The molecular basis for substrate recognition and gating in sugar transporters

Sarah McComas
The molecular basis for substrate recognition and gating in sugar transporters

Sarah McComas

Academic dissertation for the Degree of Doctor of Philosophy in Biochemistry at Stockholm University to be publicly defended on Friday 8 December 2023 at 14.30 in Vivi Täckholmssalen (Q211), NPQ-huset, Svante Arrhenius väg 20 and online via Zoom, public link is available at the department website.

Abstract
Sugar is a vital sustenance for most forms of life. For a cell to take up sugar, specialized transport proteins embedded into the membrane bilayer known as sugar porters, are required. Dysfunction of sugar porters is associated with some metabolic diseases, and their expression is upregulated in many cancers as they typically require more sugar than normal cells. Furthermore, sugar porters also play a role in the vitality of the malaria parasite.

The mechanism of sugar transport is known as a rocker-switch alternating access mechanism. Simplistically, sugar binds between two similar domains on the outside of a sugar transporter and the domains then move around the sugar, so the sugar is exposed to the inside. During this domain movement, protein mass will block the sugar binding site from both outside and inside, forming the occluded state which is essential to ensure no substrate leakage during transport. Despite this relatively simple model of transport, little is known about how different sugar porters display diverse substrate specificity, affinity, and turnover.

In the four papers making up this thesis, we structurally characterize missing pieces of the sugar transport cycle, identify how these states are connected with simulations, and assess factors contributing to sugar transport by functional assays. With simulations, we show how sugar catalyzes conformational change by interacting with the occluded state. We demonstrate our functional proteoliposome-based transport assay, which allows us to measure the effect of protein mutations, inhibitors, and lipid influences in sugar recognition and turnover. Characterization of the malaria parasite hexose transporter PfHT1 has allowed us to understand antimalarial inhibitor specificity against this protein which could have implications in combating the disease, as well as pharmacological control of sugar porters in general.

Keywords: membrane transport, sugar transporter, simulations, lipids, antimalarial drugs.

Stockholm 2023

http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-222119


Department of Biochemistry and Biophysics

Stockholm University, 106 91 Stockholm
THE MOLECULAR BASIS FOR SUBSTRATE RECOGNITION AND GATING IN SUGAR TRANSPORTERS

Sarah McComas
The molecular basis for substrate recognition and gating in sugar transporters

Sarah McComas
To my cherished colleagues, without whom this work would not be possible.
ABSTRACT

Sugar is a vital sustenance for most forms of life. For a cell to take up sugar, specialized transport proteins embedded into the membrane bilayer known as sugar porters, are required. Dysfunction of sugar porters is associated with some metabolic diseases, and their expression is upregulated in many cancers as they typically require more sugar than normal cells. Furthermore, sugar porters also play a role in the vitality of the malaria parasite.

The mechanism of sugar transport is known as a rocker-switch alternating access mechanism. Simplistically, sugar binds between two similar domains on the outside of a sugar transporter and the domains then move around the sugar, so the sugar is exposed to the inside. During this domain movement, protein mass will block the sugar binding site from both outside and inside, forming the occluded state which is essential to ensure no substrate leakage during transport. Despite this relatively simple model of transport, little is known about how different sugar porters display diverse substrate specificity, affinity, and turnover.

In the four papers making up this thesis, we structurally characterize missing pieces of the sugar transport cycle, identify how these states are connected with simulations, and assess factors contributing to sugar transport by functional assays. With simulations, we show how sugar catalyzes conformational change by interacting with the occluded state. We demonstrate our functional proteoliposome-based transport assay, which allows us to measure the effect of protein mutations, inhibitors, and lipid influences in sugar recognition and turnover. Characterization of the malaria parasite hexose transporter PfHT1 has allowed us to understand antimalarial inhibitor specificity against this protein which could have implications in combatting the disease, as well as pharmacological control of sugar porters in general.
SAMMANFATTNING


I de fyra artiklarna som presenteras i denna avhandling har jag karakteriserat transportcykeln hos sockertransportörer och har bidragit till förståelsen av vilka faktorer som styr den. I artikel III presenterades det metodiken som har lagt grunden till den laborativa delen av denna avhandling. Vi har visat att många sockertransportörer är känsliga för vilka lipider som omger dem, och att metodiken i artikel III är tillräcklig känslig för att användas till analyser av mutationer, samt studera proteinernas kinetik. Artikel I presenterades strukturen av sockertransportören från malariaparasiten Plasmodium falciparum (PfHT1), vilken också lagt grunden till arbetet under andra delar av avhandlingen. Strukturen var den första av sockertransportörsproteinerna i det ockluderade stillståndet, vilket gjorde att vi numera har strukturell information från alla de fem huvudsakliga stillstånden under transportcykeln.

I artikel II demonstrerades att det ockluderade stillståndet är nyckeln till kopplingen mellan sockerbindningen och att andra stillstånd nås. Vi tror att detta stillstånd motsvarar ett övergångstillstånd, precis som i enzymer. Skillnaden är att i en sockertransportör, är det entropin i ett proteintillstånd som gör att energibarriären är hög för byte av stillstånd, men barriären sänks när av rätt sockerart binder. Det finns fortfarande mycket kvar att utforska kring
hur enskilda sockerarter specifikt känns igen, såsom t.ex. om inducerad passformsteori gäller.

I **artikel IV** utvecklades idén om kopplingen mellan inhibition av transportörproteinet, där transporten förhindras av ett ämne, som antingen binder till sockerbindningsstället eller andra delar av proteinet. Malaria är fortfarande en dödlig och förödande sjukdom i världen och sockertransportören i denna parasit har identifierats som ett potentiell läkemedelsmål för att bekämpa sjukdomen, trots stora likheter i bindningsstället. I denna artikel har jag visat att det finns ämnen som inhiberar specifikt PfHT₁, och inte de mänskliga motsvarigheterna. Det sker på grund av interaktioner som inte är begränsade till bindningsstället, utan även via interaktioner till andra delar som är unika för PfHT₁, vilka finns en bit från bindningsstället. Nu börjar vi förstå att det finns många delar av proteinet som tillsammans bidrar likt ett nätverk och reglerar transport av socker, samt styr inhibition. Denna information kan användas för att utveckla specifika läkemedel för behandling av sjukdomar.
LIST OF PAPERS

PAPER I

The molecular basis for sugar import in malaria parasites.
Nature 578, no. 7794. (2020)

PAPER II

Determinants of sugar-induced influx in the mammalian fructose transporter GLUT5.
McComas, S. E., Reichenbach T., Mitrovic D., Alleva C., Bonaccorsi M., Delemotte L., Drew D.
Elife 12, e84808. (2023)

PAPER III

Establishing mammalian GLUT kinetics and lipid composition influences in a reconstituted-liposome system.
Suades A., Qureshi A.A., McComas S.E., Coinçon M., Rudling A., Chatzikyriakidou Y., Landreh M., Carlsson J., Drew D.
Nature Communications 10;14, 4070. (2023)

PAPER IV

Probing sugar import mechanism via multiple specific inhibitors of the malaria parasite hexose transporter.
Manuscript.

* These authors contributed equally.
† Final author list subject to change
PERSONAL CONTRIBUTIONS TO MANUSCRIPTS

PAPER I

I assisted in experiments for functional analysis of PfHT1, with protein expression, purification, and transport assays. I performed molecular dynamics simulations of PfHT1 and GLUT3, as well as analyzed the simulations. I contributed to writing and figure making of this manuscript.

PAPER II

I designed the research together with L.D. and D.D. I performed homology modeling, conventional molecular dynamics (MD) simulations, targeted MD simulations, enhanced sampling MD simulations, and free energy calculations of GLUT5. I performed the analysis of simulations, and published the code for this analysis. I (together with T.R.) performed the biochemical assays of GLUT5, including protein mutagenesis, expression, purification, and transport assays. I led the manuscript and figure preparation of this work. *

PAPER III

I performed protein expression and purification of GLUT5 for the GFP-TS assay. I performed the GFP-TS assay for both lipid extracts as well as pure lipids. I assisted in the mass spectrometry measurements. I contributed to writing and figure making of this manuscript.

PAPER IV

I designed the research together with A.S., A.G., and D.D. I performed molecular dynamics simulations of PfHT1 with MMV (all conditions), and PfHT1 with C3361 (all conditions). I performed the docking studies of WU-1 with PfHT1. I contributed to protein functional and structural measurements with protein expression and purification. I led the manuscript and figure preparation of this work. *

* First author works.
### ABBREVIATIONS

#### Proteins and Biology
- **MFS** Major Facilitator Superfamily
- **SP** Sugar Porter
- **SLC** Solute Carrier
- **GLUT** Glucose Transporter
- **HXT** Hexose Transporter
- **PfHT1** Plasmodium Falciparum Hexose Transporter 1
- **TM** Transmembrane Helices
- **IC** Intracellular
- **EC** Extracellular
- **ICH** Intracellular Helices
- **RBC** Red Blood Cell

#### Biochemistry
- **GFP** Green Fluorescent Protein
- **DDM** n-Dodecyl-β-D-maltoside
- **GFP-TS** GFP-Thermal Stability assay
- **SEC** Size Exclusion chromatography
- **FSEC** Fluorescence Size Exclusion Chromatography
- **Cryo-EM** Cryogenic electron microscopy

#### Computational
- **MSA** Multiple Sequence Alignment
- **TCDB** Transporter Classification Database
- **HMM** Hidden Markov Model
- **MD** Molecular Dynamics (simulations)
- **RMSD** Root Mean Squared Deviation
- **PCA** Principal Component Analysis
- **CGMD** Coarse Grained Molecular Dynamics
- **CV** Collective Variable
- **HPC** High Performance Computing
- **TMD** Targeted Molecular Dynamics
CONTENTS

Abstract i
Sammanfattning iii
List of papers v
Personal contributions to manuscripts vii
Abbreviations ix

1 Membranes and Membrane proteins 1
   1.1 Biological membranes 1
   1.2 Membrane proteins 2

2 Membrane Transporters 3
   2.1 Types of membrane transport 3
   2.2 The Alternating Access Mechanism 4
   2.3 Subclassifications of alternating access mechanism 5
   2.4 The importance of the occluded state in alternating access 6

3 Functional classification of transporters 9
   3.1 The Major Facilitator Superfamily 9
   3.2 Solute carrier family 10

4 Sugar Porters 11
   4.1 Motifs in sugar porters 11
   4.2 Sugar recognition and binding 12
   4.3 Conformational change in sugar porters 13
   4.4 Asymmetry in sugar porters 14
   4.5 Sugar porters and their lipid environment 15

5 The mammalian fructose transporter GLUT5 17
   5.1 GLUT5 implications in disease 18
   5.2 Substrate specificity of GLUT5 19

6 The Plasmodium falciparum hexose transporter PfHT1 21
   6.1 PfHT1 implications in disease 22
   6.2 Substrate specificity of PfHT1 22

xi
# Contents

7 Biochemical assessment of sugar porters  
7.1 Protein overexpression and purification  
7.2 Proteoliposome-based transport assay  
7.3 GFP-based thermal stability assay  
7.4 Structural biology of sugar porters  

8 Computational experiments of sugar porters  
8.1 Multiple sequence alignments  
8.2 Molecular docking  
8.3 Molecular dynamics simulations  
8.4 Enhanced sampling molecular dynamics  

Summary of papers  
Paper I  
Paper II  
Paper III  
Paper IV  

Conclusion and future perspectives  

Acknowledgements
MEMBRANES AND MEMBRANE PROTEINS

The work in this thesis aims to understand and characterize sugar transport proteins which are small, dynamic proteins that translocate sugar across a cell membrane, a vital process to all organisms, regardless of their phyla. As these proteins operate in a membrane bilayer, their function is sensitive to the lipidic environment around them, which in itself is also dynamic and is diverse between different tissues or organisms.

1.1 BIOLOGICAL MEMBRANES

All life requires some form of compartmentalization; at minimum in the form of the cell. The barrier between these and their surroundings was long thought to be a rigid barrier like a wall. Today, we realize that biological membranes, i.e. lipid bilayers, are highly dynamic and controlled lipid barriers decorated with membrane proteins [1]. In eukaryotes, these bilayers typically consist of two layers of interfacing hydrocarbon lipids, with polar head groups on either side, as seen in Figure 1.1 [2].

There is a remarkable amount of diversity in these bilayers. For instance, archaea can have a monolayer of lipids instead, which are linked more strongly to their head groups than bacteria or eukaryotes, a mechanism which has helped in their adaptation to extreme conditions [3,4]. On a smaller scale the bilayer is also diverse; the bilayer of the Golgi apparatus contains shorter hydrophobic tails

Figure 1.1: A typical phospholipid bilayer.
than that of the cell, for example [5]. This diversity is even seen within a single bilayer; lipids diffuse around the bilayer and can co-localize to different regions of the cell membrane in lipid rafts which seems to play a role in many cell processes [6–9].

1.2 Membrane Proteins

Membrane proteins are proteins embedded in the lipid bilayer, responsible for some form of connection between the outside of a bilayer to the inside. About 30% of genes in a genome encode for membrane proteins [10]. These proteins perform a large variety of functions and are usually categorized as so; they can partake in signaling, cell adhesion, enzyme activities, or molecule transport, to name a few [11]. Over 50% of small-molecule and biological drugs are targeting membrane proteins, and many membrane proteins are ubiquitous across all walks of life, highlighting their importance for organism homeostasis [10, 12, 13].

Despite their importance, membrane proteins are perpetually undercharacterized. Thus far, of the nearly 90,000 unique protein sequences which have been structurally characterized, only 1,600 are membrane proteins [14, 15]. This is not out of neglect: membrane proteins are challenging to work with due to many factors, not the least of which being their hydrophobic nature and need for a lipid environment. Many techniques have been developed to optimize the extraction of these proteins from their native environment with use of detergents for use in functional or structural work [16, 17]. In recent years the appreciation for the symbiosis between membrane proteins and their lipid environments has increased, especially with perspective on the effect of lipids on membrane proteins [18–26], with an increase in the use of lipid mimetic systems for structural studies of the proteins [21, 27], as well as improved computer simulation techniques for studying membrane protein-lipid interactions [23, 25, 26].
MEMBRANE TRANSPORTERS

Comprising roughly 25% of all membrane proteins in human [11], transport proteins are responsible for carrying in anything which cannot pass through the lipid bilayer by diffusion, such as sugars, ions, or amino acids. Transport proteins can be further categorized into either channels or transporters. Channels are selective pores embedded into the membrane; they typically open in response to a stimulus so that solutes can diffuse down their electrochemical gradient until the channel pore constricts [28]. This diffusion of substrate is the distinguishing factor between channels and transporters, as transporters instead are constantly undergoing conformational change to move substrate across the membrane [29]. In this way, transporters can be considered as enzymes, and their functional activity described by Michaelis-Menten kinetics [30]. Furthermore, transporters depend on specific substrate(s) binding to undergo a translocation event.

2.1 TYPES OF MEMBRANE TRANSPORT

Whereas channels will only allow for diffusion down a substrate’s electrochemical gradient, transporters can act as either passive transporters or active transporters [31]. Passive transporters allow for the movement of substrate down its electrochemical gradient via conformational changes of the transporter. This is distinct from diffusion through channels as translocation must be coupled to the transporter recognizing substrate and undergoing a conformational change, which means their rates are often slower than channels [32].

Active transporters are capable of pumping solutes against their concentration gradient [29,31]. Active transport can be further split into two categories: primary active transport and secondary active transport. Primary active transport typically use the hydrolysis of ATP to pump a solute against its electrochemical gradient, whereas secondary active transport couple this uphill pumping
with the movement (either via symport or antiport) of another solute (often an ion) down its electrochemical gradient. In symport, the solute providing the energy (the solute moving down its electrochemical gradient) and the solute which is being pumped against its gradient are being transported in the same direction, and in antiport they are being transported in opposite directions [29, 31]. Figure 2.1 summarizes these mechanisms.

2.2 THE ALTERNATING ACCESS MECHANISM

To partake in substrate translocation, all transporters must go through the alternating access mechanism where the substrate binding site of the protein is alternatingly revealed to solvent on different sides of the membrane. Often, transporters are organized into two domains which move against each other to perform alternating access (Figure 2.2). This mechanism was
first proposed nearly a century ago, but this model has been reinforced by the structural characterization of a variety of transporters [20,22,29,33–35]. This distinguishes the transporter from a channel as this mechanism prevents the substrate from moving freely through the protein and is thus one factor which allows transporters to be able to pump solutes against their electrochemical gradient. In a simplified model, as seen in Figure 2.2, the transporter undergoes alternating access by closure of the extracellular gating elements, which is coupled to the opening of the intracellular gating elements (these gating elements will be discussed in chapter 4). The most important feature of this is that the protein must never have both gating elements open to the solvent, preventing uncoupling of substrate binding to conformational change. Because of this, there is an intermediate protein state in the substrate translocation cycle where both gating elements are shut, known as the occluded state (middle, Figure 2.2) [29].

2.3 SUBCLASSIFICATIONS OF ALTERNATING ACCESS MECHANISM

Various passive and secondary active transporters will alternate access to the substrate binding site by different means. This introduces diversity in protein kinetics, substrate binding modes, and dependency on oligomerization, to name a few. In the simplest mode, the rocker-switch mechanism, two symmetric protein domains will rock against each other (Figure 2.3); the substrate stays roughly in the same vertical position during alternating access. The bacterial SWEETs and the mammalian GLUTs are both examples of proteins which undergo the rocker-switch mechanism [29]. There are several transmembrane helices in the transporter which dictate the opening and closure of the gates involved in this mechanism, which will be discussed in chapter 4.

The two other mechanisms in which transporters partake in alternating access involve asymmetry in two structurally distinct domains, as seen in
The rocking bundle mechanism appears similar to the rocker switch mechanism in that the substrate remains in a fixed position during protein conformational change, but one of the domains partakes in the majority of rearrangements against a more labile domain. The bacterial amino acid transporter LeuT exhibits a classic example of the rocking bundle mechanism, which was supported by crystal structures of the protein [36]. The other asymmetrical mechanism is known as the elevator mechanism, with even more dramatic asymmetry than rocking bundle mechanism. For example, this can be seen in glutamate transporters or sodium proton antiporters; one domain of the protein (the transport domain) performs all substrate translocation, moving against a scaffold domain [20, 22, 29, 37]. As such, the substrate is carried across the membrane by one of the domains known as the transport domain [29, 38]. To date, almost all known elevator proteins are oligomers, likely because this scaffold domain is required to act as an anchor to support independent movement of the transport domains [22].

### 2.4 The Importance of the Occluded State in Alternating Access

Like enzymes, transporters are thought to form a state of higher energy and higher substrate affinity in their conformational cycle, analogous to the transition state [39]. Whereas in enzymes the transition state is formed as the protein provides a catalytic conversion between substrate and product, the transition state in transporters is formed as the protein undergoes the occluded state of alternating access, facilitated by binding of substrate [13, 39–41]. In this way, substrate binding energy is used to drive a conformational change in the protein, rather than the conversion of substrate to product. Affinity for substrate is likely to be highest at the transition state (increasing enthalpy) to attempt an energetic trade-off with the entropy penalty of forming this state, as dictated by the Gibbs free energy equation $\Delta G = \Delta H - T\Delta S$. A simplified diagram for this is shown in Figure 2.4.

Formation of the occluded state is the most critical state in the alternating access mechanism. In the glucose transporters, the resetting step (i.e. moving to outward open from inward open) is slower than the glucose import process, likely due to this energetic barrier of forming the occluded state without substrate [39, 42–44]. As will be discussed in this thesis, particularly papers I and II, understanding the occluded state is likely the key to understanding how sugar binding is coupled to conformational change of
2.4 The importance of the occluded state in alternating access

A protein, as the entropic penalty will be too high if sugar is not properly coordinated.

Figure 2.4: A simplified energy diagram. The energy barrier to reach the occluded state is lowered by substrate binding.
FUNCTIONAL CLASSIFICATION OF TRANSPORTERS

Transporters can be grouped into families or superfamilies by protein sequence (phylogeny) or overall function (e.g. what substrate is being translocated). The development of high-throughput sequencing and computational biology tools have helped to establish shared connections between proteins in their sequence elements, ancestry, or topology. This information is put into databases where proteins can then be classified [10,45,46]. For example, the transporter classification database (TCDB) [46] maintains manual annotation of the proteins which allows for accurate and up-to-date functional groupings of proteins. Automatic annotation tools are also useful as the databases are large; for example, Pfam / Interpro uses hidden Markov model (HMM) based multiple sequence alignments to classify protein families and domains. This allows for protein domain classification as well as detection of distant homologs to a protein family [10,47].

3.1 THE MAJOR FACILITATOR SUPERFAMILY

The Major Facilitator Superfamily (MFS) is one of the largest families of transport proteins (currently containing 2 million different, putative proteins [47]) which partake in either uniport or secondary-active transport, and share about 12% sequence identity, at minimum to one another [13,46,48,49]. The transporters in this superfamily are ancient and found in all three domains of life [46]. MFS members share a common topology, with a cytoplasmic N- and C- terminus and 12 transmembrane (TM) helices that make up two 6-TM bundles connected by a long, cytoplasmic loop (Figure 3.1) [13,48]. In the MFS superfamily, transporters can be subclassified into smaller families (currently 17) mostly based on their function [49].

These domains arise from inverted repeats, meaning that TM helices 1-
3, 4-6, 7-9, and 10-12 are all repeats of each other, likely evolved from simpler 3-TM transporter proteins [13,50,51]. During alternating access, cavity closing contacts are typically formed between TM1-2 and TM7-8 on the outside, and TM4-5 and TM10-11 on the inside [13]. These cavity-closing helices often undergo local conformational changes to coordinate substrate and are referred to as gating helices. To enable one half of the helix to move somewhat independently from the other half, these TMs often contain helix-breaking residues, splitting the TMs into two half-helices 3.1 [13,50].

![Figure 3.1: A typical MFS fold, with inverted repeat symmetry. TMs 1, 4, 7, and 10 are shown as half-helices.](image)

### 3.2 SOLUTE CARRIER FAMILY

The solute carrier family (SLC) is a mammalian-specific classification of small molecule transporter proteins. In human, there are roughly 450 SLCs that can be grouped into 65 subfamilies sharing 25% sequence identity to each other [13,29,52]. The sequence identity is the main means of clustering each family member, rather than function although these often go hand-in-hand. It is important to note that not all SLC proteins belong to MFS, or vice versa. SLC and MFS serve as separate means of classification of proteins, and SLC pertains only to mammals. Currently, 13 of the 65 SLC families have the MFS transporter fold [52].
The Sugar Porter (SP) family is a group of proteins which partake in the transport of different sugars, typically from the outside of the cell to the inside, via either uniport or symport. The family currently contains 133 annotated proteins in the TCDB, and 125,000 putative proteins \[13,46,48,49\]. The SP family belongs to a subfamily of the MFS superfamily, and the mammalian SP proteins also belong to the SLC2 subfamily of solute carriers \[13,52,53\]. The 14 isoforms of the mammalian sugar porters are known either as SLC2A or GLUT (glucose transporter)1 - 14, and are generally responsible for organism-wide glucose homeostasis. The 14 GLUTs vary in tissue localization as well as substrate specificity (some GLUTs do not recognize glucose) and turnover rates \[53\]. Other frequently-studied sugar porters relevant to this work include the *Escherichia coli* proton-coupled xylose transporter XylE \[35,54\], the yeast hexose transporters (HXTs) \[55–57\], and the *Plasmodium falciparum* hexose transporter PfHT1 \[58\].

### 4.1 Motifs in Sugar Porters

Sugar porters display the MFS fold with an addition of 4-5 small helices outside the intracellular (IC) region of the protein (Figure 4.1), known as the intracellular helices (ICH) \[13,34\]. SP proteins contain the sugar porter motif, a feature which was in fact historically codified a protein as a sugar porter \[59\]. The sugar porter motif is a network of charged residues located in intracellular loops between transmembrane helices. They are capable of forming salt-bridges \[59\], often with IC helices as seen in Figure 4.1. Similarly to discussed in chapter 3, these salt bridge residues also appear to be related by inverted repeat symmetry.
4.2 SUGAR RECOGNITION AND BINDING

In addition to conserved ICH1-5 and salt bridges, many thus-far characterized sugar porters share an overall conserved substrate binding site, remarkably so given their functional divergence [13, 34, 35, 60]. This functional diversity is often specific; i.e. most sugar porters can only transport one type of sugar. Even further puzzling is how the proteins can maintain this diversity while also displaying a generally low affinity (low mM range) for their sugars. There have been several studies thus far in aiming to identify what determines this, particularly for the yeast HXT transporters [55–57] and some of the GLUTs [61, 62]. Although these studies have identified several binding site residues critical for substrate recognition, completely altering preference of substrate has proven challenging.

Currently, there are 30 published structures of sugar porters, 10 of which are distinct proteins [49]. Several of these structures show substrate interacting with residues in the binding site which are overall identical in sugar porters.
4.3 Conformational Change in Sugar Porters

Sugar porters all undergo the rocker-switch mechanism of alternating access (first addressed in Figure 2.3). During alternating access, four main transmembrane helices partake in the majority of the occlusion of the substrate binding site (Figure 4.3), known as the extracellular (EC) and intracellular (IC) gates. The half-helices of TM1 and TM7 (TM1b and TM7b) form the EC gate, and the half helices of TM4 and TM10 (TM4b and TM10b) form the IC gate. As discussed in chapter 3, these helices are related as inverted repeats.
In alternating access, sugar porters undergo 5 major states upon substrate binding (Figure 4.4). With the publishing of paper I, all 5 states have now been structurally characterized by at least one sugar porter protein [34, 35, 60, 64–66]. The substrate typically binds the protein in the outward open state. This state is the most energetically favorable for sugar porters in the absence of substrate (as seen in paper II), likely because of the formation of intracellular salt bridges; perturbation of these arrest the transporter in an inward-facing conformation [67]. Upon substrate binding, the extracellular gate begins to close as TM7b approaches TM1b in the outward occluded state. Once the extracellular gate is shut, the occluded state has been reached. At this point, TM1ob will move away from TM4b in the inward occluded state. Lastly, once the intracellular gate is fully open, the intracellular salt bridges will break apart and there is displacement of the IC helices in the inward open state, where sugar will be released into the cell. The protein then must go through the same steps in reverse in the absence of substrate. As mentioned in chapter 2, this is the rate-limiting step of the reaction, approximately 100-fold slower than the sugar import step [13, 42, 43, 53]. Papers I and II in this thesis both have contributed to the characterization of this transport cycle.

### 4.4 Asymmetry in sugar porters

In addition to asymmetry in the sugar binding site (that the C terminal domain contains more sugar binding residues than the N terminal domain), sugar porters also exhibit asymmetry in their binding direction. Sugar porters will typically bind substrate in an outward facing conformation (Figure 4.4) to then bring sugar from the outside to the inside of the cell. With an inverse electrochemical gradient, sugar porters work in the reverse direction, transporting from inside to outside of cell. However, these proteins typically have a weaker affinity for substrate when it binds from the inside compared...
Figure 4.4: The substrate transport cycle in sugar porters. Gating TMs 1&4 (light blue) and 7&10 (dark blue) are visible. The orange circle represents sugar substrate.

to outside [13,53,68]. When sugar binds from the outside, sugar-coordinating residues on TM7b will induce extracellular gate kinking, encouraging the protein to enter the occluded state, as explored in paper II. However, when sugar binds from the inside, there are no equivalent residues on TM10b, and furthermore TM10b does not form the same kink angle as TM7b when the intracellular gate is closed. It seems therefore likely that the coupling of sugar binding to conformational change is responsible for binding asymmetry in these proteins [13].

4.5 SUGAR PORTERS AND THEIR LIPID ENVIRONMENT

Both specific lipid-protein interactions and membrane physical properties (such as fluidity, compression, and thickness) can play a role in membrane protein or transporter function [5,23,69]. There are three main categories of lipid-protein interfaces, listed here in order of relative diffusion speed: bulk lipids which diffuse the quickest and do not interact with the protein, annular lipids which form a shell around the membrane protein and do not interact specifically, and nonannular lipids which tend to be bound to a protein for long periods of time as they interact specifically [24]. Nonannular lipids can be considered almost as ligands to the protein. There are plenty of examples of nonannular lipids being critical to MFS transporter
function [20, 22, 70] but to date, sugar porters do not seem to be one of those proteins, as will be discussed in paper III.

This does not mean that sugar porters are not sensitive to their lipid environment. Hydrogen-deuterium exchange mass spectrometry experiments have showed that sugar porters’ conformational dynamics depends on their lipid surroundings, and that interaction with the IC salt bridge-forming residues seems to play a role in this [71]. Research, including work presented in paper III, has shown that sugar porters display different transport function when incorporated into liposomes of varied lipid composition and that the proteins are sensitive to lipid shape or tail saturation [69]. Diversity of lipid sensitivity within the sugar porter family (such as why XylE transports best in E. Coli lipids), or how lipids are implicated in disease phenotypes which involve GLUTs (such as the change in lipid metabolism/ profiles in obesity and diabetes) remain to be elucidated [72, 73]. Furthermore, the effect of lipids on sugar porter dimerization, which has been shown for other MFS transporters, is another avenue to explore [74].
THE MAMMALIAN FRUCTOSE TRANSPORTER GLUT5

GLUT5 is one of the 14 GLUT isoforms in mammals, and therefore belongs both to the SLC and MFS superfamilies [49]. Like most other GLUTs, it is a uniporter, but it transports fructose instead of glucose with a Km of roughly 10 mM [34, 53, 75, 76]. GLUT5 expression levels are highest in the small intestine, but modest expression levels can also be found in testes, skeletal muscle, and brain [77]. To date, there are two structures available of GLUT5 in the outward open and inward open conformations (see Figure 4.4) [34]. GLUT5 maintains the same features canonical for a sugar porter and MFS protein, with about a 30% sequence identity to the other GLUTs (Figure 5.1), and sharing a nearly conserved binding site with the other GLUTs. The work in papers II and III contribute to characterization of this protein.

Figure 5.1: A phylogenetic tree of the human GLUTs. Made in iTOL [78].
5.1 GLUT5 IMPLICATIONS IN DISEASE

Because of human’s increasing consumption of high-fructose corn syrup, there have been studies exploring the altered expression of GLUT5 in links to obesity, inflammatory bowel diseases, and fructose malabsorption. Although there is a correlation between these factors, the causative effect of altered GLUT5 expression has not yet been linked to these conditions [79,80].

On the other hand, it is well known that GLUT5 is overexpressed in many cancers [79,81], so much so that using radiolabeled fructose can be a means of measuring breast cancer progression [82]. One of the hallmarks of cancer is the ability to rapidly proliferate, which requires a dramatic increase in sugar metabolism [83]. This phenomenon is known as the Warburg effect, where the cancer cells switch to lactic acid fermentation to upkeep rapid ATP production and require more sugar for this metabolism [84]. Thus, there is a correlation between the overexpression of GLUT5 and the malignancy level of many cancers [79–81].

Figure 5.2: The GLUT5 sugar binding site (orange), overlaid on GLUT3 as in Figure 4.2.
5.2 SUBSTRATE SPECIFICITY OF GLUT5

It is unclear what makes GLUT5 specific to fructose, especially given the fact that the substrate binding site of GLUT5 is nearly identical to other GLUTs (Figure 5.2). In fact, glucose is not recognized by GLUT5, surprising given that XylE (a xylose specific sugar porter) is inhibited by the sugar [13, 34, 55]. Furthermore, GLUT7 which can potentially transport both glucose and fructose [62], shows a much higher sequence identity with GLUT5 (60% compared to 30%) and clusters together phylogenetically (Figure 5.1).

Although exact determinants of fructose specificity in GLUT5 have yet to be revealed, GLUT5 recognizes fructofuranose over fructopyranose (Figure 5.3), and epimers of fructose are not transported [85, 86]. Work presented in paper II indicates allosteric coupling between substrate binding and turnover in GLUT5, involving residues which are specific to the protein.

![Figure 5.3: β-D-fructose epimers.](image-url)
**THE PLASMODIUM FALCIPARUM HEXOSE TRANSPORTER PFHT₁**

PFHT₁ a sugar porter protein from the malaria parasite *Plasmodium falciparum*, containing the typical motifs and topology of SP and MFS proteins. It is distantly related to the human GLUTs, with about a 25% sequence identity to them. It is the only sugar import protein in the *Plasmodium* genome, and is primarily important during the parasite’s growth in the red blood cell (RBC), i.e. during the ring or trophozoite stage of the malaria infection (see Figure 6.1) [87,88]. At this stage, the parasite is taking over the cell and PFHT₁’s importance is clear as glucose consumption in infected RBC’s is 100-fold increased, and proteins are expressed by *Plasmodium* to dephosphorylate the glucose which has been metabolized by the RBC for import by PFHT₁ [89–92]. Furthermore, PFHT₁ can transport a variety of different hexoses [87,89] and to date, there are several published structures of PFHT₁ available [58]. **Papers I and IV** both contribute to characterization of this protein.

![Diagram](image)

**Figure 6.1:** The trophozoite stage of a *Plasmodium* infected RBC, where the vacuole will grow in the RBC, and PFHT₁ is the most expressed [88]. The EXT2 protein is a potential means of which the sugar enters from the parasite vacuole to the parasite membrane [91].
6.1 \textit{pfHT1} implications in disease

Even now, malaria remains a deadly infectious disease, with 600,000 deaths from 250,000,000 infections in 2022 alone [93]. This is caused by several species of \textit{Plasmodium} parasites, with the most deadly being \textit{Plasmodium falciparum} [93]. Artemisinin-based therapies to combat malaria have been popular and successful in recent years [94,95]. However, an increasing resistance of \textit{Plasmodium falciparum} to the drug is being reported in East Asia, with the potential of spread to Africa causing ‘major concern’ at the World Health Organization [93,96].

This is not the first time \textit{Plasmodium falciparum} has developed resistance to antimalarial drugs, where chloroquine resistance is also established [97]. Both of these drug therapies are imported into the parasite (Figure 6.1); here the parasite has been able to develop mechanisms for resistance by exportation or neutralization [96,97].

There is therefore a strong interest in targeting membrane proteins for antimalarial drugs, of which \textit{PfHT1} is a candidate being the only hexose transporter in the \textit{Plasmodium} genome and clearly exhibiting a critical role in the parasite growth [87–89,98]. In fact, there are several compounds which have adverse effects on parasite growth and target \textit{PfHT1} specifically; several of which will be discussed in the work presented in this thesis in \textit{paper IV} [58,98–101].

6.2 \textit{pfHT1} substrate specificity

Apart from its implications in antimalarial drug development, \textit{PfHT1} is an interesting model for substrate specificity as it can recognize and transport a variety of sugars [89]. As discussed in chapter 4, the substrate binding site of sugar porters is overall conserved, of which \textit{PfHT1} is no exception. Although it is yet to be understood how to alter substrate specificity of \textit{PfHT1}, the work presented in \textit{paper I} reveals that \textit{PfHT1} seems to be promiscuous to many hexoses based on its unique polar extracellular gate, which can shut more easily upon substrate binding.
Structural and functional assessment of membrane proteins remains a challenging and time-consuming process; one reason being that these studies often require large quantities of purified protein. Native quantities of these proteins are typically too low for any meaningful assessment and therefore require overexpression, which can lead to cytotoxicity and low protein yield \[102\]. Removing membrane proteins from their lipid environment for isolation in functional or structural assays is not always straightforward and can result in protein aggregation or unfolding \[16, 103, 104\]. Additionally, for characterization of a novel or unfamiliar membrane protein, extensive screens of conditions to yield a functional and well-folded protein for functional studies is often necessary \[16\]. Despite these challenges, structural and functional studies of GLUT5 and PfHT1 has contributed to great insights into these transporters’ function during the work shown in this thesis, relevant in all papers (paper I - IV).

### 7.1 Protein Overexpression and Purification

In the functional work presented in this thesis, PfHT1 and GLUT5 are expressed in *Saccharomyces cerevisiae*. *Saccharomyces cerevisiae* is an excellent tool for membrane protein overexpression as yeast participates in post-translational modification and has a comparable lipid bilayer to that of mammals \[105, 106\]. Furthermore, yeast is inexpensive, easy to upscale, and well-studied meaning that alterations to the organism are easier attainable; knock-out strains are useful in ensuring that the gene containing the protein of interest (POI) has been properly incorporated \[16, 107\]. In novel protein characterization, the screening of different constructs to optimize protein expression and dispersion is typically necessary; utilizing an expression system such as yeast which quickly yields results is ideal \[16, 107, 108\].
In the purification of sugar porters (or any membrane protein), detergent is needed to extract the protein from their membranes. As the property of the detergent chosen for a purification can yield different protein dispersity or purification yield, a screen of detergents is commonplace, especially for the more-sensitive eukaryotic proteins [104]. Using fusion partners such as green fluorescent protein (GFP), which is included in the polypeptide sequence of a POI, can aid membrane protein purification tremendously. The fluorescent signal from this is correlated to protein quantity (in any step of the expression and purification process), is visualizable for detecting protein localization, and can be removed by cleavage during purification [16,107,109]. Size exclusion chromatography (SEC) is a useful tool to measure dispersity throughout the purification of a transporter and fluorescence size exclusion chromatography (FSEC) is an adaptation of this which allows for a easier, more sensitive screening of proteins tagged with GFP [110,111].

### 7.2 Proteoliposome-Based Transport Assay

Using cell-based assays is a common tool for measuring function of sugar porters [43,53,112,113]. This has the advantage that the POI need only be overexpressed in a cell, but determining proper kinetics of the transporters can be difficult as it is difficult to accurately quantify the amount of protein in the cell, nor is it possible to isolate conditions (such as bilayer lipid composition or association of other proteins) for assessing external influence on the protein. Furthermore, for measuring glucose transport, 2-deoxy-D-glucose is used instead of glucose to avoid being metabolized by the cells [114]. Based on this, sugar porter function should be measured using an *in-vitro* system where a measurable quantity of purified protein is embedded into a liposome and transport of sugar can be measured by amount taken up, without the need for sugar analogs such as 2-deoxy-D-glucose. For this, proteoliposomes are used, vesicles which mimic the lipid bilayer containing a POI. A proteoliposome assay for measuring sugar porter activity is a main focus in paper III of this thesis.

*Considerations in proteoliposome preparation*

Notwithstanding their usefulness, the path to a proper proteoliposome transport assay is painstaking. The optimization of this assay for GLUT5 and other
sugar porters is the main topic for paper III, several features of this assay will be discussed here.

Figure 7.1: Basic pipeline of proteoliposome-based transport assay as described in paper III. 
i: Purified protein in detergent micelles is mixed with pre-formed liposomes, ii: the purified protein is incorporated into the liposomes by the rapid dilution method, iii: radiolabeled sugar is added to the proteoliposomes, iv: the excess sugars are washed away, sugars transported into the proteoliposome are detected by scintillation counting.

Besides protein purification, liposome preparation is the first step in proteoliposome assays where the objective is to produce liposomes of roughly equal size by means of freeze/thaw cycles and extrusion through a 400 nm filter. Liposomes larger than 400 nm can be multilamellar, meaning that several bilayers are inside the liposome, which is not ideal for measuring protein transport [115]. Smaller liposomes are also undesirable as the curvature of the membrane should be comparable to in vivo measurements [116, 117]. Purified proteins in detergent micelles are incorporated into liposomes by the rapid dilution method (see Figure 7.1), where an overabundance of lipids are present compared to the amount of protein, allowing the protein to spontaneously reconstitute into the liposome; a negligible amount of detergents could be reconstituted as well [118, 119].

Considerations about protein or lipid concentration in these preparations should be made, though no generalizable model has been proposed. Still after optimization, the observed reconstitution efficiency for sugar porters into proteoliposomes in the work in this thesis (paper III) is roughly 35%, and the protein orientation for most sugar porters is roughly 80% with the N and C termini inside the liposome. Exactly what determines the reconstitution efficiency or protein orientation is not known and furthermore, activity in proteoliposomes differs between various sugar porters. That being said, as discussed in chapter 4, activity of these proteins are sensitive to their
lipid surroundings, with factors such as membrane fluidity and presence of negatively charged lipids playing a role in transporter function [69, 71, 120].

**Measuring sugar transport in a proteoliposome assay**

Once the protein has been incorporated into the liposome, there are several means of measuring sugar uptake. The use of radiolabeled sugars as substrate is popular, effective because the property of the sugar is practically unchanged and detection of sugar uptake is simple in whole cells and proteoliposomes [120, 121]. This method is used in the work encompassing this thesis in papers I-IV. Another technique in sugar import measurement is the FLIP (FLII^{12}Pglu – 700μδ) sensor, which emits specific fluorescence resonance energy transfer (FRET) signals when glucose binds to it inside a cell [101, 122]. This has the benefit of being able to measure time-resolved glucose import kinetics which is much more cumbersome in radiolabeled-based measurements, though is only optimized for whole-cell experiments with glucose.

Assays using radiolabeled sugars is done in the work in this thesis using zero-trans uptake, where the sole sugar present is radiolabeled sugar added from the outside of the liposome at a low concentration (low µ range), thus the amount of sugar present inside the liposome is zero at the start of the experiment. Counterflow measurements are another popular uptake method, where liposomes are pre-loaded with unlabelled sugar before the radiolabeled sugar is added, meaning that the transporters are effluxing sugar from the inside to the outside of the liposome at the start of the experiment [58, 123]. Counterflow setup is useful when analyzing proton-coupled sugar transporters like XylE, since mutants which abolish proton-coupled symport may still work in counterflow, as long as they have not perturbed sugar binding [123].

### 7.3 GFP-BASED THERMAL STABILITY ASSAY

The folding of a protein is dictated largely by the hydrophobic effect as well as hydrogen or covalent bonding, disulfide bridges, van der Waals contacts, and more [124]. For membrane proteins in particular, this is further complicated by their embedding in a lipid bilayer, where a large portion of the protein exterior will be exposed to a bilayer’s hydrophobic surface. The
irreversible process of a folded protein leaving this state is called denaturation. Intentionally denaturing proteins by altering environment (such as pH) or applying heat to the system allows for one to measure the amount of work needed to put into the system to perform the denaturation. For heat-based metrics, the melting temperature $T_m$ is a useful metric, defined as the temperature in which half of an initially folded protein has become denatured. Factors binding to a protein (such as substrate, inhibitors, lipids, etc) can increase or decrease the $T_m$ of a protein by introducing new interactions that will change the proteins thermostability \cite{18,111,125}. For some proteins which have an altered $T_m$ upon ligand binding, this can be used as a proxy for a binding assay \cite{18,126}.

Measuring a $T_m$ can be performed using SEC post-heating to measure the dispersity or relative quantity of folded protein compared to an unheated control. Although effective, this is time-consuming and cumbersome. A faster alternative to this is to use a detergent to help precipitate aggregated/denatured protein after heating and measure the relative amounts of protein remaining in the supernatant (i.e. solution which remains after centrifugation). GFP fusion protein is highly suitable for this, as the protein amount is apparent by the relative fluorescence from a sample \cite{18}. Although the $T_m$ of sugar porters such as GLUT5 or PfHT1 does not change in the presence of sugar binding (data not presented), the $T_m$ does change the presence of different lipids, which highlights the key interplay between lipids which are critical for protein function and for protein folding, as discussed in paper III.

7.4 Structural Biology of Sugar Porters

Determining protein structure is an increasingly accessible tool for membrane proteins as techniques specific to these proteins develop \cite{127–130}. A protein structure typically contains a model of the 3D arrangement of atoms in a protein which representing a single protein conformation. Depending on a variety of factors such as sample heterogeneity, each structure has a certain resolution, which defines the level of confidence in placing atoms in a position in 3D space \cite{131}. A protein structure is used to visualizing interactions of residues in the protein, assessing protein state, interactions with ligands, and so on, but also facilitates the ability to simulate these structures using molecular dynamics simulations, as will be discussed in chapter 8.
**X-ray crystallography**

Historically, X-ray crystallography has been the classic method for structure determination of membrane proteins. In this, a protein is purified and crystallized and then subjected to an X-ray beam, generating a diffraction pattern which can yield an electron density map. From this map, a protein structure can be reconstructed: a more highly ordered protein can yield a higher resolution map, meaning that more confidence can be assigned to the position of the atoms in the resulting model of the protein [130]. This entire process is challenging, especially for membrane proteins [108,130]. For instance, crystallization of the protein requires high protein concentration, detergent micelles around the protein disrupt crystal-forming contacts, and the conformational heterogeneity of these proteins yields poor densities in the resulting maps [108,130]. Despite this, techniques have been developed to overcome these barriers for membrane proteins, and thus most structures to date of sugar porters are obtained using X-ray crystallography [34,35,58,65]. This is a method used both in paper I and paper IV here.

**Cryo-EM**

Recent years has seen an explosion of cryogenic electron microscopy (cryo-EM), which has far superseded X-ray crystallography in number of published membrane protein structures per year. As seen in Figure 7.2, in 2022 alone, 86% (of 1027) of published structures of membrane proteins were determined using cryo-EM, compared to 51% (of 631) in 2019, and 22% in 2016 (of 259). It is also apparent from this data that the increase of cryo-EM has not fully replaced X-ray crystallography in membrane protein structure biology, more that cryo-EM has led to an increased accessibility of membrane protein structure determination. This is largely due to the fact that cryo-EM does not require crystallization nor is sample...
heterogeneity as detrimental of a problem [22, 127, 129, 132]. In cryo-EM, protein samples are quickly frozen in a thin layer of ice and many thousand 2D images are captured of the sample using an electron beam, where noise in protein conformation heterogeneity can be averaged out during data processing [133]. In fact, tools capitalizing on this heterogeneity is becoming popular in cryo-EM, where an ensemble of states can be captured from datasets [134]. As is the case in X-ray crystallography, the end result of processing the data coming from this is a 3D map (in this case, a Coulomb potential map instead of electron density) which will be used to model atoms in 3D space [131, 132].

Because the sample does not need to be crystallized or homogenous in cryo-EM, this tool can also be used to capture structures of entire protein complexes, or purified membrane proteins can be reconstituted into a lipid bilayer mimetic known as a lipid nanodisc [128, 135]. Lipid nanodiscs are useful both because one can potentially capture a nonannular lipid interacting with a membrane protein, and that they can shift the conformational dynamics of the protein to capture a distinct protein state or oligomeric population [22, 66, 128, 136]. In paper IV, cryo-EM structures using lipid nanodiscs were used and aided in capturing a distinct state of inhibitors bound to PfHT1.
Computational modeling of protein structure, function, or dynamics can take many forms but the overarching theme among techniques is their predictive power. Given the central dogma of molecular biology that a DNA sequence will ultimately give rise to a functional protein, it should be possible to predict protein structure or behavior based on its sequence. However, proteins don’t exist in isolation, and the effects their environment has on them typically has to be simplified in computational modeling which is one factor which can effect the accuracy of this modeling [137]. Regardless, computational techniques have been successful in performing experiments which are otherwise costly or impossible to perform in vitro, and the increasing availability of data (such as more published protein structures) has served to strengthen their power [138]. For example, these methods have succeeded in predicting protein domains, structure, dynamics of a transition state, binding partners to a protein, or lipid effect on oligomeric states of membrane proteins, to name a few [137–140]. In silico and in vitro methods do best in complementation with each other, as has been evident by their combined use in all papers I, II, III, and IV.

8.1 Multiple Sequence Alignments

A multiple sequence alignment (MSA) is one highly used method in the in silico modeling toolkit, where many (three to millions) of sequences are aligned together, typically with a heuristic algorithm which balances computational time with alignment accuracy. Each alignment algorithm varies, though all introduce gaps in the construction of alignments so that conserved regions between sequences can be highlighted [141]. MSAs have been used in identifying protein motifs or domains in related proteins [10], but can also aid in the opposite way by identifying distinguishing features of related
proteins which display diverse function (such as identifying determinants of substrate specificity) [142,143]. The information gained from MSAs provides evolutionary context, aiding in construction of phylogenetic trees (for example, the tree seen in figure 5.1) as well as classifying otherwise uncharacterized proteins [11,47,49,78]. The evolutionary information present in a MSA improves the quality of secondary or tertiary structure prediction tools because proteins related in sequence often fold in a similar manner, which has been utilized in topology prediction tools and the highly popular AlphaFold [10,139]. Multiple sequence alignments have been useful tools for the research presented in papers I, II and IV.

8.2 Molecular Docking

Docking is a molecular modeling technique which (typically) predicts the binding of a small molecule ligand to a protein [144]. Docking starts with defining atom types for both protein and ligand, establishing a box in a 3D space where the docking can take place (for example, one can restrict docking to a known binding area or allow the ligand to bind anywhere on the protein), and identifying rotatable bonds in a ligand [144,145]. Once this is established, a docking simulation will then place the ligand somewhere in this docking box in interaction with a macromolecule, moving the ligand about its rotatable bonds, and calculating a score of a the pose based on desolvation effects, entropy penalties of binding, and interactions between the ligand and protein such as hydrogen bonding, electrostatics, or hydrophobic interactions [146,147].

Although some docking algorithms allow for side chain flexibility for a protein residue [148], the protein is overall static meaning that the results from docking algorithms are sensitive to the structural state of the protein which can limit their predicative power if the relevant binding conformation is not known. Docking is nonetheless a quick and affordable means of predicting ligand binding mode, and has been useful in the research presented both in papers III and IV.
8.3 MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics (MD) simulations are a computational modeling method which is used to understand the behavior of a protein in atomic detail [149]. The foundation of a MD simulation is the forcefield, a means of describing how all atoms in a system behave, describing their mass and charge for example. Pairs or groups of atoms interacting are also described in a forcefield, giving parameters for electrostatic or van der Waals interactions, as well as covalent interactions where bond length and angle are also described [149–151]. The force that every atom in the system has on each other is calculated based on this forcefield, and Newton’s law of motion $F_i = m_i a_i$ can be solved for atom $i$, where $F_i$ is the force that is being exerted on an atom, $m_i$ is the atom’s mass, and $a_i$ is the acceleration of that atom (i.e. the second derivative of a position over time). From this, the position of the atoms are updated and the calculations are performed again, resulting in a movie-like trajectory for the simulation where a protein can explore local minima in its energy landscape. MD simulations have been used in research presented here in papers I, II, and IV.

It is worth noting that several approximations must be made during simulations. The forcefield is an approximation of atom behavior based on calculations from experiments (both in vitro and in silico), and is best parameterized for common features in macromolecular simulations (i.e. water, amino acids, some lipids) [150]. Generalizable forcefields also exist, these are used in approximating parameters for small molecules in simulation, but results here should be interpreted with more caution than usual [152]. However, this approach is helpful when simulations of ligand-protein complexes is desired, and has been used in paper IV. Another approximation is the time in which the atom position should be updated. This is known as a time step which is on the femtosecond scale in order to balance the computational load of simulations and their accuracy, typically shorter than the fastest vibration in a system (i.e. hydrogens) [153].

Simulation pipeline in the context of sugar porters

A typical preparation of a MD simulation for membrane proteins takes place over several steps, as summarized in Figure 8.1. Initially, the forcefield for the 3D model of the protein is established, and hydrogens are added to the atoms in the protein. For titratable residues, the hydrogens added are based on their protonation state at pH 7.5, an approximation as covalent bonds are not broken during classic MD simulations [149,151]. From here, the
Figure 8.1: Basic pipeline of membrane protein simulation preparation. 

\( i \): The model of the protein is parameterized, 
\( ii \): the protein model is embedded into a lipid bilayer, 
\( iii \): a water box with ions is added.

protein is embedded into a bilayer of lipids; many biological lipids have been parameterized in forcefields, but the ‘classic’ lipid for a conventional MD simulation is often phosphatidylcholine (POPC) due to its net zero charge, abundance in eukaryotic membranes, and frequent use in simulations (rendering comparison between conditions easier) \([154,155]\). This protein and bilayer is then hydrated in a water box, where ions (typically sodium and chloride) are present.

Given the amount of atoms in these systems (typically a few hundred thousand) and the femtoseconds-long time step taken, MD simulations of sugar porters tend to be between 100 - 1000 nanoseconds long \([58,156]\). From the trajectories, perhaps the most typical analysis of the simulations would be measuring the root mean squared deviation (RMSD) which is defined by

\[
RMSD(t) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} d_i(f)^2}
\]

The starting frame of the simulation is typically the reference for which the distance deviation \((d_i)\) of an atom \(i\) is measured throughout each frame of the simulation \(f\), for \(n\) atoms. This gives a good approximation of the movement of the entire protein in a simulation:
for sugar transporters a RMSD of about 0.4 nm is expected in a metastable system.

**Collective variables**

The type of analysis performed on MD simulations will depend on the study. Regardless, a common interest is to identify which state along some path a protein is in. For example, measurements of the distance between the extracellular or intracellular gating helices (see Figure 4.3) is one useful metric to determine where a transporter is in alternating access, as seen in paper I [54]. Features such as the extracellular and intracellular gates are examples of **collective variables** (CVs, sometimes also referred to as a *reaction coordinate*); features that can both distinguish metastable states along some transition path while also reducing the high-dimensional motion of a simulation in a more interpretable low-dimension way [157–159]. For this reason, CVs are useful both in understanding as well as driving protein conformational dynamics in simulations.

The extracellular and intracellular gates are examples of a biological approach to CV choice, but data-driven methods are increasingly popular with the advancement of machine learning. **Principal component analysis** (PCA) and **time-lagged independent component analysis** (TICA) are both unsupervised techniques in dimensionality reduction which will transform high-dimension data into lower-dimension components. They differ in that PCA will find the maximum variance in data, whereas TICA will find slowly varying data over time which has proven particularly suitable for state separation in MD simulations [159,160]. Supervised methods in CV selection are also popular where information about protein states or important contacts is provided to the model which can then identify the features most important for forming these states/contacts. This does require *a priori* knowledge about a system, but has proven to be very successful at producing meaningful collective variables [143,158].

**8.4** **ENHANCED SAMPLING MOLECULAR DYNAMICS**

Especially with growing computer power, infrastructure for *high performance computing* (HPC), as well as optimization of MD software to work on HPC environments, the amount of information one can obtain from MD simu-
lations has skyrocketed in the past decade [161–163]. Nonetheless, even if simulations can be extended to reach several thousand nanoseconds, this is generally not enough time for sampling a biological process. For instance, the turnover of sugar from the outside to the inside of the cell for PfHT1 is reported to be about 19 molecules per second (paper I), a timescale that conventional MD simulations do not currently reach.

To reach longer timescales in molecular dynamics simulations, several different approaches can be used. One such approach is the use of coarse-grained molecular dynamics (CGMD), where atoms are grouped into beads and the forcefield is much more approximate [164]. The timestep in these simulations can therefore take a longer stride, easily reaching microseconds-long timescales. Because of the atom approximations, CGMD systems can also be larger, therefore this is particularly suitable for assessing lipid bilayer dynamics and their long-term interactions with membrane proteins [23, 26, 70]. Although proper sampling of protein conformation is difficult due to the approximation of atoms in the system, CGMD is useful in characterizing state-dependent interactions with proteins, such as ligand or lipid binding [25].

To accelerate conformational sampling of a protein, enhanced sampling molecular dynamics is often used, of which there are many techniques to choose from depending on the desired results and a priori information available [165]. One way to categorize enhanced sampling simulation methods is by their use of collective variables or not. In the latter, no a priori knowledge is required about the system and often involve the exploitation of temperature to move a system around its energy landscape quickly. In collective variable-based methods, bias is introduced to the system to move the protein around a conformational space, which requires a priori knowledge of how this could happen. The bias applied to the system is usually not high-dimensional, using one or several CVs to drive conformational change [165]. These methods are the more popular choice for enhanced sampling MD of macromolecules [157], as presented in paper II of this thesis.

String method with swarms of trajectories

One such CV-based enhanced sampling method is known as string method with swarms of trajectories (string simulations for short), which has been successful in understanding conformational dynamics of several membrane proteins, including sugar porters both in this thesis (paper II) and in other works [54, 166, 167].
Figure 8.2: A simplified graphic of string simulations. Endpoints are colored in light grey, beads are represented as dots, and strings are represented as the curves connecting the dots. As seen in the left insert, swarms are launched from a bead, which will dictate how the string moves across an energy landscape. If sampling has been correct, the final string (yellow) should be visiting energy minima in a landscape.

As seen in Figure 8.2, string simulations begin with defining protein conformations (the beads) along an initial path (the string), between a starting and ending conformation in a collective variable space. From these beads, many short simulations (the swarms, left insert in Figure 8.2) are launched. The average drift from these simulations is calculated, and the string is reparameterized to follow this drift in CV space. After many iterations, the strings will find a local minima and convergence has been reached, for which additional sampling can generate a free energy surface (FES) based on the transitions made in each iteration [168].

The generation of this initial path between end states is a critical step in string simulations, as the transition between protein states is usually unknown. Occasionally, long MD simulations may have already sampled this conformational space, [167], but more often an initial estimation of the path is needed [166], as used in paper II. This typically requires enhanced sampling simulations themselves, using steered MD (biasing along some CV) or targeted MD (biasing all atoms in a 3D space) to push a system between states [169, 170]. If this transition is not estimated properly, for instance if the chosen collective variables do not describe conformational changes well enough or state intermediates are not well known, errors will accumulate in these simulations (known as hysteresis). These errors will propagate
through iterations of string simulations, leading to nonsensical free energy surfaces [171, 172]. Hysteresis is not a problem specific to string simulations, but means that initial path generation should be considered carefully.

These simulations therefore require more *a priori* knowledge than many other CV-based enhanced sampling simulations, given the need for an initial path. For this reason, the structural characterization of sugar porters was paramount to the work in paper II where string simulations were used. Notwithstanding, string simulations also have several advantages, as much of the simulations are performed in equilibrium and therefore the method is flexible to the choice of collective variable and results are easy to interpret [166].
PAPER I


The hexose transporter for the malaria parasite Plasmodium falciparum, named PfHT1, imports sugar, mainly glucose and fructose, from an infected erythrocyte into the parasite. PfHT1 can recognize and transport many hexoses, a feature that stands out among sugar porters which typically only transport one type of sugar. In this paper, we determined the crystal structure of PfHT1 in the fully occluded state with D-glucose bound, and saw that the substrate binding site and glucose coordination is nearly identical to the GLUTs. In transport assays of PfHT1, we found via mutagenesis that many residues in the binding site are critical for substrate transport, but we could not find a particular residue which altered substrate specificity.

The structure of PfHT1 is in the fully occluded state, and we could see that the extracellular gate of PfHT1 was much more polar compared to the GLUTs. Mutations to residues on this gate rendered PfHT1 unable to transport sugars, highlighting the importance of this gate for protein function. Furthermore, we observed that during MD simulations of the occluded state structure of PfHT1 that the extracellular gate remains shut, compared to simulations of a previously published outward-occluded state structure of GLUT3, where the gate eventually opens and glucose unbinds.

The results of the structure, simulations, and mutagenesis led us to conclude that PfHT1 has an extracellular gate which more easily shuts compared to the GLUTs. This readily shut gate confers substrate polyspecificity and is also the reason why we were able to capture PfHT1 in the occluded state, which had not yet been structurally characterized for sugar porters yet. These results also help compose a complete picture of the five major states of the transport cycle in sugar porters.

* These authors contributed equally.
Knowing the five major states of the transport cycle in sugar porters, we aimed to obtain a free energy landscape of a sugar porter to understand how substrate binding and conformational dynamics are coupled. For this system we used the fructose-transporting sugar porter GLUT5, both for technical reasons as well as because fructose coordination had not yet been structurally characterized for GLUT5. We performed string simulations with swarms of trajectories with the help of other modeling techniques to reconstruct the alternating access mechanism in GLUT5 in the presence and absence of substrate, and obtain a free energy surface (FES) using the extracellular and intracellular gate distances as collective variables.

The FES in the absence of fructose indicated that the outward open state was the most energetically favorable state, which was in line with previous research. The FES in the presence of fructose showed a flattened energy surface, with inward-facing states being similarly favorable as outward-facing states and the occluded state of GLUT5 (modeled after the occluded structure of PfHT1) being precisely on the barrier between these states, indicating a transition state. We determined that fructose had a higher coordination in this occluded state compared to an outward open state, indicating that the occluded state is the essential state for sugar recognition.

In the conformations of the occluded state, residues which were specific to GLUT5 on the extracellular gate were interacting with GLUT5-specific residues on the same helix as the intracellular gate (i.e. TM7b-TM10a interactions). Mutations to these residues or residue pairs yielded nonfunctional GLUT5 in our transport assays. We therefore concluded that these residue interactions couple sugar recognition to the alternating access mechanism, and that this information can be used to further understand substrate specificity.
PAPER III


Establishing working functional assays for sugar porters remains a challenging task, as whole cells assays are not ideal for kinetic measurements and proteoliposome assays typically show a very low signal to noise ratio. In this work, we established an in vitro proteoliposome transport assay. With these results we show: the importance of result interpretation based on these assays, difference in GLUT kinetics, and importance of membrane fluidity on GLUT activity.

In the proteoliposome assays, purified protein is mixed with pre-formed liposomes and incorporated by a rapid dilution method. Radiolabeled sugar will be taken up by the protein into the liposome, where sugar quantity can later be measured by scintillation counting. We found in optimizing this assay that certain extracts of lipids gave higher specific uptake, particularly with the crude extract brain fraction VII. Proteoliposomes comprising this extract gave kinetic values matching in vivo estimates, and the effect of mutations or inhibitor to GLUT\(5\) could be measured for the first time.

Lipidomics profiles of different crude lipids showed that another extract, brain fraction I, contained more phosphatidylserine (PS) compared to brain fraction VII. Using the GFP-TS method, we found that brain fraction I increased the melting temperature of GLUT\(5\), as did pure PS lipids. By mimicking the brain VII extract using synthetic lipids, we could measure specific effects of lipids on transporter function, and found that removal of phosphatidylethanolamine (PE) was the most detrimental to transport. This seems related mostly to membrane dynamics, as altering lipid tail saturation shows a similar effect. Based on these results and the GFP-TS, sugar porters seem sensitive to lipid composition mostly out of membrane dynamics; there seems no specific lipid binding tightly to the protein and conferring strong function and increased melting temperature.
PAPER IV


The *Plasmodium falciparum* hexose transporter *PfHT1* is a potential antimalarial drug target given its importance in parasite growth; at least three compounds which have an adverse effect on parasite growth target *PfHT1* specifically. In this work, we structurally characterized *PfHT1* bound to these compounds, propose how these drugs can target *PfHT1* when it is so similar to the human GLUTs, and explored how structure determination conditions can change the interpretation of small-molecule interactions with proteins.

We determined a crystal structure of *PfHT1* in complex with the inhibitor MMV009085, which was found in the sugar binding site. Many compounds analogous to MMV009085 did not inhibit *PfHT1* in our functional assays; from this we concluded that the length of the symmetric four-carbon butanol tails of MMV009085 were critical for inhibition. However, we were not confident of the binding pose from the crystal structure, based on functional assays and simulations, so we determined a structure of *PfHT1* with an inhibiting analog of MMV009085 using cryo-EM. In this, the MMV009085 analog was bound in the substrate binding site, but more near to the intracellular gate. We also determined the structure of *PfHT1* in complex with another inhibitor named WU-1 and observed a similar binding mode. To understand if the conditions in two different structure determination methods shift the binding interfaces between proteins and small molecules, we determined a cryo-EM structure of *PfHT1* bound to the inhibitor C3361, of which a crystal structure had previously been published. Indeed, the inhibitor was bound in a different orientation and thus the physiologically relevant binding mode of C3361 is unclear.

From simulations and functional assays of MMV009085 and WU-1, we concluded that specificity of these competitive inhibitors stemmed from residues near the binding site. These residues, unlike the conserved sugar binding site, are unique to *PfHT1* and therefore are targetable, which could provide inspiration for targeting other GLUTs with competitive inhibitors.

† Final author list subject to change
In the work presented in this thesis, I have demonstrated the value of a multifaceted approach to understanding sugar transporters by examining their conformational dynamics, which links protein structure to their function.

**Paper III** has set the foundations for the functional work in my PhD. The establishing of a transport assay for sugar porters allowed us to accumulate a large quantity of functional data: without a proper signal to noise ratio in these assays, mutagenesis and kinetic measurements would not be possible. Furthermore, the occluded state structure of PfHT1 from paper I has set the foundation for the theoretical work in my PhD as it provided the structural template completing the sugar transport cycle.

The occluded state is the key to coupling sugar binding to conformational change in the protein. We believe this is a transition state in sugar porters. The affinity for substrate appears to be highest at this point: as seen in **paper II**, fructose is the most coordinated when passing through the occluded state. Although this still needs validation, we propose that this coordination is necessary to properly couple the motions between substrate binding and conformational change; without the precise coordination of sugar, this process will not happen. A transition state should not be metastable [40]; the formation of the occluded state is not entropically favorable which should encourage the protein to leave this state. However, this state must be visited in the absence of sugar (i.e. the resetting step) and the unfavorable energy of the occluded state forming is likely the reason why passing through the cycle without sugar is 100-fold slower [43].

It is still not known whether the sugar binding couples to conformational change by induced fit versus conformational selection. In the induced fit model, akin to enzymes, the protein undergoes the conformational change as substrate is bound (see Figure 3), whereas in conformational selection, the substrate would select for the protein state [173]. One argument for conformational selection would be that a sugar porter is visiting an occluded state in the absence of sugar, but it is not known whether this occluded state is the exact conformation that the sugar-bound state visits. As a final note on this topic, it should be noted the context in which we define a state. We refer
to the structure of \( PfHT_1 \) from paper I as being in the occluded state, but this is a metastable state that is very near the transition state, not actually the transition state which is not observable with structure biology techniques.

The coupling of the substrate binding site to the rest of the protein, which we first explored in paper I and revisited in paper II, has also contributed to our understanding of specific inhibition of sugar porters in paper IV. Here, we now understand that a competitive inhibitor can still be specific, even when the area it is binding to is not unique. In the case of the inhibitors of \( PfHT_1 \), unique residues adjacent to the binding site can affect how the binding site residues interact with the ligand, conferring inhibitor specificity. There is still much to explore in this, but this could be contributing to substrate specificity in the sugar porters, and could be exploitable for developing GLUT isoform-specific inhibitors.

Lastly, the gates and regions adjacent to substrate binding site are not the only regions which provide specificity to the sugar porters. We do not know why many show different rates of transport, resetting, or binding asymmetry. As discussed in paper III, the effect of lipids on sugar porters seems to be largely mechanical, but studies have been done to show that different lipids shift the conformational populations of these proteins [71], and we have seen in our results that XylE (a bacterial sugar porter) is the only sugar porter that transports best in E. Coli lipids. Clearly, there are interactions between the lipids and proteins that alter their conformations, but what these are remain a mystery.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor David Drew for taking me on as a PhD student in your lab. You have given me a lot of both guidance and trust through the years on my projects, and I am very thankful for this. I also would not have managed any of this if it weren’t for my project supervisor Lucie Delemotte who accepted me as a project worker so long ago. You helped me pave the way both into and throughout the PhD, and have provided so much support during my time. A big thanks to Albert, Ashu, Marta, Claudia, Magnus, and David for helping correct my thesis.

I would not have managed this PhD without the support and help of many Drew and Delemotte lab members. Firstly, the biggest thank you to Albert who graciously received the task to supervise me in the lab when I knew nothing. You have also been a super colleague to collaborate with on these projects through the years, I could not have asked for anything more. Ashu, thank you for being such a joyous colleague, you can make me smile even on the worst days. Your support on these projects has also been invaluable. Tom, I have learned a lot from you in the lab but you have also taught me how to keep my cool (I am still not good at it) and find humor when things are getting wild, thank you very much for that. I hope I can babysit your kitties again some time. Darko, thank you so much for your help and encouragement during the GLUT5 project, I think I would not have managed without you. I can’t wait to see what you will do in the future. Magnus, you manage to wrangle this Drew lab around DBB and SciLife, well done. You know how to keep your head up and are always willing to give guidance when needed, thank you.

Being shared between the Drew and Delemotte labs got much less lonely when my Italian girls came around. Marta, your laughter is one of my favorite things to hear and Claudia, you are such a kind soul, I know I always have your ear. I really love working with both of you, the way you both are always willing to help and support me and each other is really inspiring.

I am so thankful for all other current Drew and Delemotte lab members who I don’t work direct with, but who have always given me happiness at
work. Sukkyeong, I love our morning desk conversations and will miss them terribly. Ved, I am happy to have a fellow beer lover around in the lab and I will never forget the time you threw the mint exactly into my coffee cup, well done. Surabhi your energy and laughter is truly infectious to be around. I am cheering for you as you round the corner on your PhD. Dohwan, you are so kind and considerate, my favorite times with you were always the DBB pub. John, I am always happy to have a fellow American around, your understated ways of saying things are my favorite. Yue you are just the funniest person in the world, I like to sit next to you during meetings because I know you will always say something great. Antoni you are a very considerate and inquisitive colleague, the Delemotte lab is truly extremely lucky to have you. Andrea and Adrien, our time together in Lucie’s lab has been too short but I know you are going to do very well there, I look forward to seeing it.

To all past Drew and Delemotte lab members, Yurie, Roman, Pascal, Hyunku, Aziz, Rachel, Iven, Jan, Mathieu, Rei, Emmanuel, Joseph, Koushik, Will, Annie, Oliver, you guys have been so fun to work with and I miss having each and every one of you around. Especially thanks to Marina who taught me so much about simulations and provided many encouraging words through the years, and Sergio who helped with the string simulation analysis so that even a biochemist could learn!

I am eternally thankful for everyone else in the MBS environment who has contributed to such a nice working place, especially Reba who is always willing to go out of her way to help others. Thanks to my fellow bioinformatics TA’s, especially Eloy, Marie, and Gabriele of whom we shared many years of pain and also laughter.

Pia, Gunnar, Erik, thank you for your mentorship through the PhD program. The biggest thank you ever to Matt who goes above and beyond for this department constantly, including booking this very room for my defense.

Lastly, I would like to acknowledge my partner Iven, my friends and family, especially my parents and sister Emily. You have all listened to more than your fair share of my struggles during the past 5 1/2 years, have always lent encouraging words, and given me so much joy in my life. Thank you.


[49] “Interpro - major facilitator superfamily ipr011701.”


