factor- and Vascular Endothelial Growth Factor165-Induced Angiogenesis by Competing for Heparin-Like Binding Sites on Endothelial Cells. *Am. J. Pathol.* **169**, 1753–1766
24. Grandin, M., Meier, M., Delcros, J. G., Nikodemus, D., Reuten, R., Patel, T. R., Goldschneider, D., Orriss, G., Krahn, N., Boussouar, A., Abes, R., Dean, Y., Neves, D., Bernet, A., Depil, S., Schneiders, F., Poole, K., Dante, R., Koch, M., Mehlen, P., and Stetefeld, J. (2016) Structural decoding of the Netrin-1/UNC5 interaction and its therapeutical implications in cancers. *Cancer Cell*. **29**, 173–185
25. Stetefeld, J., McDougall, M., Loewen, P. C., Moya-Torres, A., and Meier, M. (2023) Structural elucidation of the Ectodomain of mouse UNC5H2. *Protein Data Bank 6OOL*
26. Zhang, J. H., Chung, T. D., and Oldenburg, K. R. (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **4**, 67–73
27. Meier, M., Nikodemus, D., Reuten, R., Patel, T. R., Orriss, G., Okun, N., Koch, M., and Stetefeld, J. (2020) X-ray Crystal Structure of Mouse Netrin-1. *Protien Data Bank 4OVE*
28. Matsumoto, Y., Irie, F., Inatani, M., Tessier-Lavigne, M., and Yamaguchi, Y. (2007) Netrin-1/DCC Signaling in Commissural Axon Guidance Requires Cell-Autonomous Expression of Heparan Sulfate. *J. Neurosci.* **27**, 4342–4350
29. Priest, J. M., Nichols, E. L., Smock, R. G., Hopkins, J. B., Mendoza, J. L., Meijers, R., Shen, K., and Özkan, E. (2024) Structural insights into the formation of repulsive netrin guidance complexes. *Sci. Adv.* **10**, eadj8083
30. Kappler, J., Franken, S., Junghans, U., Hoffmann, R., Linke, T., Müller, H. W., and Koch, K.-W. (2000) Glycosaminoglycan-Binding Properties and Secondary Structure of the C-Terminus of Netrin-1. *Biochem. Biophys. Res. Commun.* **271**, 287–291
31. Finci, L. I., Krüger, N., Sun, X., Zhang, J., Chegkazi, M., Wu, Y., Schenk, G., Mertens, H. D. T., Svergun, D. I., Zhang, Y., Wang, J., and Meijers, R. (2014) The Crystal Structure of

Netrin-1 in Complex with DCC Reveals the Bifunctionality of Netrin-1 As a Guidance Cue. *Neuron*. **83**, 839–849

32. Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., and Shi, J. (2020) Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduct. Target. Ther.* **5**, 1–23
33. Moya-Torres, A., Gupta, M., Heide, F., Krahn, N., Legare, S., Nikodemus, D., Imhof, T., Meier, M., Koch, M., and Stetefeld, J. (2021) Homogenous overexpression of the extracellular matrix protein Netrin-1 in a hollow fiber bioreactor. *Appl. Microbiol. Biotechnol.* **105**, 6047–6057
34. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (Walker, J. M. ed), pp. 571–607, Humana Press, Totowa, NJ, 10.1385/1-59259-890-0:571

Chapter 6

Conclusion and future directions

6.1 The importance of studying protein – ligand interactions

Understanding protein - ligand interactions is of fundamental importance for understanding the molecular mechanisms of biological processes, and for developing therapies for diseases or other valuable biotechnologies. This thesis presents the study of the protein – ligand interactions involved in three unique biological systems. This work advances the understanding of these protein – ligand interactions, demonstrates improved methods for identifying inhibitors of protein – ligand interactions and develops inhibitors of protein – ligand interactions.

6.2 Identifying the molecular basis of Laminin N-terminal domain Ca^{2+} binding using a hybrid approach: Conclusions and future directions

The Ca^{2+} binding Laminin N-terminal (LN) domain is a protein domain found in members of the Netrin and Laminin protein families. It has previously been shown that Ca^{2+} binding is required for the proper function of several proteins containing LN domains. This requirement for Ca^{2+} has largely been attributed to the role of Ca^{2+} in structuring the Ca^{2+} binding loop of these LN domains (1, 2). Through detailed structural and computational analysis, we identified a unique non-continuous Ca^{2+} binding motif within the Netrin-1 LN domain. Bioinformatic analysis revealed the conservation of this Ca^{2+} binding motif across known Ca^{2+}-coordinating LN domains. This finding enables the sequence - based prediction of Ca^{2+} binding ability in LN domains. Biophysical investigation using ITC and TSA, demonstrate this LN domain Ca^{2+} binding motif exhibits high affinity and specificity for Ca^{2+} and is required for the thermal stabilization of the Netrin-1 and Laminin-γ LN domains. These findings enhance our understanding of LN domain structure and demonstrate the role of Ca^{2+} binding with respect to LN domain function. Furthermore, the characterization of this LN domain Ca^{2+} binding on a fundamental level may help improve the understanding and study of protein – cation interactions.

Recently computational methods such as AlphaFold and RoseTTAfold have made major advancements in the prediction of protein structure from a protein's amino acid sequence (3, 4). In addition to being able to accurately predict protein structures, these methods can be easily and conveniently used by researchers (5). Despite the remarkable ability of these programs to predict protein structure, they lack the ability to include the ligands important to biological function such as co-factors and coordinated metal ions. Researchers have recently begun developing computation approaches to include these ligands in predicted protein structures (6). Development and validation of these computational methods relies on existing experimental data. As such, the understanding of the LN domain – Ca^{2+} interaction developed in this work may serve as a valuable resource for researchers seeking to understand metal ion binding in related proteins and protein domains and may aid the development of computational tools for predicting metal ion binding sites.

6.3 Improved SARS-CoV-2 main protease high-throughput screening assay using a 5-carboxyfluorescein substrate: Conclusions and future directions

Coronaviruses are a family of viruses which can cause disease in humans and animals, with several coronaviruses capable of causing epidemics and pandemics. These coronaviruses have a similar life cycle which involves the use of a viral protease for the proteolytic processing of translated polyproteins into functional viral proteins. Inhibiting these viral proteases with protease inhibitors presents a promising therapeutic strategy for combating viral infections in both humans and animals (7). High throughput screening with a robust and repeatable enzymatic assay is a commonly used approach to identify and characterize viral protease inhibitors, however the methods used in many publications lack this robustness and reproducibility. By performing a thorough biophysical and enzymatic analysis of the SARS-CoV-2 Main protease with various substrates we identified optimal conditions for performing protease assays. Guided by our assessment of these substrates, we develop a substrate with improved performance in a high throughput screening assay due to incorporation of a brighter fluorophore which is less susceptible to interference and false positives than existing substrates.

The use of protease inhibitors has been successfully employed to treat viral diseases including hepatitis C virus (HCV) and human immunodeficiency virus (HIV). Since the emergence of SARS-CoV-2 in 2019, multiple protease inhibitors have been used for treatment of COVID-19(8). While the use of protease inhibitors can be a successful approach to treating viral infections, viruses can develop resistance to these protease inhibitors, in part due to the high mutation rate of these viruses. Therefore, there is a continued need for the development of new protease inhibitors to treat protease resistant viral strains. A recent study has shown that naturally occurring SARS-CoV-2 M^{pro} variants confer partial resistance to two clinically approved protease inhibitors for the treatment of COVID-19(9). The authors propose that these naturally occurring variants may enable the evolution of fully drug-resistant variants. Therefore, the development of additional M^{pro} inhibitors with different resistance profiles from existing drugs will be required to ensure the long-term effectiveness of protease inhibitors for treatment of COVID-19.

In addition to developing new protease inhibitors there are other approaches which may help ensure the long-term effectiveness of protease inhibitors for treatment of COVID-19. Treatment of HCV relies on the combined use of a viral protease inhibitor and a viral polymerase inhibitor(8). The combination of these two drugs greatly reduces the probability that a viral strain will evolve the mutations required to develop resistance to two drugs targeting distinct targets. A similar strategy referred to as combination antiretroviral therapy is used for treatment of HIV and has greatly improved clinical outcomes in HIV patients(8). Given the success of utilizing a combination of antiviral drugs with distinct mechanisms for the treatment of HCV and HIV, a similar strategy for the treatment of COVID-19 merits exploration. The use of drugs targeting SARS-CoV-2 viral RNA replication have been explored and are strong candidates for use in a combination therapy for treatment of COVID-19(10).

New and emerging zoonotic viruses pose a substantial threat to human health. While SARS-CoV-2 has caused the COVID-19 pandemic, closely related SARS-CoV has caused disease outbreak in the past and MERS-CoV continues to cause fatal disease in humans. Over the past 25 years many other viruses have emerged or have spread rapidly such as West Nile virus which was first introduced into north America in 1999 and continues to pose a public health threat. Zika virus has caused multiple outbreaks across the Americas, Asia, and Africa with an epidemic occurring in Brazil in 2015. Zaire ebolavirus caused an epidemic in west Africa in 2014 and an outbreak in 2018 resulting in more than 11,500 deaths. More recently a global outbreak of Monkeypox virus began in 2022(11). Given this recent history of viral outbreaks and epidemics, solutions to limit or prevent future viral outbreaks are needed.

Many of the scientific advancements leading to the development of protease inhibitors and vaccines against COVID-19 relied of past scientific discoveries made from studying closely related coronaviruses. For example, a major breakthrough in the development of a stabilized prefusion SARS-CoV-2 spike protein for vaccine development utilized two proline substitutions in the S2 subunit which had previously been shown to stabilize the prefusion spike proteins of MERS-CoV and SARS-CoV(12). Similarly, much of the work done to recombinantly express SARS-CoV-2 M^{pro} for enzymatic and crystallography studies relied on methods developed for SARS-CoV M^{pro} expression(13). Furthermore, the first reported crystal structure and protease inhibitors developed to target SARS-CoV-2 M^{pro} were based on previous research aimed at developing broad spectrum protease inhibitors effective against a range of closely related viral proteases(14). These examples show that fundamental research into viruses which do not pose an immediate public heath threat has proved incredibly valuable in handling the COVID-19 pandemic. Therefore, expanding this fundamental research on viruses that pose a potential future risk to human health should be explored further.

6.4 Targeting the heparin co-receptor for the disruption of Netrin-1 UNC5B interaction: Conclusions and future directions

Netrin-1 is a cell signaling molecule which interacts with the cell surface receptor UNC5B to regulate critical processes such as axon guidance, cell survival and cancer progression (15). Despite the importance of this interaction in physiological and pathological processes, the understanding of the nature of the interaction between Netrin-1 and UNC5B is limited. In the final chapter of this work, we used a biophysical approach to explore the Netrin-1 UNC5B interaction in more detail. We demonstrate that the Netrin-1 UNC5B interaction requires a heparin co-receptor for complex formation. We then developed peptides designed to mimic Netrin-1 structure and block the Netrin-1 UNC5B interaction by binding to the heparin co-receptor. Given that disruption of the Netrin-1 UNC5B interaction has been shown to be a feasible therapeutic strategy for treatment of certain cancers, there is potential that these inhibitory peptides may also serve as a lead molecule for the development of new therapeutics for cancer treatment. The finding that the Netrin-1 UNC5B interaction also requires a heparin co-receptor also has significant implications for advancing the understanding of Netrin-1 signaling.

Glycosaminoglycans like heparin are well known to play a role in the modulation of cell signaling processes. This modulation occurs through multiple mechanisms, including serving as a co-receptor for ligand-receptor binding, inducing ligand or receptor multimerization, and facilitating ligand immobilization and gradient formation (16). Previous studies have shown that Netrin-1 DCC signaling requires heparan sulfate, and the disruption of Netrin-1 heparin binding interferes with the Netrin-1 signaling ability in *C. elegans* (17–19). In showing that heparin serves as a co-receptor for Netrin-1 UNC5B complex formation indicates that the effect of heparin expression on Netrin-1 UNC5B signaling should be examined in detail. This examination may lead to a better understanding of the way Netrin-1 cell signaling processes are modulated.

Glycosaminoglycans including heparan sulfate are known to play a role in cancers(20). As an example, in multiple myeloma, studies have demonstrated that syndecan-1 (SDC1) interacts with hepatocyte growth factor (HGF) through its heparan sulfate (HS) chains. This interaction activates signaling cascades which are crucial for cell proliferation and survival(21). In addition, the activation of the Wnt/β-catenin cascade in multiple myeloma is also promoted by HS chains of SDC1, leading to cancer cells proliferation(22). Based on the clearly defined role of heparin in NET1 function as described here and in recent studies, it is worthwhile to investigate the role of heparan sulfate in NET1-related cancers. Furthermore, our in-vitro demonstration of the use of heparin-targeting peptides to disrupt the NET1-UNC5 interaction, suggests that targeting heparin's role in NET1 function could be a viable therapeutic approach.

In this work we demonstrate that inhibition of the Netrin-1 UNC5B interaction in vitro can be accomplished by targeting the heparin co-receptor. The use of heparin binding peptides to disrupt heparin mediated cell signaling processes has been used to target other cell signaling pathways. For example, peptides derived from the heparin binding regions of chemokines have been used to reduce neutrophile adhesion and migration during inflammation (23–25). The use of heparin mimetics that interfere with heparin mediated cell signaling processes is however a more commonly used approach to disrupt heparin mediated cell signaling processes (26). Future research aimed at understanding the sulfation patten requirements for Netrin-1 heparin binding and heparin mediated Netrin-1 UNC5B complex formation should be performed. Performing these studies could aid in determining if heparin mimetics are capable of disrupting and modulating the Netrin-1 UNC5B binding interaction and could produce lead molecules for future inhibitor development.

The most prevalent example of the use of heparin or heparin mimetics for therapeutic purposes is the use of heparin as an anticoagulant. Heparins function as an anticoagulant is a result of its role in binding to and activating antithrombin, which then inactivates proteases involved in the coagulation cascade(27). Through studying the interaction between antithrombin and heparin using biophysical and structural methods, it was identified that antithrombin binds to a pentasaccharide sequence of heparin with a specific sulfation pattern(28). Antithrombin binding to this pentasaccharide inactivates factor Xa of the coagulation cascade, but not thrombin. In order to inactivate thrombin, antithrombin must bind to an 18 saccharide unit of heparin enabling the formation of a ternary complex between thrombin, antithrombin and heparin(29). This understanding of the heparin requirements for antithrombin activation enabled the development of heparin preparations with specific anticoagulant effects and reduced side effects(30). Similarly, a better understanding of the structural features of heparin with respect to their effect on the Netrin-1 UNC5B interaction would serve to advance the understanding of the Netrin-1 UNC5B heparin interaction and could enable the development of heparin-based therapeutics targeting this interaction.

6.5 References

1. Kulczyk, A. W., McKee, K. K., Zhang, X., Bizukojc, I., Yu, Y. Q., and Yurchenco, P. D. (2023) Cryo-EM reveals the molecular basis of laminin polymerization and LN-lamininopathies. *Nat Commun*. **14**, 317
2. Xu, K., Wu, Z., Renier, N., Antipenko, A., Tzvetkova-Robev, D., Xu, Y., Minchenka, M., Nardi-Dei, V., Rajashankar, K. R., Himanen, J., Tessier-Lavigne, M., and Nikolov, D. B. (2014) Structures of netrin-1 bound to two receptors provide insight into its axon guidance mechanism. *Science*. **344**, 1275–1279
3. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589
4. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. **373**, 871–876
5. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat Methods*. **19**, 679–682
6. Hekkelman, M. L., de Vries, I., Joosten, R. P., and Perrakis, A. (2023) AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat Methods*. **20**, 205–213
7. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol*. **19**, 155–170
8. Shyr, Z. A., Cheng, Y.-S., Lo, D. C., and Zheng, W. (2021) Drug combination therapy for emerging viral diseases. *Drug Discov Today*. **26**, 2367–2376
9. Moghadasi, S. A., Heilmann, E., Khalil, A. M., Nnabuife, C., Kearns, F. L., Ye, C., Moraes, S. N., Costacurta, F., Esler, M. A., Aihara, H., von Laer, D., Martinez-Sobrido, L., Palzkill, T., Amaro, R. E., and Harris, R. S. (2023) Transmissible SARS-CoV-2 variants with resistance to clinical protease inhibitors. *Science Advances*. **9**, eade8778
10. Wang, X., Sacramento, C. Q., Jockusch, S., Chaves, O. A., Tao, C., Fintelman-Rodrigues, N., Chien, M., Temerozo, J. R., Li, X., Kumar, S., Xie, W., Patel, D. J., Meyer, C., Garzia, A., Tuschl, T., Bozza, P. T., Russo, J. J., Souza, T. M. L., and Ju, J. (2022) Combination of antiviral drugs inhibits SARS-CoV-2 polymerase and exonuclease and demonstrates COVID-19 therapeutic potential in viral cell culture. *Commun Biol*. **5**, 1–14
11. Keusch, G. T., Amuasi, J. H., Anderson, D. E., Daszak, P., Eckerle, I., Field, H., Koopmans, M., Lam, S. K., Das Neves, C. G., Peiris, M., Perlman, S., Wacharapluesadee, S., Yadana, S., and Saif, L. (2022) Pandemic origins and a One Health approach to preparedness and prevention: Solutions based on SARS-CoV-2 and other RNA viruses. *Proceedings of the National Academy of Sciences*. **119**, e2202871119
12. Hsieh, C.-L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H.-C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C.-W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., Lavinder, J. J.,

Ippolito, G. C., Maynard, J. A., Finkelstein, I. J., and McLellan, J. S. (2020) Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*. **369**, 1501–1505

13. Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., Bartlam, M., and Rao, Z. (2007) Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *Journal of Molecular Biology*. **366**, 965–975
14. Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., von Brunn, A., Leyssen, P., Lanko, K., Neyts, J., de Wilde, A., Snijder, E. J., Liu, H., and Hilgenfeld, R. (2020) α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **63**, 4562–4578
15. Boyer, N. P. & Gupton, S. L. (2018) Revisiting Netrin-1: one who guides (axons). *Frontiers in Cellular Neuroscience* **12**, 221
16. Linhardt, R. J., and Toida, T. (2004) Role of Glycosaminoglycans in Cellular Communication. *Acc. Chem. Res.* **37**, 431–438
17. Matsumoto, Y., Irie, F., Inatani, M., Tessier-Lavigne, M., and Yamaguchi, Y. (2007) Netrin-1/DCC Signaling in Commissural Axon Guidance Requires Cell-Autonomous Expression of Heparan Sulfate. *J Neurosci*. **27**, 4342–4350
18. Meier, M., Gupta, M., Akgül, S., McDougall, M., Imhof, T., Nikodemus, D., Reuten, R., Moya-Torres, A., To, V., Ferens, F., Heide, F., Padilla-Meier, G. P., Kukura, P., Huang, W., Gerisch, B., Mörgelin, M., Poole, K., Antebi, A., Koch, M., and Stetefeld, J. (2023) The dynamic nature of netrin-1 and the structural basis for glycosaminoglycan fragment-induced filament formation. *Nat Commun*. **14**, 1226
19. Priest, J. M., Nichols, E. L., Smock, R. G., Hopkins, J. B., Mendoza, J. L., Meijers, R., Shen, K., and Özkan, E. (2024) Structural insights into the formation of repulsive netrin guidance complexes. *Science Advances*. **10**, eadj8083
20. Elgundi, Z., Papanicolaou, M., Major, G., Cox, T. R., Melrose, J., Whitelock, J. M., and Farrugia, B. L. (2020) Cancer Metastasis: The Role of the Extracellular Matrix and the Heparan Sulfate Proteoglycan Perlecan. *Front. Oncol.* 10.3389/fonc.2019.01482
21. Derksen, P. W. B., Keehnen, R. M. J., Evers, L. M., van Oers, M. H. J., Spaargaren, M., and Pals, S. T. (2002) Cell surface proteoglycan syndecan-1 mediates hepatocyte growth factor binding and promotes Met signaling in multiple myeloma. *Blood*. **99**, 1405–1410
22. Ren, Z., van Andel, H., de Lau, W., Hartholt, R. B., Maurice, M. M., Clevers, H., Kersten, M. J., Spaargaren, M., and Pals, S. T. (2018) Syndecan-1 promotes Wnt/β-catenin signaling in multiple myeloma by presenting Wnts and R-spondins. *Blood*. **131**, 982–994
23. Vanheule, V., Janssens, R., Boff, D., Kitic, N., Berghmans, N., Ronsse, I., Kungl, A. J., Amaral, F. A., Teixeira, M. M., Damme, J. V., Proost, P., and Mortier, A. (2015) The Positively Charged COOH-terminal Glycosaminoglycan-binding CXCL9(74–103) Peptide Inhibits CXCL8-induced Neutrophil Extravasation and Monosodium Urate Crystal-induced Gout in Mice *. *Journal of Biological Chemistry*. **290**, 21292–21304
24. McNaughton, E. F., Eustace, A. D., King, S., Sessions, R. B., Kay, A., Farris, M., Broadbridge, R., Kehoe, O., Kungl, A. J., and Middleton, J. (2018) Novel Anti-Inflammatory Peptides Based on Chemokine–Glycosaminoglycan Interactions Reduce Leukocyte Migration and Disease Severity in a Model of Rheumatoid Arthritis. *The Journal of Immunology*. **200**, 3201–3217
25. Martínez-Burgo, B., Cobb, S. L., Pohl, E., Kashanin, D., Paul, T., Kirby, J. A., Sheerin, N. S., and Ali, S. (2019) A C-terminal CXCL8 peptide based on chemokine–glycosaminoglycan interactions reduces neutrophil adhesion and migration during inflammation. *Immunology*. **157**, 173–184
26. Ghiselli, G. (2019) Heparin Binding Proteins as Therapeutic Target: An Historical Account and Current Trends. *Medicines*. **6**, 80

27. Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N., and Carrell, R. W. (1997) The anticoagulant activation of antithrombin by heparin. *Proceedings of the National Academy of Sciences*. **94**, 14683–14688
28. Thunberg, L., Bäckström, G., and Lindahl, U. (1982) Further characterization of the antithrombin-binding sequence in heparin. *Carbohydrate Research*. **100**, 393–410
29. Casu, B. (1985) Structure and Biological Activity of Heparin. in *Advances in Carbohydrate Chemistry and Biochemistry* (Tipson, R. S., and Horton, D. eds), pp. 51–134, Academic Press, **43**, 51–134
30. Petitou, M., Hérault, J.-P., Bernat, A., Driguez, P.-A., Duchaussoy, P., Lormeau, J.-C., and Herbert, J.-M. (1999) Synthesis of thrombin-inhibiting heparin mimetics without side effects. *Nature*. **398**, 417–422