2009

Evolution of Substrate Specificity in the Rok Superfamily and the Origin of Allosteric Regulation in Human Pancreatic Glucokinase

Mioara Larion
FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

EVOLUTION OF SUBSTRATE SPECIFICITY IN THE ROK SUPERFAMILY AND THE ORIGIN OF ALLOSTERIC REGULATION IN HUMAN PANCREATIC GLUCOKINASE

By
MIOARA LARION

A Dissertation submitted to the Department of Chemistry and Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Degree Awarded:
Summer Semester, 2009

Copyright © 2009 Mioara Larion
All Rights Reserved
The members of the committee approve the dissertation of Mioara Larion defended on June 29, 2009.

____________________________________
Brian G. Miller
Professor Directing Dissertation

____________________________________
Ross W. Ellington
Outside Committee Member

____________________________________
Penny J Gilmer
Committee Member

____________________________________
Lei Zhu
Committee Member

Approved:

____________________________________
Joseph B. Schlenoff, Chair, Chemistry and Biochemistry

____________________________________
Joseph Travis, Dean, Arts and Sciences

The Graduate School has verified and approved the above-named committee members.
To my family: Stefan, Rodica, Loredana, Ionut, Eusebiu, Florentina, Felicia and Alina

and to Adrian
ACKNOWLEDGEMENTS

The beginning of my doctoral studies at Florida State University was largely due to my college professor Dr. Elena Gonzales, who believed in my interest for science. Therefore I want to take the opportunity and thank Dr. Elena Gonzales with all my heart. Without her guidance as a college student I would not choose to pursue a doctorate.

I also want to thank my mother Rodica Larion and my father Stefan Larion for their continuous support and encouragement.

I want to thank Dr. Miller, my advisor and mentor, for his permanent support, availability help, kindness and consideration throughout all these years. I am forever grateful for accepting me in his group and giving me a second chance, advice, helpful ideas and a very interesting subject to study: human glucokinase.

I want to thank my colleagues Kevin Desai, Casey O'Brien, Kristen Ramsey, for their help throughout the last years.

I want to thank Dr. Claudius Mundoma for all the hours of training on most of the biophysical instruments in the PBF, especially for the help on the AUC set up and analysis and for very useful discussions.

I would like to acknowledge Margaret Seavy for her help with N-terminal sequencing of human glucokinase.

I would like to thank my committee members, Professors' Ellington, Zhu, Gilmer and Miller, for their time to review my dissertation and their help during the committee meetings.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>xii</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Bacterial Sugar Kinases: Examples of Modern Enzymes with Broad Substrate Specificity</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Human Hexokinases: Regulatory Enzymes Evolved via Gene Duplication</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Human Glucokinase: a Monomeric Enzyme with Positive Cooperativity</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Sugar Specificity of Glucokinase</td>
<td>13</td>
</tr>
<tr>
<td>1.5 Aims of the study</td>
<td>14</td>
</tr>
<tr>
<td>2. BACTERIAL GLUCOKINASES: DIVERGENT EVOLUTION OF FUNCTION IN THE ROK-SUPERFAMILY</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Materials and Methods</td>
<td>19</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>23</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>29</td>
</tr>
<tr>
<td>3. KINETICS OF GLUCOSE BINDING TO HUMAN PANCREATIC GLUCOKINASE</td>
<td>33</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>34</td>
</tr>
<tr>
<td>3.2 Materials and Methods</td>
<td>38</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>41</td>
</tr>
<tr>
<td>3.4 Discussion of the Glucose Binding Model Proposed</td>
<td>50</td>
</tr>
<tr>
<td>4. INVESTIGATION OF THE CONFORMATIONAL HETEROGENEITY OF HUMAN PANCREATIC GLUCOKINASE VIA HSQC-NMR</td>
<td>55</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>56</td>
</tr>
<tr>
<td>4.2 Materials and Methods</td>
<td>57</td>
</tr>
<tr>
<td>4.3 Results</td>
<td>59</td>
</tr>
<tr>
<td>4.4 Discussion and Future Directions</td>
<td>80</td>
</tr>
</tbody>
</table>
5. THE ROLE OF THE C-TERMINAL ALPHA13 HELIX TO THE ALLOSTERIC REGULATION OF HUMAN PANCREATIC GLUCOKINASE

5.1 Introduction ................................................................. 83
5.2 Materials and Methods .................................................. 85
5.3 Results ...................................................................... 90
5.4 Discussion .................................................................. 96

6. CONCLUSIONS AND FUTURE DIRECTIONS .................. 101

REFERENCES ..................................................................... 106

BIOGRAPHICAL SKETCH ................................................... 117
LIST OF TABLES

1.1: Kinetic characterization and localization of different human hexokinase isozymes .......................................................... 3

2.1: Kinetic parameters for the phosphorylation of native and ambiguous substrates by wt-NanK and selected variants .................. 25

2.2: Kinetic parameters for the phosphorylation of native and ambiguous substrates by wt-AlsK and selected variants .......................... 26

2.3: Specificity profile of wild-type AlsK and A73G AlsK ........................ 27

2.4: Specificity profile of wild-type NanK and L84P NanK .................. 28

4.1: Activity assays for GK before and after the HSQC experiments ...... 59

4.2: Estimated molecular weights of the peaks eluted from the Superose 6 column ........................................................................... 68

4.3: Hydrodynamic parameter dependence on the angular velocity ...... 77

4.4: Calculated hydrodynamic parameters of wild-type GK in the super-open and closed conformations .................................................. 79

5.1: Kinetic and thermodynamic parameters of wild-type GK, truncated GK, and the truncated D205A variant ................................. 91

5.2: Effects on $k_{cat}$ upon peptide addition to truncated GK at 25°C and 35°C ......................................................................................... 91

5.3: Catalytic effects of C-terminal polyalanine addition .................... 93

5.4: The identity and frequency of amino acids observed at positions 450-456 following genetic selection for E. coli growth of a randomized α13 helix library ......................................................... 94

5.5: Kinetic and thermodynamic characterization of selected α13 variants ......................................................................................... 95
## LIST OF FIGURES

1.1: Glucose phosphorylation reaction catalyzed by *E. coli* glucokinase (Glk), human glucokinase (GK) and the ambiguous *E. coli* kinases AlsK, NanK, NagK and Mak................................................................. 1

1.2: Chemical structures of the ambiguous substrate D-glucose and the natural substrates D-allose, N-acetyl D-mannosamine, D-fructose and N-acetyl-D-glucosamine for AlsK, NanK, Mak and NagK............... 2

1.3: The link between the kinetic properties of Hexokinases I-IV found in rat and their physiological roles................................................................. 5

1.4: Steady state kinetics of human pancreatic glucokinase fitted to the Hill equation rather than the Michaelis-Menten equation ............... 6

1.5: Computer simulations of the steady-state velocity done by Ricard and colleagues for the case of a two substrate enzyme displaying positive cooperativity................................................................. 8

1.6: Crystal structure of unliganded GK (super-open) and liganded GK (in the presence of glucose and the allosteric activator) ......................... 9

1.7: Kinetic model for glucose binding to human pancreatic GK as proposed by Heredia *et al.* based on KIMSIM simulations..................... 10

1.8: Ligand induced slow transition model and the results of computer calculations of double reciprocal plots for a Hill coefficient of 1.28 ..... 11

1.9: Kim *et al.*, model includes a preexisting equilibrium to describe the transient state glucose binding traces................................................. 11

1.10: Antoine *et al.*, model includes a preexisting equilibrium to describe the transient state glucose binding traces................................. 12

2.1: Multiple sequence alignment of ROK superfamily members............. 24

2.2: Effects of the A73G AlsK loop substitution upon the kinetic parameters for a variety of carbohydrate substrates............................... 29

2.3: The active site of inorganic polyphosphate/ATP-glucomannokinase from *Arthrobacter* sp. strain KM in complex with D-glucose, revealing the position of the highly conserved glycine residue (Gly-84) with respect to the O3 position of the bound substrate................................. 32
3.1: Schematic representation of the minimal mnemonical model (A) and ligand induced slow transition model (B) proposed to explain the kinetic cooperativity of human glucokinase ................................................. 34

3.2: Crystal structures of human pancreatic glucokinase depicting the location of the tryptophan residues responsible for generating the intrinsic fluorescence ........................................................................... 37

3.3: Hyperbolic dependence of GK intrinsic fluorescence upon glucose addition .......................................................................................................................... 42

3.4: Attempt to fit the 96 mM glucose binding to human glucokinase curve to a double exponential fit .............................................................. 43

3.5: Attempt to fit the 96 mM glucose binding to human glucokinase curve to a triple exponential fit .............................................................. 44

3.6: Superposition of the 96 mM glucose binding curve to a quadruple exponential fit ...................................................................................... 45

3.7: The dependence of observed microscopic rate constants on the glucose concentration .......................................................................................... 46

3.8: Representative time traces for the 0.35mM-96mM range of glucose concentrations ........................................................................................................ 47

3.9: Stopped-flow binding trace of 40 μM human glucokinase with 250mM 2-deoxy glucose overlaid with a quadruple exponential fit ..... 48

3.10: Proposed mechanism along with the microscopic rate constants to describe glucose binding curves ......................................................... 49

3.11: Proposed mechanism along with the microscopic rate constants to describe glucose binding curves ......................................................... 50

3.12: Average trace of glucose dissociation kinetics superimposed with the fit composed of four exponentials to verify the authenticity of the four transients ........................................................................ 52

4.1: HSQC of 1.6 mM (A) and 120 μM (B) human glucokinase in the absence of ligand after Ni^{2+} column at 25 °C .......................................................... 60

4.2: Sedimentation velocity of 24 μM ^{15}N-GK conducted from the recovered sample (120 μM) after the HSQC NMR experiment ........ 61
4.3: Snapshot of the van Holde and Weischet analysis of human glucokinase after Ni$^{2+}$ column showing a fast moving component sedimenting faster than the rest of the sample ........................................... 63

4.4: Extrapolation plot of 26 μM GK (A) described in Figure 4.3 and the corresponding distribution plot (B) of the sedimentation coefficient across the boundary fraction, after the time invariant noise was subtracted ........................................................................................................ 64

4.5: Finite element solutions (pairs of S and S/f0) for the 26 μM GK displayed in UltraScan (Genetic Algorithm Initialization module) ........... 66

4.6: Calibration of the Superose 6 10/300 gel filtration column .................. 67

4.7: Elution profile of 100 μl of human glucokinase after overnight run at 0.02ml /min on Superose 6 10/300 column.............................................. 68

4.8: MALDI-TOF spectra of peak 1, 2, 3 and 4 with the y-axis aligned (m/z), indicating the molecular weight of the peaks .......................... 69

4.9: Sedimentation velocity (50000 rpm, 20°C) of 31 μM wild-type human glucokinase in the absence of glucose................................. 71

4.10: Sedimentation velocity (50000 rpm, 20°C) of 31 μM wild-type human glucokinase in the absence of glucose after 2DSA .................... 72

4.11: Sedimentation velocity (50000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 50 mM glucose ......................... 73

4.12: Sedimentation velocity (50000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 50 mM glucose after 2DSA......... 74

4.13: Sedimentation velocity (50000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 100 mM glucose ...................... 75

4.14: Sedimentation velocity (50000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 100 mM glucose after 2DSA ......... 76

4.15: Models generated by SOMO to calculate the hydrodynamic parameters of human glucokinase based on their X-ray diffraction structures .................................................................................................. 79

5.1: The crystal structure of unliganded human glucokinase (left) and glucokinase in complex (right) with glucose (yellow) and an allosteric activator (not shown) ............................................................... 84
5.2: Comparison of the steady-state kinetics of cooperative wild-type human glucokinase (A) and a non-cooperative truncated variant lacking helix α13 (B)................................................................. 90

5.3: Comparison between the fluorescence pattern of wild-type GK and the truncated GK.............................................................................................................. 92

5.4: The position of helix α13 (red) and the connecting loop (blue) in the structure of (A) unliganded human glucokinase and (B) glucokinase crystallized in the presence of glucose and an allosteric activator ...... 93

5.5: The structure of human hexokinase I, an isozyme of glucokinase, depicting the location of the connecting helix (red) that facilitates communication between the allosteric N-terminal domain and the catalytic C-terminal domain................................................................. 99
ABSTRACT

Sugar kinases are enzymes known to catalyze the phosphorylation of different sugar substrates. The sugar kinases encompassed by the ROK (repressor, open reading frame, kinase) superfamily offer a unique opportunity to explore the evolutionary origins of enzyme specificity. We probed evolutionary events that occur during the optimization of glucokinase activity in two members of the ROK superfamily, AlsK and NanK, which share only 21% sequence identity with one another. Following random mutagenesis and in vivo functional selection, we identified two structurally overlapping mutational “hot spots” in the sugar kinase scaffold. Steady state kinetic analyses of the selected variants demonstrate that the native activities of AlsK and NanK are largely unaffected by the glucokinase-enhancing substitutions. Furthermore, the variants have acquired an increased ability to phosphorylate a variety of nonnatural carbohydrate substrates as a result of the evolutionary process. This finding is consistent with an evolutionary process that includes the formation of intermediates possessing relaxed substrate specificities during the initial steps of enzyme functional divergence.

Highly evolved enzymes display low $K_m$ values for their substrates as a result of their functional refinement toward increased specificity. An anomalous behavior of higher order enzymes is presented by human glucokinase, a monomeric enzyme possessing kinetic cooperativity and a high glucose $K_m$ value. The sigmoidal response of the turnover rate as a function of glucose concentration allows human pancreatic glucokinase to regulate glucose concentration in the blood stream. The importance of glucokinase proper function is reflected in the different diseases associated with genetic mutations in glk. Maturity onset diabetes of young (MODY) is associated with genetic lesions in glk that reduce catalytic activity, while persistent hyperinsulinemia of infancy (PHHI) is found in hyperactive glucokinase variants. Different mechanisms have been postulated to explain glucokinase kinetic cooperativity, several of which involve the existence of multiple enzyme species interconverting with a rate slower than catalysis. Our kinetic studies of glucose binding, obtained via stopped-flow, suggest the existence of at least two species in equilibrium in the absence of glucose, which are able to bind glucose and subsequently isomerize. This postulated mechanism was further tested experimentally by analytical ultracentrifugation (AUC) and nuclear magnetic resonance (NMR). Analytical ultracentrifugation results show that human pancreatic glucokinase is a monomer in solution with a sedimentation coefficient of 3.5 S. Unfortunately, AUC does not provide the resolution needed to separate the different conformations of human glucokinase. Preliminary
HSQC-NMR spectra are characteristic of either (a) a partially unfolded protein or (b) multiple species slowly interconverting. Additional investigations conducted on AUC eliminated the possibility of aggregates as an explanation for the observed NMR spectra. Future experiments to test the existence and rates of GK conformers in solution are discussed.

Structurally, glucose binding to human pancreatic glucokinase involves large conformational changes of the small domain, towards formation of a more compact structure. Of particular interest was the secondary structural element located at the C-terminus (helix α13), which, upon ligand binding, moves from a solvent exposed position to a hydrophobic pocket. Deletion of helix α13 abolishes cooperativity and restores Michaelis-Menten kinetics. Addition of a thirteen amino acid synthetic peptide in trans does not restore the $k_{cat}$ value of the truncated variant. Elongation of α13 via the addition of a C-terminal polyalanine tail does not affect the glucokinase steady-state kinetics. Randomization studies of residues 450-456 of helix α13, followed by selections in a glucokinase-deficient bacterium BM5340(DE3), identified variants that possessed lower $K_{0.5}$ glucose values, Hill coefficient near unity and enhanced equilibrium binding affinity for glucose. Our studies on helix α13 demonstrate the essential role played by this structural element in governing cooperativity and establish a link between its primary amino acid sequence and the functional dynamics of the glucokinase scaffold that are required for allostery.
CHAPTER 1
INTRODUCTION

1.1 Bacterial Sugar Kinases: Examples of Modern Enzymes with Broad Substrate Specificity

The sugar kinases encompassed by the ROK superfamily include four ambiguous kinases discovered in the *Escherichia coli* K-12 genome: allokinase (AlsK) (1, 2), *N*-acetyl-D-mannosamine kinase (NanK) (3, 4), *N*-acetyl-D-glucosamine kinase (NagK) (5), and manno(fructo)kinase (Mak) (6). Allokinase (AlsK) phosphorylates the sugar allose, an epimer of glucose and was found to play a secondary role in D-allose metabolism in *Escherichia coli* (1). Phosphorylation of *N*-acetyl-D-mannosamine by NanK is followed by epimerization to *N*-acetyl-D-glucosamine-6-phosphate (7, 8), which is subsequently metabolized via fructose-6-phosphate (9-11) in the glycolytic pathway or combined with phosphoenolpyruvic acid to form sialic acid (12). NagK (*N*-acetyl-D-glucosamine kinase) phosphorylates *N*-acetyl-D-glucosamine (GlcNAc), which is the main substrate for the synthesis of UDP-GlcNAc. UDP-GlcNAc is important in the biosynthesis of N/O-glycans (13), in signal transduction processes via modification of serine and threonine residues of both cytosolic and nuclear proteins (14) and finally it is involved in the biosynthesis of sialic acids (15, 16). Mak phosphorylates fructose to form fructose-6-phosphate that is converted to fructose 1, 6 bisphosphate (by phosphofructokinase), to form a central carbohydrate in glycolysis and gluconeogenesis (17).

Figure 1.1 Glucose phosphorylation reaction catalyzed by *E.coli* glucokinase (Glk), human glucokinase (GK) and the ambiguous *E.coli* kinases: AlsK, NanK, NagK and Mak.
Although the four sugar kinases are contemporary kinases with relatively high catalytic efficiencies for their native substrates, and are involved in different biosynthetic pathways, they were found to possess a low level of glucokinase activity that allowed the survival of auxotropic BM5430 (DE3) on glucose minimal media (18, 19). This catalytic competence in phosphorylating an alternate carbohydrate that is similar in structure (Figure 1.2) to the native substrate is termed substrate ambiguity and is characteristic of many ancient enzymes (20). The ability of modern enzymes to possess broad substrate specificity (21) raises questions about their prevalence and use in modern metabolic pathways, where precise regulatory controls impede the anomalous reactions to happen.

![Chemical structures of the ambiguous substrate D-glucose and the natural substrates D-allose, N-acetyl D-mannosamine, D-fructose and N-acetyl-D-glucosamine for AlsK, NanK, Mak and NagK.](image)

The recruitment of modern enzymes like NanK and AlsK (18, 19) which share very low sequence identity (21%), to phosphorylate the alternative sugar substrate, D-glucose, allowed us to look for common mechanisms that occur during the early stages of evolutionary events that lead to the acquisition of highly specialized functions.
1. 2 Human Hexokinases: Regulatory Enzymes Evolved via Gene Duplication

In humans there are four hexokinase isoforms specialized for four different tissues. The distinct isozymes of hexokinase have been separated by ion exchange chromatography (22) or electrophoresis (23) from extracts of various mammalian tissues. Due to their localization, their kinetic behavior is also adapted. Hexokinases I, II and III are mainly localized in the brain, muscle and erythrocytes, and are inhibited by the product glucose-6-phosphate. Isozymes I - III are referred to as "low-\(K_m\)" isozymes because of their high affinity for glucose (Table 1.1).

Table 1.1 Kinetic characterization and localization of different human hexokinase isozymes

<table>
<thead>
<tr>
<th>Tissue distribution</th>
<th>HK-I(^a)</th>
<th>HK-II(^b)</th>
<th>HK-III(^c)</th>
<th>HK-IV(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinetics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(k_{cat}) (s(^{-1}))</td>
<td>100</td>
<td>318</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>(K_{0.5, \text{glucose}} ) (M)</td>
<td>6.1 x 10(^{-5})</td>
<td>3.4 x 10(^{-4})</td>
<td>3.8 x 10(^{-5})</td>
<td>7.6 x 10(^{-3})</td>
</tr>
<tr>
<td>(k_{cat}/K_{m, \text{glucose}}) (M(^{-1}) s(^{-1}))</td>
<td>1.7 x 10(^{6})</td>
<td>0.9 x 10(^{6})</td>
<td>4.2 x 10(^{6})</td>
<td>8.2 x 10(^{3})</td>
</tr>
<tr>
<td>(K_{m, \text{ATP}}) (M)</td>
<td>1.2 x 10(^{-3})</td>
<td>1.0 x 10(^{-3})</td>
<td>3.1 x 10(^{-3})</td>
<td>2 x 10(^{-3})</td>
</tr>
<tr>
<td>(K_i, \text{glu-6P}) (M)</td>
<td>1.5 x 10(^{-5})</td>
<td>2.1 x 10(^{-4})</td>
<td>1.3 x 10(^{-4})</td>
<td>no inhibition</td>
</tr>
<tr>
<td>Release of inhibition by Pi</td>
<td>yes</td>
<td>none</td>
<td>none</td>
<td>-</td>
</tr>
<tr>
<td>Cooperativity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperativity for glucose</td>
<td>none</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>Cooperativity for ATP</td>
<td>none</td>
<td>negative</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Cooperativity for Glu-6P</td>
<td>negative</td>
<td>negative</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

*Data taken from \(^a\) (24), \(^b\) (25), \(^c\) (26), \(^d\) (27)*

Hexokinase I is the main isozyme in the human brain (28), but it is also expressed in muscle and erythrocytes. HK-I displays a low \(K_m\) for glucose similar to isozyme III (Table 1.1), a high \(K_{m, \text{ATP}}\) compared with HK-IV, and tight inhibition by glucose-6-phosphate, which is abolished via addition of \(P_i\). The increase in HK-I activity with an increased ratio between \([P_i]\)
and [G6P], along with its vast distribution along the mammalian tissues, led to the suggestion that HK-I plays a catabolic role (to produce ATP) (29, 30).

HK-II plays an important role in muscle, which has the continuous need for glycolysis to support work, and thus displays the highest turnover rate, $k_{\text{cat}}$ (Table 1.1). The negative cooperativity for glucose was speculated to allow this enzyme to respond to conditions that result in high levels of glucose uptake. Its localization, in the insulin sensitive tissues (skeletal muscle; cardiac muscle and adipose tissue) as well as the response to G6P and $P_i$ led Wilson et al., (30, 31) to postulate an anabolic role for HK-II. An interesting kinetic adaptation is the unique negative cooperativity with respect to ATP observed in HK-II.

HK-III is important in red blood cells. Based on similarities with HK-II, Wilson et al. suggested that isozyme III could be associated with anabolic pathways as well (32). The enzyme has modest positive cooperativity for glucose, and it is the only isozyme shown to be inhibited by its own sugar substrate at high glucose concentrations (26, 33). The low cooperativity for glucose is rationalized (34) as necessary due to low glycolytic activity in erythrocytes compared to muscle or brain. Thus a more sensitive glucose response was necessary, as glucose concentrations in red blood cells is about 0.5 mM.

In contrast to HK-I and II, which are known to bind specifically and reversibly to the outer mitochondrial membrane via a 15-residue hydrophobic sequence at the N-terminus, HK-III has lost this ability (35-37). Association with the mitochondrial membrane has an important regulatory function with respect to the use of ATP resulting from the oxidative pathways inside the mitochondrion.

HK-IV is mainly expressed in the liver and pancreas and at lower levels in the hypothalamus and brainstem, in enteroendocrine cells and in gonadotropes and thyrotropes of the pituitary gland (38-40). Glucokinase (HK-IV) expressed in the pancreatic β-cell is responsible for maintaining constant levels of glucose (5 mM) throughout the whole body by controlling the amount of insulin that is being released into the bloodstream. Increased ATP/ADP resulting from glycolysis triggers a series of signaling pathways, including inhibition of $K_{\text{ATP}}$ channels and activation of voltage-dependent $\text{Ca}^{2+}$ channels, which increases intracellular $\text{Ca}^{2+}$ concentration. This stimulates insulin release into the bloodstream from the granules present at the plasma membrane. In adipose and muscle cells, insulin stimulates glucose uptake and metabolism by triggering the translocation of the glucose transporter isoform GLUT4 from intracellular storage vesicles to the plasma membrane. In the liver, glucokinase facilitates glucose utilization and storage as glycogen, which is stimulated by insulin and thus decreases the blood glucose concentration (41, 44).
The rate of phosphorylation in pancreatic β-cells and the liver by GK depends on the concentration of glucose and not on the glucose-6-phosphate needed; thus, there is no feedback inhibition by the product. The small amounts of glucose-6-phosphate that are needed are mainly satisfied by the basal levels of glucokinase or the other hexokinases present. Because the main function of glucokinase is to maintain whole body homeostasis, GK has to display high sensitivity to the physiological glucose concentration (4 - 10 mM). Sigmoidal kinetics yields a higher sensitivity to glucose concentration than does hyperbolic kinetics, and explains the physiological importance of cooperativity.

Structurally, HK I-III are 100 kDa enzymes, constructed from two similar 50 kDa halves, suggesting that they evolved via duplication and fusion of a gene encoding an ancestral enzyme of 50 kDa (35). A structure-function analysis of the N-terminal and C-terminal domains in isozymes I-III adds interesting aspects to their regulatory behavior. HK-II has preserved the catalytic activity in both domains, suggesting that it may be the result of an early duplication event of a 50 kDa glucokinase without extensive refinement (25). The loss of catalytic ability of the N-terminal domain, and its subsequent optimization to acquire membrane binding ability, might indicate that HK-I and III were further evolved after the initial duplication event (34). The existence of a regulatory G6P site in the N-terminal domain in HK-I allows allosteric modulation of catalytic activity in the C-terminal domain via interdomain contacts, and this mode of regulation was believed to be a consequence of refinement after the gene duplication event.

![Figure 1.3. The link between the kinetic properties of hexokinases I-IV found in rat and their physiological roles. The shaded region emphasizes the range of glucose concentrations in the blood. Insets show the different kinetic behavior of the isozymes with respect to glucose phosphorylation: HK I and II display Michaelis-Menten kinetics, HK-III is inhibited by high glucose concentrations, HK-IV shows kinetic cooperativity with Hill coefficient of 1.7. Figure adapted from Cardenas et al., (1998) (47).](image)
1.3 Human Glucokinase: a Monomeric Enzyme with Positive Cooperativity

Glucokinase (HK-IV, hexokinase D) is a monomeric enzyme whose activity is allosterically regulated by its substrate glucose. The phosphorylation of glucose by glucokinase is the rate-limiting step of glucose metabolism in the human liver and pancreas, which makes the enzyme the central regulator of glucose homeostasis (42). The ATP generated after G6P undergoes glycolysis initiates a series of cellular events leading to insulin secretion (48). Genetic lesions in the human glucokinase gene resulting in decreased rates of glucose phosphorylation, and thus hyperglycemia, lead to maturity onset diabetes of young 2 (MODY2) (49). Over-stimulation of glucokinase activity via activating mutations leads to Hyperinsulinemia of Infancy (HI), which arises as a result of oversecretion of insulin under hypoglycemic conditions (50). These disease states emphasize the importance of precisely regulating glucokinase activity in vivo. The steady state velocity of the glucokinase-catalyzed reaction is not hyperbolic, but rather is described by the Hill equation with an exponent of 1.7-1.8. The glucokinase-catalyzed rate follows Michaelis-Menten kinetics with respect to the concentration of MgATP\(^2\) (51). The slight positive cooperativity as a function of glucose concentration is intriguing since GK functions as a monomer (52, 53).

![Figure 1.4 Steady state kinetics of human pancreatic glucokinase fit best to the Hill equation rather than the Michaelis-Menten equation.](image)
The classical models developed to understand cooperativity involve the existence of multiple subunits (54-56) on a polypeptide chain, and were proposed to explain a thermodynamic effect. The lack of additional sites for glucose binding, as well as the observed kinetic effect, allowed two models to be postulated to describe the kinetic cooperativity in GK – the mnemonical model and the ligand induced slow transition model. The mnemonical model was developed by Richard, Meunier and Buc (57) based on an earlier concept of regulatory monomeric enzymes by Rabin (58) and has the following features:

1) The monomeric enzyme exists in two different conformations in the absence of the ligand with an equilibrium constant that strongly favors one conformation.

2) Binding of a ligand induces a different conformation than the initial thermodynamically favored conformation, which is able to undergo catalysis. Only one catalytically competent enzyme exists.

3) The last product stabilizes the active conformation following catalysis; thus the enzyme remembers, for a short time, the conformation that is active for catalysis. This conformation, being less stable, will slowly relax to the alternate, more thermodynamically stable conformation.

The emphasis of the mnemonical model is upon the slow conformational relaxation step that follows catalysis. The catalytic step and the release of products are faster than the equilibration between the two enzyme conformations, thus explaining the appearance of kinetic cooperativity. Another important feature is that only one binary complex exists in this case, thus one catalytic cycle.

Simulations of rate constants available to describe the mnemonical model include the calculation of deviation from linearity in the reciprocal plot ([E₀]/v versus 1/[S]) and thus are based on the steady-state velocity equation. Richard et al., was the first to simulate this deviation and to apply the predictions to the case of wheat germ hexokinase L1 (57, 59- 61). The extent of cooperativity does not depend on the equilibrium constant between the conformations, but instead upon the on rates of substrate binding (Figure 1.5).
Figure 1.5 Computer simulations of the steady-state velocity done by Richard and colleagues for the case of a bisubstrate enzyme displaying positive cooperativity. (The simulation outputs are reciprocal plots ([(E_0)v versus 1/[S]]) and show a lag before steady-state is achieved). The values have arbitrary units and are displayed only for a comparison of their magnitude. Positive cooperativity arises due to fact that the rate of forward glucose binding to E is larger than the rate of glucose binding to E' (57).

Initial evidence to support the mnemonical model for the description of glucokinase sigmoidal kinetics came from early studies on the rat liver glucokinase (51, 62-64). Based upon inhibition studies with product and substrate analogs, the observation of decreased glucose cooperativity at low MgATP^2- concentration was interpreted by Storer et al., (62) in terms of non-equilibration of two enzyme forms in the absence of any ligand under steady-state conditions. Pollard-Knight et al. (65) tested one of the predictions of the mnemonical model applied to glucokinase: glucose cooperativity depends on the ability of MgATP^2- to react rapidly with the binary GK-glucose complex. As expected, they found a loss in cooperativity when a poor nucleotide (Mg ITP^2-) replaced MgATP^2-.

Structural studies by Kamata and colleagues (46), using X-ray crystallography, identified two conformations of human hepatic glucokinase. The structure in the absence of any ligand was termed super-open and was attributed to the thermodynamically favored conformation. Addition of glucose and the synthetic allosteric activator (N-Thiazol-2-yl-2-amino-4-fluoro-5-(1 methylimidazol-2-yl) thiobenzamide) produced GK crystals whose structure was assigned as the closed conformation, similar to the presumed structure of GK in the ternary complex (glucose and ATP bound). Based on structural similarities with hexokinase I, Kamata et al., (46) postulated the existence of an open conformation that is intermediate in structure between the closed and the super-open forms. These newly identified structures were used to map the GK kinetic features in the mnemonical model as schematically represented in Figure 1.6.
Figure 1.6 Crystal structure of unliganded GK (super-open) and liganded (in the presence of glucose and the allosteric activator). In black and blue (sticks) the allosteric activator and glucose are displayed, respectively. The structurally uncharacterized open form is assumed to be in the absence of any ligands and to resemble the C-terminal structure of HK-I in the absence of ligands. Slow and fast steps are predicted based on the degree of structural rearrangements. Pymol was used to generate the picture using the PDB entries 1V4T and 1V4S (148).

The transition between the super-open and closed conformations of glucokinase was assigned as the slow step based on the high order of structural rearrangements that occur, while the open-closed transition was considered fast and reversible. In accordance with the theoretical predictions of the mnemonic model, the super-open, thermodynamically favored conformation would exist in the absence of glucose and would interconvert with the open form more slowly than the catalytic cycle. The open form has high affinity for glucose and upon glucose binding undergoes a fast catalytic cycle. At high glucose concentrations, the open form, realized following the release of glucose-6-phosphate and ADP, would be locked into a fast cycle, while at low glucose concentrations (below $K_{0.5}$) the open form would relax to the super-open conformation thus creating a slow catalytic cycle. This model postulates that cooperativity arises from the differential ratio of the two cycles, which involves formation of a single ternary complex.

The mnemonic model was also in close agreement with the pre-steady-state measurements of glucose and 2-deoxy-glucose binding to GK and Y214C GK done by Heredia et al. (66). The calculated microscopic rate constants for glucose binding are indicated in Figure 1.7. Two major assumptions are present in the model. The $E'$ to $E$ relaxation is set at 1 s$^{-1}$ and considered irreversible, and binding of glucose to the high affinity state $E'$ is 20 000 M$^{-1}$ s$^{-1}$ so that it is faster than the turnover rate.
Figure 1.7 Kinetic model for glucose binding to human pancreatic GK as proposed by Heredia et al., based on KIMSIM simulations. The rate of glucose binding to the high affinity state (E) was assumed to be 20000 M$^{-1}$s$^{-1}$ so that $k^* [S] >> k_{cat}$ (1000 s$^{-1}$ >> 38 s$^{-1}$ for 50 mM glucose concentration). $k_{relax}$ was also assumed to be 1 s$^{-1}$ thus, making the relaxation of E to E' almost irreversible (66).

The isomerization of E'G to EG occurs with a forward rate constant of 0.45 s$^{-1}$, which is 85-fold lower than the $k_{cat}$ value. Based on this model, the conformational change after the initial collision event is considered to happen outside the catalytic cycle.

The second theoretical model, developed by Neet and Ainslie, (67) is more general and has the distinctive assumption that two conformations of the enzyme can undergo catalysis at different rates. The steady state velocity is the sum of the rates of both cycles. The emphasis of the ligand induced slow transition model (LIST) is placed upon the existence of slow conformational changes upon ligand binding that occur more slowly than $k_{cat}$, the non-equilibration of substrate binding due to this slow conformational change, and the change in the preferred catalytic route with changes in glucose concentrations. Neet and Ainslie used the LIST model to simulate ten cases where positive and negative cooperativity exists, with bursts and lags observed in the approach to steady-state. We have schematically represented the positive cooperativity with a lag case, which is applicable to glucokinase, in the Figure 1.8.
Any perturbation of the rate constants that results in equilibration of the species E, E', EG and EG', would eliminate cooperativity. Ligand binding is faster than the rates of conformational transitions and is considered to be at equilibrium. At different ligand concentrations, different catalytic cycles are operational (67-69). The existence of a slow transition was demonstrated using fluorescence studies on the rat liver glucokinase with a half life of 30 seconds (70, 71).

Two recent reports of transient state kinetic binding provide strong support for a preexisting equilibrium between two or more conformations of human glucokinase, which supports the LIST model.
Kim *et al.* (72) modeled the biphasic behavior of the observed rate with glucose concentration, as illustrated in Figure 1.9. In their model, Kim assumes the preexisting equilibrium between $E$ and $E'$, which have different affinities for glucose, and uses this preexisting equilibrium to explain the exponential decrease in the observed rate with the increase in glucose concentrations up to 3 mM followed by a linear increase. The tight binding species $E$ displays a $K_D$ of 90 $\mu$M while the low affinity state $E'$ has a $K_D$ of 73 mM.

Very recently Antoine and colleagues (73) found that transient state glucose binding traces take longer to reach equilibrium (>100 seconds) and are best described by a sum of four exponents. The phase with the highest amplitude exhibited the same biphasic behavior as the observed rate previously determined by Kim *et al.*, (72). They introduced the idea that glucokinase has to sample multiple conformations between the super-open and the closed states and proposed a model that qualitatively includes a minimum of three intermediate conformers. The results of their simulations are shown in Figure 1.10.

![Figure 1.10](image.png)

*Figure 1.10. The Antoine et al., model includes a preexisting equilibrium to describe the transient state glucose binding traces. The kinetic values shown above are obtained from the simulated dependence of $k_{obs3}$ with increasing glucose concentration (73).*

Lately, more experimental evidence has been obtained that supports the LIST model. As outlined in this section of the introduction, the distinction between the models relies upon the number of GK conformations able to undergo catalysis and the relationship between the microscopic rate constants. The mnemonic model clearly states that the $K_{eq}$ between $E$ and $E'$ states has to favor one conformation in the absence of ligand and that both unliganded conformations lead to the same binary complex. The number of conformations undergoing
catalysis, (one or two), which is the clear distinction between the models, has not been investigated.

1.4 Sugar Specificity of Glucokinase

Although called glucokinase, human GK is not a true glucokinase, as defined for the bacterial glucokinases as an enzyme with high specificity for glucose. The accepted name based upon the sugar specificity of GK would be Hexokinase IV or D. Although both names have been frequently encountered, we use the term glucokinase to emphasize its role in glucose homeostasis and the unique kinetic features observed during glucose phosphorylation. Sugar specificity, $k_{\text{cat}}/K_{0.5}$, the analogous constant to $k_{\text{cat}}/K_m$, measured by Xu et al., (74), shows that glucokinase phosphorylates other carbohydrates as well, in the following order: glucose = mannose > deoxyglucose > fructose = glucosamine (the $k_{\text{cat}}/K_{0.5}$ values in M$^{-1}$ s$^{-1}$ are 11000, 12000, 3000, 490, and 330, respectively). Based on carbohydrate structural similarities, Xu et al., (74) suggest that the 1-OH is critical for sugar binding and phosphorylation ($K_{0.5}$ of methyl-$\alpha$-glucopyranose or methyl-$\beta$-glucopyranose were $> 1$ M and 2, 5-anhydroglucitol was not a substrate). The orientation and hydrogen bonding capabilities of the 3-OH and 4-OH were also critical based on the fact that allose and galactose, which differ from glucose only in the orientation of the 3-OH or the 4-OH, were not found to be substrates. Flexibility in both the existence and orientation of the 2-OH position was observed, due to the ability of GK to phosphorylate mannose, deoxy-glucose, and glucosamine.

Using intrinsic fluorescence Zelent et al. (75) found the following preference of sugar binding to GK: N-acetyl-D-glucosamine > D-glucose > D-mannose > D-mannoheptulose > 2-deoxy-D-glucose >> L-glucose.
1.5 Aims of the Study

During the natural evolution of new enzyme activities, the initial stages of the evolutionary process often include the recruitment of a promiscuous catalyst, followed by the optimization of the needed activity via sequential rounds of mutational drift and natural selection. The evolutionary processes responsible for the optimization of promiscuous catalysts are typically too slow for scientists to monitor in real time. Thus, directed evolution studies that mimic the natural evolutionary process promise to overcome this problem. Mechanistic investigation of the intermediates generated in a superfamily of enzymes during evolution might provide clues to any common structural features that are targeted during alteration of substrate specificity. One aim of this study is to understand the molecular mechanism(s) responsible for specialization of function in a family of modern sugar kinases found to possess substrate ambiguity. This aim is accomplished in Chapter 2 of the thesis. Based on the postulate that common mechanisms can be identified in a superfamily of protein catalysts that share a common ancestor and possess an overlapping alternate function, we combined random mutagenesis and in vivo functional selection, to identify two structurally overlapping mutational “hot spots” in the AlsK and NanK sugar kinase scaffold.

The main interest of my research is focused on understanding the modes by which kinetic cooperativity can be achieved in a monomeric enzyme. This interesting question has origins in the 1970s when new models were proposed for the newly discovered case of a polypeptide with a single active site whose kinetics deviate from classical Michaelis-Menten behavior. Thus, the second aim of the dissertation is to understand the mechanistic bases for the kinetic cooperativity in human glucokinase. More specifically, our goal is to provide a qualitative working kinetic model from which we can extend our investigations via other biophysical techniques. Our current understanding of the cooperativity with respect to a particular kinetic model is described in Chapter 3. Preliminary HSQC-NMR and analytical ultracentrifugation experiments aimed at describing the sample composition of human glucokinase are discussed in Chapter 4.

The third aim of the dissertation is to provide a molecular description of the conformational transitions that give rise to cooperativity in human glucokinase. In Chapter 5 we focus on the role of the C-terminal secondary structural element, helix α13, to the kinetic cooperativity of human glucokinase. Inspection of the crystal structures in the super-open and
closed forms suggests that helix α13 plays a role in mediating the conformational transitions responsible for generating kinetic cooperativity. We were interested in investigating the extent to which helix α13 contributes to the unique kinetic characteristics of human glucokinase and to establish a link between the primary sequence of this structural element and the dynamics of glucokinase conformational rearrangements.

**Chapter 6** summarizes our findings with respect to the evolutionary origins of specificity, current advances in the elucidation of a specific model to describe the kinetic cooperativity of human pancreatic glucokinase and also outlines future directions of the project.
CHAPTER 2

BACTERIAL GLUCOKINASES - DIVERGENT EVOLUTION OF FUNCTION IN THE ROK-SUPERFAMILY

ABSTRACT

The D-allose and N-acetyl-D-mannosamine kinases of Escherichia coli K-12 are divergent members of the functionally diverse ROK (repressor, open reading frame, kinase) superfamily. Previous work in our laboratory has demonstrated that AlsK and NanK possess weak phosphoryl transfer activity toward the alternate substrate D-glucose. To gain insight into the evolutionary mechanisms that fuel the specialization of individual enzyme function, experimental laboratory evolution was conducted to improve the glucokinase activities of AlsK and NanK. Error-prone PCR was combined with in vivo functional selection in a glucokinase-deficient bacterium to identify four independent single nucleotide substitutions in the alsK and nanK genes that improve the glucokinase activity of each enzyme. The most advantageous substitutions, L84P in NanK and A73G in AlsK, enhance the $k_{cat}/K_m$ values for phosphoryl transfer to glucose by 12-fold and 60-fold, respectively. Both substitutions co-localize to a variable loop region located between the fourth $\beta$-sheet and the second $\alpha$-helix of the ROK scaffold. A multiple sequence alignment of diverse ROK family members reveals that the A73G substitution in AlsK recapitulates a conserved glycine residue present in many ROK proteins, including some transcriptional repressors. Steady-state kinetic analyses of the selected AlsK and NanK variants demonstrate that their native activities toward D-allose and N-acetyl-D-mannosamine are largely unaffected by the glucokinase-enhancing substitutions. Substrate specificity profiling reveals that the A73G AlsK and L84P NanK variants display systematic improvements in the $k_{cat}/K_m$ values for a variety of nonnative carbohydrates. This finding is consistent with an evolutionary process that includes the formation of intermediates possessing relaxed substrate specificities during the initial steps of enzyme functional divergence.
2.1 Introduction

Modern experimental techniques in molecular biology and microbial genetics have made possible the rapid evolution of protein catalysts within the laboratory in a manner that closely resembles the process of natural divergent evolution (76-79). These methods offer a unique opportunity to identify underlying principles that govern which features of a progenitor catalyst are targeted for alteration during the specialization of enzyme function. A detailed investigation of laboratory-based experimental enzyme evolution promises to reveal common mechanisms for the generation of new catalytic activities during sequential rounds of mutational drift and natural selection. In addition, the ability to understand how single amino acid substitutions within a given enzyme scaffold stimulate changes in individual kinetic constants should provide unique insight into the linkage between a protein's three-dimensional structure and its biological function. If common structural and/or kinetic features are targeted by evolution during the specialization of function, one might expect these to be most easily identified in a superfamily of protein catalysts that share a common ancestor and possess an overlapping alternate function (80, 81).

Recently, our laboratory has discovered a series of divergent bacterial sugar kinases that belong to the ROK superfamily (Pfam 00480), which fits this description (18, 19). The ROK superfamily represents a functionally diverse group of prokaryotic proteins that includes carbohydrate-responsive transcriptional repressors and sugar kinases (82). ROK transcriptional repressors are signified by a conserved N-terminal helix-turn-helix motif that affords DNA binding, whereas ROK kinases contain an N-terminal ATP-binding motif denoted by the sequence DXGXT (83). Presently, the Pfam database contains more than 1600 ROK members, many of which possess unknown physiological functions. Within the last three years, the crystal structures of several ROK polypeptides have been determined. Included among these are unliganded structures of *Escherichia coli* N-acetyl-D-mannosamine kinase (PDB 2AA4) and *Salmonella typhimurium* N-acetyl-D-glucosamine kinase (PDB 2AP1), as well as the structure of the *Escherichia coli* transcriptional repressor Mlc (PDB 1Z6R) (84-86). Notably, the unique inorganic polyphosphate/ATP-glucomannokinase from *Arthrobacter* sp. strain KM is the only ROK kinase whose structure has been determined with a carbohydrate bound at the active site (PDB 1WOQ) (87). This enzyme lacks the CXCGXXGC sequence motif present in the founding members of the ROK family first described by Saier and colleagues, but it retains the core ROK family signature pattern (88). The crystal structure of *Escherichia coli* glucokinase has also been recently determined in the presence of glucose; however, this enzyme is distantly related to
other ROK proteins and it is not currently classified as a member of this superfamily (89). The *Escherichia coli* K-12 genome harbors the coding sequences for four ROK sugar kinases: alko kinase (AlsK) (1, 2), N-acetyl-D-mannosamine kinase (NanK) (3, 4), N-acetyl-D-glucosamine kinase (NagK) (5), and manno(fructo)kinase (Mak) (6). These four enzymes are derived from a common ancestor and each possesses a low level of phosphoryl transfer activity toward the alternate substrate, D-glucose (Figure 1.2). The latent glucokinase activities of these enzymes, which range from ~100 to 1000 M$^{-1}$ s$^{-1}$, are several orders of magnitude lower than the catalytic efficiency displayed by native *Escherichia coli* glucokinase (18, 19). In contrast, the $k_{cat}/K_m$ value of each enzyme, acting upon its native substrate exceeds 10$^5$ M$^{-1}$ s$^{-1}$, as expected for a physiologically relevant activity. How differences in substrate specificity have arisen within the ROK superfamily of bacterial sugar kinases, given the high level of structural similarity between the native carbohydrate substrates of each member (Figure 1.2), is unknown.

The evolutionary origin of enzyme specificity is of particular interest to our laboratory. Most contemporary enzymes possess the ability to recognize subtle structural differences in individual substrate molecules, often at the single-atom level, and to transform these differences into large variations in catalytic efficiency. Indeed, substrate selectivity is a distinguishing characteristic of protein catalysts, one that is largely unrealized in the world of synthetic small-molecule catalysts (90). Specificity is defined by the ratio of second-order rate constants, $k_{cat}/K_m$, that govern the transformations of two competing substrates. Several models have been put forth to explain the structural, thermodynamic, and kinetic basis for enzyme specificity, often amid much discussion (91, 92). One particularly controversial issue is the extent to which enzyme conformational changes participate in substrate discrimination (93-95). A consensus has yet to be reached on this issue (96-98); however studies aimed at uncovering the evolutionary origins of substrate discrimination in protein catalysts may provide additional insight.

In this chapter, we probe evolutionary events that occur during the optimization of glucokinase activity in two members of the ROK superfamily, AlsK and NanK, which share only 21% sequence identity with one another. Following random mutagenesis and *in vivo* functional selection, we identify two structurally overlapping mutational “hot spots” in the sugar kinase scaffold. Steady-state kinetic analyses of the selected variants demonstrate that the native activities of AlsK and NanK are largely unaffected by the glucokinase-enhancing substitutions. Furthermore, the variants have acquired an increased ability to phosphorylate a variety of nonnatural carbohydrate substrates as a result of the evolutionary process. These findings provide insight into the determinants of substrate specificity in the ROK superfamily and support
previous reports suggesting that alterations in enzyme-substrate selectivity proceed through nonspecialized intermediates (99, 100).

2.2 Materials and Methods

2.2.1 Random Mutagenesis of alsK and nanK

Error-prone PCR of the native alsK and nanK genes was carried out with the Gene Morph II random mutagenesis kit (Stratagene) using the following oligonucleotide primers: AlsK forward: 5’-CTT TAA GAA GGA GAT ATA CCA TG-3’; AlsK reverse: 5’-GGT GCT CGA GTG CGG CCG-3’; NanK forward: 5’-GGC GGT GCC ATG GCC ACA CTG GCG ATT GAT ATC G-3’; NanK reverse: 5’-CTG TAC TTC ACC CAT CCG GCC GAT TTT TCT CCC-3’. 1 mg of template plasmid (pBGM101-alsK or pBGM101-nanK) (18, 19) was used in each reaction and 25 amplification cycles were performed to achieve an average mutation rate of 1 base pair change per gene. Following amplification, the reaction mixture was treated with DpnI (20 U) for 1 h at 37°C to digest parental DNA and the products were purified using the Promega PCR clean-up kit. The resulting mutagenized product was digested with EagI (5 U) and NcoI (5 U) for 2 h at 37°C, purified again with the PCR clean-up kit, and added to a ligation reaction containing appropriately digested pBGM101.1 vector (1 mg) and T4 DNA ligase (40 U). The ligation reaction was incubated overnight at 17°C, and then the reaction was quenched by heating at 70°C for 10 min. The ligation products were precipitated via the addition of sodium acetate (0.3 M) and ethanol (70%) with incubation at -20°C for 1 h. The reaction mixture was centrifuged at 16,000g for 10 min to collect the DNA pellet, which was subsequently washed with 200 mL of ice-cold ethanol (70%). Following a second centrifugation step at 16,000g, the pellet was isolated and resuspended in 15 ml of 37°C nuclease free water.

To evaluate the quality and mutational frequency of the alsK and nanK libraries, a small aliquot of each library was transformed into electrocompetent BM5340(DE3) cells via electroporation. Transformed cells were plated on Luria-Bertani plates containing ampicillin (150 μg/mL) chloramphenicol (25 μg/mL) and kanamycin (40 μg/mL) and single colonies were isolated following overnight incubation at 37°C. NcoI and EagI double digests of mini-prep DNA prepared from individual clones demonstrated that more than 80% of library members possessed an insert. A total of 18 clones were selected from rich LB plates and the plasmid
DNA from each was sequenced to reveal a mutation frequency of 0.9 ± 0.6 bps per gene. Serial dilutions of transformed BM5340(DE3) cells were plated on antibiotic supplemented Luria-Bertani plates to yield estimated library sizes of 7.3 x 10^6 transformants for the alsK library and 8.6 x 10^6 transformants for the nanK library. These library sizes are sufficiently large to ensure that all possible single base pair changes are sampled in each library.

### 2.2.2 Genetic Selection Experiments

To select for amino acid substitutions that improved the glucokinase activity of AlsK and NanK, error-prone PCR libraries were transformed into electrocompetent BM5340(DE3) cells. Following recovery at 37°C for 1.5 h in SOB media supplemented with glucose (20 mM) and MgCl₂ (20 mM), transformed cells were pelleted by centrifugation at 3000g, washed twice with 1 mL of M9 minimal media and plated on M9 minimal plates containing glucose (0.005% w/v), ampicillin (150 μg/mL), chloramphenicol (25 μg/mL) and kanamycin (40 μg/mL). Selection plates were also supplemented with IPTG at a final concentration of either 50 μM for NanK selection experiments or 10 μM for AlsK selection experiments to induce protein production. The total fraction of complementing clones was 0.128% for the nanK library and 0.064% for the alsK library. Approximately 1% of the total complementing clones from each library afforded colonies within 2 days of growth at 37°C on glucose supplemented M9 minimal plates, and these subpopulations were selected for further study. Control experiments verified that expression of the wild-type nanK and alsK genes does not produce colonies within this time frame under the selective growth conditions described above, and the wild-type genes were never recovered in the library selection experiments. Mini-prep DNA prepared from individual colonies was retransformed into electrocompetent BM5340(DE3) cells and the selection was repeated as described above to verify the growth phenotype of primary selectants. The fastest growing library members from this second round of selection were submitted for sequencing.

### 2.2.3 Enzyme Expression and Purification

Enzymes were produced in the glucokinase-deficient *Escherichia coli* strain BM5340(DE3) to avoid contamination with endogenous glucokinases. Cultures were inoculated to an initial OD₆₀₀nm of 0.01 and cells were grown in Luria-Bertani media at 37°C containing ampicillin (150 μg/mL), chloramphenicol (25 μg/mL) and kanamycin (40 μg/mL) until the OD₆₀₀nm reached 0.9. IPTG was added to a final concentration of 1 mM to induce gene expression and cultures were grown for an additional 2-3 hours at 37°C. Bacteria were harvested by centrifugation at 7000g for 15 min and the resulting cell pellet (3 g) was resuspended in 12 mL
of buffer A, containing sodium phosphate (75 mM, pH 7.6), glycerol (5% w/v) and imidazole (25 mM). Cell extracts were prepared by passage through a model 110L Microfluidizer (Microfluidics Corp) and the resulting cell lysate was cleared by centrifugation at 15000g for 30 min at 4°C. The supernatant was syringe-tip filtered (0.2 μm), and then loaded onto a 5 mL HisTrip FF crude column that had been previously equilibrated with buffer A. Following loading, the column was washed with ten additional column volumes of buffer A containing 40 mM imidazole. Purified enzymes were eluted in buffer A containing 250 mM imidazole and were dialyzed overnight against two sequential 2 L aliquots of Tris-HCl buffer (25 mM, pH 7.6) containing MgSO₄ (5 mM), glycerol (5% w/v) and DTT (0.5 mM). Protein concentrations were estimated from absorbance readings at 280 nm using previously determined molar extinction coefficients of 15,000 M⁻¹ cm⁻¹ for NanK and 34,900 M⁻¹ cm⁻¹ for AlsK (18, 19). The purity of all enzymes was judged to be ≥ 95% on the basis of polyacrylamide gel analysis.

2.2.4 Enzyme Assays

The glucokinase activity of individual enzymes was measured by coupling the production of glucose 6-phosphate to the reduction of NADP via glucose 6-phosphate dehydrogenase (34). To determine apparent \( K_m \) values for glucose, reaction mixtures contained Tris (0.20 M, pH 7.6), NADP (0.50 mM), DTT (1.0 mM), ATP (7.5 mM), MgCl₂ (8.5 mM) and G6PDH (7.5 U). Enzyme activity with allose, altrose, 2-deoxyglucose, mannose and N-acetyl-D-mannosamine as substrates was measured by coupling the production of ADP to the oxidation of NADH via the combined action of pyruvate kinase (15 U) and lactate dehydrogenase (25 U) (35). Assays contained Tris (0.2 M, pH 7.6), NADH (0.5 mM), DTT (1 mM), PEP (5 mM), ATP (5-25 mM), MgCl₂ (6-26 mM) and KCl (5 mM). To determine the apparent \( K_m \) values for ATP, the same experimental conditions were employed at saturating concentrations of carbohydrate. Assays were conducted at 25°C and were initiated by addition of the enzyme under investigation. Data were fitted to standard Michaelis-Menten kinetic equations with each data point representing the average of at least two individual rate determinations. Assays of wild-type and variant NanK acting upon 1,5-anhydroglucitol were conducted using the lactate dehydrogenase/pyruvate kinase linked assay at saturating ATP concentrations (25 mM) and subsaturating concentrations of carbohydrate. The \( k_{cat}/K_m \) value was determined from the slope of the linear plot of the rate of NAD production as a function of 1, 5-anhydroglucitol concentration.
2.2.5 Multiple Sequence Alignment and Phylogenetic Analysis

An initial ClustalW (101) alignment of *Escherichia coli* AlsK and NanK was used to produce a HMM profile (102) with HMMerBuild. This HMM profile was then used to search for homologous sequences in UniProt Release 9.6 with HMMerSearch. A natural break in Expectation values was apparent within the results between 1x10^{-76} and 3x10^{-50}. Therefore, in spite of the presence of hundreds of homologous sequences, all sequences with E values higher than 1x10^{-76} were eliminated from the dataset. HMMerAlign was then used to align the remaining dataset to the initial HMM. Nearly redundant sequences were then eliminated and a new HMM profile was built. This new HMM was then used to merge MLC_Ecoli (P50456 E value 1.5x10^{-2} in initial search), YDHR_Bacsu (O05510 E value 2.9x10^{-6} in initial search), PGMK_Arts (Q7WT42 E value 6.9x10^{-4} in initial search), MAK_Ecoli (P23917 E value 5.7x10^{-7} in initial search), NAGK_Ecoli (P75959 E value 7.5x10^{-9} in initial search), and NAGK_Salty (Q8ZPZ9 E value 5.1x10^{-8} in initial search) into the alignment. Finally, this alignment was sent to the Advanced T-Coffee Expresso (103) server to be refined using structural consistency methodology, and then minor manual adjustments were made using the GCG SeqLab editor. The alignment was next masked to eliminate homoplastic columns of amino acids with less than 10% similarity, and exported to FastA format, which was then converted to PHYLIP format with ReadSeq (104). The final data matrix consisted of twenty-five taxa by 311 characters.

ProML (version 3.67) (105) was used to infer the maximum likelihood (ML) tree for the dataset using a JTT protein model (106) with a four discrete rate gamma correction model using an alpha value of 1.00. The optimal model was determined with ProtTest (107). The heuristic search strategy consisted of ten random additions with global tree rearrangements for each replicate. SeqBoot (version 3.67) (105) was then used to produce 100 pseudoreplicate datasets, which were then passed back to ProML for bootstrapped maximum likelihood analysis. An identical evolutionary model and search strategy was employed. The resulting one-hundred bootstrap ML trees were then run through Consense (version 3.67) (105) to produce the majority-rule consensus tree for the bootstrap analysis. The single maximum likelihood tree, available in supporting information, had a negative log likelihood value of 9148.29.

All sequence analysis (with the exception of the format converter ReadSeq (104)) was performed within SeqLab (108) from the GCG (109) package. Phylogenetic analyses were done with the PHYLIP (105) package.
2.3 Results

2.3.1 Co-localization of Selected AlsK and NanK Variants

To investigate the accessible evolutionary pathway(s) for improving the glucokinase activities of *Escherichia coli* AlsK and NanK, error-prone PCR was conducted on the wild-type alsK and nanK genetic templates. In these directed evolution experiments, a low rate of mutation was chosen in an attempt to mimic the single nucleotide changes that likely power the evolutionary refinement of weak promiscuous catalytic activities in natural biological systems. Single amino acid substitutions that increased the glucokinase activity of AlsK and NanK were selected by the ability of their encoded genes to efficiently complement the glucokinase auxotrophy of BM5340(DE3). Two single amino acid substitutions in NanK, L84P and V138M, produced the fastest growth rate of BM5340(DE3) cells on glucose minimal medium. Similarly, two single amino acid variants of AlsK, A73G and F145L, were most effective at complementing the glucokinase deficiency of BM5340(DE3).

A comparison of the primary amino acid sequences of AlsK and NanK reveals that the sites of substitution co-localize to similar positions within the sugar kinase scaffold. A sequence alignment of individual ROK family members, shown in Figure 2.1, indicates that Leu-84 of NanK is located near Ala-73 of AlsK. Similarly, Val-138 of NanK nearly overlaps with the position of Phe-145 in the AlsK primary sequence. When the conserved secondary structural elements of the ROK scaffold were mapped onto the sequence alignment, the glucokinase enhancing substitutions correspond to the same secondary regions. In particular, Leu-84 of NanK and Ala-73 of AlsK reside in a variable loop region that is located between the fourth β-sheet and the second α-helix of the ROK scaffold. Conversely, Val-138 of NanK and Phe-145 of AlsK both reside in the seventh β-sheet. The pair-wise co-localization of these selected substitutions suggests the existence of two "hot spots" within the polypeptide scaffold, which are targeted by evolution during the optimization of glucokinase activity.
Figure 2.1 Multiple sequence alignment of ROK superfamily members. The location of the amino acid substitutions observed in this study are shown on a black background and their identities are depicted in bold text and color. (red for A73G and L84P and green for F145L and V138M). The secondary structural elements of the ROK scaffold are represented above the sequence alignment and conserved residues are shown on a grey background. An asterisk denotes ROK family members whose structures are available in the Protein Data Bank. A description of the construction of this alignment is provided in the Materials and Methods chapter 2.2.5.
2.3.2 Activity of Selected Variants Toward Glucose and Their Native Substrates

To assess the enhancements in glucokinase activity afforded by individual AlsK and NanK variants, kinetic assays were performed on highly purified preparations of each protein. Steady state kinetic analysis of the most active NanK variant, L84P, demonstrates that this substitution increases the second-order rate constant for glucose phosphorylation, \( k_{\text{cat}}/K_{\text{m glucose}} \), by 10-fold, from \( 1.1 \times 10^3 \) M\(^{-1}\) s\(^{-1}\) to \( 1.3 \times 10^4 \) M\(^{-1}\) s\(^{-1}\) (Table 2.1). This increase is achieved by a 3-fold decrease in the \( K_{\text{m}} \) value for glucose and a 4-fold increase in the \( k_{\text{cat}} \) value. The L84P NanK variant also displays a modest 2-fold decrease in the catalytic efficiency towards its native substrate, \( N \)-acetyl-D-mannosamine. V138M, the second NanK variant selected from the error-prone PCR library, displays a 6-fold increase in the \( k_{\text{cat}}/K_{\text{m}} \) value for glucose, while the \( k_{\text{cat}}/K_{\text{m}} \) value for \( N \)-acetyl-D-mannosamine remains unchanged. To investigate the consequences of combining the L84P and V138M substitutions in a single polypeptide, the doubly mutated nanK coding sequence was constructed and the resulting gene product was purified. The \( k_{\text{cat}}/K_{\text{m}} \) value for glucose phosphorylation by the L84P-V138M variant increased 5-fold compared to wild-type NanK, from \( 1.1 \times 10^3 \) M\(^{-1}\) s\(^{-1}\) to \( 5.3 \times 10^3 \) M\(^{-1}\) s\(^{-1}\). Thus, the effects of the individual L84P and V138M substitutions were non-additive.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Parameter</th>
<th>Wild-type NanK</th>
<th>L84P NanK loop region</th>
<th>V138M NanK β-sheet region</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )-acetyl D-mannosamine</td>
<td>( k_{\text{cat/sugar}} ) (s(^{-1}))</td>
<td>( 56 \pm 1 )</td>
<td>( 18 \pm 1 )</td>
<td>( 20 \pm 1 )</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{m/sugar}} ) (M)</td>
<td>( (3.6 \pm 0.3) \times 10^{-4} )</td>
<td>( (2.3 \pm 0.3) \times 10^{-4} )</td>
<td>( (1.6 \pm 0.1) \times 10^{-4} )</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{m/ATP}} ) (M)</td>
<td>( (2.6 \pm 0.1) \times 10^{-4} )</td>
<td>( (8.4 \pm 0.6) \times 10^{-5} )</td>
<td>( (1.1 \pm 0.1) \times 10^{-4} )</td>
</tr>
<tr>
<td></td>
<td>( k_{\text{cat}}/K_{\text{m/sugar}} ) (M(^{-1}) s(^{-1}))</td>
<td>( (1.5 \pm 0.1) \times 10^{3} )</td>
<td>( (7.8 \pm 0.7) \times 10^{4} )</td>
<td>( (1.2 \pm 0.1) \times 10^{5} )</td>
</tr>
<tr>
<td>D-glucose</td>
<td>( k_{\text{cat/sugar}} ) (s(^{-1}))</td>
<td>( 22 \pm 1 )</td>
<td>( 84 \pm 1 )</td>
<td>( 60 \pm 1 )</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{m/sugar}} ) (M)</td>
<td>( (2.0 \pm 0.1) \times 10^{-2} )</td>
<td>( (6.4 \pm 0.1) \times 10^{-3} )</td>
<td>( (8.6 \pm 0.2) \times 10^{-3} )</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{m/ATP}} ) (M)</td>
<td>( (1.2 \pm 0.1) \times 10^{-3} )</td>
<td>( (9.2 \pm 1) \times 10^{-4} )</td>
<td>( (6.3 \pm 1) \times 10^{-4} )</td>
</tr>
<tr>
<td></td>
<td>( k_{\text{cat}}/K_{\text{m/sugar}} ) (M(^{-1}) s(^{-1}))</td>
<td>( (1.1 \pm 0.1) \times 10^{3} )</td>
<td>( (1.1 \pm 0.1) \times 10^{4} )</td>
<td>( (7.0 \pm 0.1) \times 10^{3} )</td>
</tr>
</tbody>
</table>

\( ^{a} \) Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6. \( ^{b} \) Determined by a pyruvate kinase/lactate dehydrogenase coupled assay. \( ^{c} \) Determined by a glucose-6-phosphate dehydrogenase coupled assay (see Materials and Methods).
Steady state kinetic analysis of the most active AlsK variant, A73G, reveals a 60-fold increase in the \( k_{\text{cat}}/K_m \) value for glucose phosphorylation (Table 2.2). The effect of this mutation is exclusively due to an improved \( K_m \) value for substrate glucose, which decreases from 29 mM to 650 \( \mu \)M. The second AlsK variant selected from our randomized library is F145L, an enzyme that displays a 10-fold increase in glucokinase activity, compared with the \( k_{\text{cat}}/K_m \) value of the wild-type enzyme. Similar to the selected NanK variants, the A73G and F145L substitutions in AlsK had no significant effect upon the second-order rate constant for phosphoryl transfer when the native carbohydrate, D-allose, was the substrate.

| Table 2.2: Kinetic Parameters for the Phosphorylation of Native and Ambiguous Substrates by wt-AlsK and Its Selected Variants\(^a\) |
|---|---|---|---|
| substrate | parameter | wild-type AlsK | A73G AlsK putative loop region | F145L AlsK putative \( \beta \)-sheet region |
| D-allose\(^b\) | \( k_{\text{cat}, \text{sugar}} \) (s\(^{-1}\)) | 47 ± 1 | 40 ± 1 | 49 ± 1 |
| | \( K_m, \text{sugar} \) (M) | (1.9 ± 0.1) \times 10^{-4} | (2.0 ± 0.1) \times 10^{-4} | (1.5 ± 0.1) \times 10^{-4} |
| | \( K_m, \text{ATP} \) (M) | (2.7 ± 0.2) \times 10^{-4} | (6.0 ± 0.5) \times 10^{-5} | (1.4 ± 0.1) \times 10^{-4} |
| \( k_{\text{cat}}/K_m, \text{sugar} \) (M\(^{-1}\) s\(^{-1}\)) | (2.5 ± 0.1) \times 10^{5} | (2.0 ± 0.1) \times 10^{5} | (3.2 ± 0.2) \times 10^{5} |
| D-glucose\(^c\) | \( k_{\text{cat}, \text{sugar}} \) (s\(^{-1}\)) | 10 ± 1 | 13 ± 1 | 28 ± 1 |
| | \( K_m, \text{sugar} \) (M) | (2.9 ± 0.2) \times 10^{-2} | (6.5 ± 0.9) \times 10^{-4} | (7.2 ± 0.5) \times 10^{-5} |
| | \( K_m, \text{ATP} \) (M) | (5.4 ± 0.5) \times 10^{-3} | (4.5 ± 0.2) \times 10^{-4} | (9.6 ± 0.1) \times 10^{-4} |
| \( k_{\text{cat}}/K_m, \text{sugar} \) (M\(^{-1}\) s\(^{-1}\)) | (3.4 ± 0.3) \times 10^{2} | (2.1 ± 0.1) \times 10^{4} | (3.9 ± 0.3) \times 10^{3} |

\(^a\) Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6. \(^b\) Determined by a pyruvate kinase/lactate dehydrogenase coupled assay. \(^c\) Determined by a glucose-6-phosphate dehydrogenase coupled assay (see Materials and Methods).

2.3.3 Substrate Specificity Profiles of AlsK and NanK Loop Variants

To understand the basis for the 60-fold increase in glucokinase activity provided by the AlsK loop substitution, A73G, several carbohydrates were tested as potential substrates. Altrose, 2'-deoxyglucose and mannose were found to be substrates for phosphoryl transfer using both wild-type and A73G AlsK. Together with the native substrate, D-allose, this collection provides a group of carbohydrates that are transformed by the wild-type enzyme with a range of catalytic efficiencies that span more than 5 orders of magnitude. As shown in Table 2.3, substitution of Ala-73 with Gly in AlsK enhances the \( k_{\text{cat}}/K_m \) for all non-native substrates tested. This includes a 12-fold improvement in activity toward altrose, a 38-fold increase in activity toward mannose and a 90-fold increase in the \( k_{\text{cat}}/K_m \) value for 2'-deoxyglucose. These improvements are achieved, in part, via a 2-fold, 4-fold and 8-fold decrease in the \( K_m \) values for
mannose, altrose and 2’-deoxyglucose, respectively. Similarly, the A73G loop substitution increases the $k_{\text{cat}}$ value for altrose, 2’-deoxyglucose and mannose by 3-fold, 13-fold and 20-fold, respectively. These observations stand in stark contrast to the unchanged value of $k_{\text{cat}}$ for glucose phosphorylation resulting from the A73G substitution. A summary of the effects of the A73G substitution upon individual catalytic constants for all substrates examined in this study is provided in Figure 2.2.

<table>
<thead>
<tr>
<th>enzyme</th>
<th>substrates</th>
<th>$k_{\text{cat}}$ (s$^{-1}$)</th>
<th>$K_{\text{M}}$ (M)</th>
<th>$k_{\text{cat}}/K_{\text{M}}$ (M$^{-1}$ s$^{-1}$)</th>
<th>fold increase in $k_{\text{cat}}/K_{\text{M}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type AlsK</td>
<td>D-allose</td>
<td>$47 \pm 0.7$</td>
<td>$(1.9 \pm 0.1) \times 10^{-4}$</td>
<td>$(2.5 \pm 0.1) \times 10^{2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-glucose</td>
<td>$10 \pm 0.2$</td>
<td>$(2.9 \pm 0.2) \times 10^{-2}$</td>
<td>$(3.4 \pm 0.3) \times 10^{2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-altrose</td>
<td>$34 \pm 2.2$</td>
<td>$(2.1 \pm 0.3) \times 10^{-1}$</td>
<td>$(1.6 \pm 0.2) \times 10^{2}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2’-deoxy-D-glucose</td>
<td>$1.5 \pm 0.1$</td>
<td>$(3.8 \pm 0.4) \times 10^{-1}$</td>
<td>$4.1 \pm 0.5$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-mannose</td>
<td>$0.5 \pm 0.1$</td>
<td>$(3.9 \pm 0.7) \times 10^{-1}$</td>
<td>$1.2 \pm 0.2$</td>
<td></td>
</tr>
<tr>
<td>A73G AlsK</td>
<td>D-allose</td>
<td>$40 \pm 0.5$</td>
<td>$(2.0 \pm 0.1) \times 10^{-4}$</td>
<td>$(2.0 \pm 0.1) \times 10^{5}$</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>D-glucose</td>
<td>$13 \pm 0.5$</td>
<td>$(6.5 \pm 0.9) \times 10^{-4}$</td>
<td>$(2.1 \pm 0.3) \times 10^{4}$</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>D-altrose</td>
<td>$94 \pm 2.1$</td>
<td>$(4.8 \pm 0.3) \times 10^{-2}$</td>
<td>$(2.0 \pm 0.1) \times 10^{3}$</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>2’-deoxy-D-glucose</td>
<td>$19 \pm 0.5$</td>
<td>$(5.1 \pm 0.4) \times 10^{-2}$</td>
<td>$(3.7 \pm 0.3) \times 10^{2}$</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>D-mannose</td>
<td>$9.5 \pm 0.4$</td>
<td>$(2.1 \pm 0.2) \times 10^{-1}$</td>
<td>$(4.5 \pm 0.4) \times 10^{1}$</td>
<td>37.5</td>
</tr>
</tbody>
</table>

$^a$Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6, by a glucose-6-phosphate dehydrogenase coupled assay for glucose or a pyruvate kinase/lactate dehydrogenase coupled assay for all other sugars (see Materials and Methods)
The effect of the L84P loop substitution upon the specificity profile of native NanK was also measured using two poor substrates, D-mannose and 1,5-anhydroglucitol. It was impossible to achieve saturating concentrations of 1,5-anhydroglucitol with both wild-type and L84P NanK, thus only a value for $k_{cat}/K_m$ is reported for this substrate. Similar to the effects of the AlsK loop substitution, the L84P substitution enhanced the catalytic efficiency of NanK toward both 1,5-anhydroglucitol and D-mannose by 6-fold and 10-fold, respectively. In the case of mannose, this increase stems from a modest decrease in the $K_m$ value and a slight increase in the $k_{cat}$ value (Table 2.4).

<table>
<thead>
<tr>
<th>enzyme</th>
<th>substrate</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (M)</th>
<th>$k_{cat}/K_m$ (M$^{-1}$ s$^{-1}$)</th>
<th>fold increase in $k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type NanK</td>
<td>N-acetyl-D-mannosamine</td>
<td>56 ± 1.0</td>
<td>(3.6 ± 0.3) × 10$^{-4}$</td>
<td>(1.5 ± 0.1) × 10$^5$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-glucose</td>
<td>22 ± 0.5</td>
<td>(2.0 ± 0.1) × 10$^{-2}$</td>
<td>(1.1 ± 0.1) × 10$^3$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D-mannose</td>
<td>25 ± 0.2</td>
<td>(8.4 ± 0.7) × 10$^{-2}$</td>
<td>(3.0 ± 0.1) × 10$^2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,5-anhydro-glucitol</td>
<td>n.d.</td>
<td>n.d.</td>
<td>1.7 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>L84P NanK</td>
<td>N-acetyl-D-mannosamine</td>
<td>18 ± 0.4</td>
<td>(2.3 ± 0.3) × 10$^{-4}$</td>
<td>(7.8 ± 0.7) × 10$^4$</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>D-glucose</td>
<td>84 ± 0.2</td>
<td>(6.4 ± 0.1) × 10$^{-3}$</td>
<td>(1.3 ± 0.6) × 10$^4$</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>D-mannose</td>
<td>62 ± 0.7</td>
<td>(2.3 ± 0.2) × 10$^{-2}$</td>
<td>(2.8 ± 0.2) × 10$^3$</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>1,5-anhydro-glucitol</td>
<td>n.d.</td>
<td>n.d.</td>
<td>(1.0 ± 0.1) × 10$^1$</td>
<td>5.9</td>
</tr>
</tbody>
</table>

$^a$Kinetic parameters were determined at 25 °C, in 0.2 M Tris, pH 7.6, by a glucose-6-phosphate dehydrogenase coupled assay for glucose or a pyruvate kinase/lactate dehydrogenase coupled assay for all other sugars (see Materials and Methods).
Figure 2.2 Effects of the A73G AlsK loop substitution upon the kinetic parameters for a variety of carbohydrate substrates. Fold improvements were calculated as follows: \( \frac{k_{\text{cat, A73G AlsK}}}{k_{\text{cat, wt-AlsK}}} \); \( \frac{K_{\text{m, wt-AlsK}}}{K_{\text{m, A73G AlsK}}} \); \( \frac{k_{\text{cat}}/K_{\text{m, A73G AlsK}}}{k_{\text{cat}}/K_{\text{m, wt-AlsK}}} \).

2.4 Discussion

Previous work in our laboratory has identified a divergent superfamily of bacterial sugar kinases that possess ambiguous substrate specificities toward the alternate substrate, D-glucose \((18, 19)\). This family includes the AlsK and NanK polypeptides of *Escherichia coli* K-12, which display \( k_{\text{cat}}/K_{\text{m}} \) values for glucose phosphorylation of \( 3.4 \times 10^2 \text{ M}^{-1} \text{ s}^{-1} \) and \( 1.1 \times 10^3 \text{ M}^{-1} \text{ s}^{-1} \),...
respectively. The identification of an overlapping alternate function in these enzymes, in combination with the availability of a sensitive genetic selection system linking glucokinase activity to bacterial survival, provided a unique opportunity to trace the evolutionary pathway(s) leading to the specialization of individual enzyme function within the ROK superfamily. In particular, we wished to investigate whether common structural and/or kinetic features of the sugar kinase scaffold could be targeted for alteration during the evolutionary refinement of glucokinase activity in this superfamily. Here, we show that substitutions at two similar sites within the AlsK and NanK scaffold can improve catalytic activity toward an alternate substrate. This result is similar to the findings of a previous investigation in the enolase superfamily, which demonstrated that single amino acid replacements at overlapping active site locations in the L-Ala-D/L-Glu epimerase from *Escherichia coli* and the muconate lactonizing enzyme II from *Pseudomonas* sp. P51 enhance the alternate *o*-succinylbenzoate synthase activity of these enzymes (80). Similarly, substitutions at structurally superimposable active site residues in two divergent aminohydrolase family members were found to increase the promiscuous esterase activity of each enzyme (81). Taken together, these examples strongly suggest that mutational events that target a common structural region of a polypeptide scaffold often power the initial steps of the divergent evolution of alternate function within a superfamily of related protein catalysts.

Two distinct categories of enhancing substitutions were selected from our random libraries. The first class of variants includes F145L of AlsK and V138M of NanK, both of which produced modest increases in glucokinase catalytic efficiency. Phe-145 and Val-138 reside on a β-sheet whose amino terminus includes two highly conserved glycine residues that form close interactions with the C6 hydroxyl of glucose in the crystal structures of the inorganic polyphosphate/ATP glucomannokinase from *Arthrobacter* sp. strain KM (PgmK) and *Escherichia coli* glucokinase (89, 101). In fact, Phe-145 of AlsK is directly adjacent to one of these glycine residues. Replacement of this bulky residue with a smaller leucine side chain could potentially perturb the positioning of the sugar substrate within the AlsK active site. Although the Val-138 side chain is further removed from the NanK active site, subtle perturbations at this location could also impact positioning of carbohydrates within the active site via the collective reorientation of intervening residues.

The second category of substitutions, which produced larger improvements in glucokinase activity, occurred at positions occupied by Ala-73 of AlsK and Leu-84 of NanK. These two residues are located in a variable region of the ROK scaffold located between the fourth β-sheet and the second α-helix. To better understand the nature of the glucokinase-
enhancing substitutions, we compared the amino acid sequences of NanK and AlsK in this region with a diverse range of other ROK family members (Figure 2.1). There was no detectable conservation in other ROK proteins of the position analogous to Leu-84 in NanK. Several gaps appear in the multiple sequence alignment in the region surrounding Leu-84, consistent with significant alterations occurring within this loop during the divergent evolution of function in this superfamily. The structure of the unliganded Escherichia coli NanK is available (84), however the absence of a structure that reveals how the carbohydrate substrate binds in the active site makes inferences regarding the impact of the L84P replacement upon NanK specificity difficult. In contrast to the variability in the sequence alignment near Leu-84, Figure 2.1 reveals that Ala-73 of AlsK is a highly conserved glycine residue in numerous other superfamily members. Indeed, the conservation of glycine at this position persists across the entire phylogeny of the ROK superfamily and is not limited to sugar kinases, as many transcriptional repressors also retain this residue. Unfortunately, the structure of AlsK is unknown, however, the structure of another ROK sugar kinase, the inorganic polyphosphate/ATP glucomannokinase (PgmK) from Arthrobacter sp. strain KM, was recently determined in the presence of substrate glucose (87). The active site of this enzyme reveals that the Cα atom of the conserved glycine residue (Gly-84 in PgmK) is located within 3.3 Å of the O3 atom of bound glucose (Figure 2.3). Notably, the structure of D-allose differs from that of D-glucose at the 3-OH position (Figure 1.2). On the basis of the available sequence and structural data, we postulate that removal of the Ala-73 side chain in AlsK provides sufficient space near the 3-OH of D-glucose to allow this allose epimer to bind in a position that is conducive to efficient phosphoryl transfer.

To understand the consequences of the evolutionary process more generally, the L84P NanK and A73G AlsK enzymes were tested for activity with a variety of non-native sugar substrates. Interestingly, the A73G and L84P replacements enhanced the catalytic efficiencies for all substrates examined.
Figure 2.3. The active site of inorganic polyphosphate/ATP-glucomannokinase from *Arthrobacter* sp. strain KM in complex with D-glucose, revealing the position of the highly conserved glycine residue (Gly-84) with respect to the O3 position of the bound substrate. Images were created with the program PyMOL (Delano Scientific) and PDB entry 1WOQ (87).

This observation is consistent with the results of other experimental evolution studies that support the formation of non-specific protein catalysts during the early stages of functional divergence (99, 100). In the present study, the effects of the A73G and L84P substitutions upon the steady state catalytic parameters of individual carbohydrates depended upon the nature of the substrate. For example, the A73G substitution increased the second-order rate constants for the promiscuous phosphorylation of D-altrose, D-mannose and 2-deoxyglucose, in part, by enhancing the $k_{cat}$ value of each substrate. By contrast, the value of $k_{cat}/K_m$ for glucose was enhanced solely via an altered $K_m$ value. This observation suggests that different intrinsic steps may limit the transformation rates of distinct carbohydrate substrates by each enzyme. For one carbohydrate, 2-deoxyglucose, the A73G substitution produced a greater increase in this substrate’s $k_{cat}/K_m$ value (90-fold) than for D-glucose (60-fold). This observation was unexpected since the evolved AlsK was isolated under glucose selective growth conditions, and to our knowledge the *in vivo* selection conditions imposed in this study provided no impetus to enhance activity toward 2-deoxyglucose. The present results demonstrate how the mutational expansion of enzymatic function via the formation of non-specific intermediates can serve to amplify additional promiscuous reactions, which could prove advantageous to a host organism.
CHAPTER 3
KINETICS OF GLUCOSE BINDING TO HUMAN PANCREATIC GLUCOKINASE

ABSTRACT

transitions is independent of the Hill coefficient measured via steady-state kinetic studies. A model that includes two enzyme conformations in the absence of glucose, two distinct collision complexes and five isomerization steps best describes glucose binding traces. Although simpler models were tested, the sum of square values associated with the fit suggested inaccuracies and these models were not considered to be valid. We propose a kinetic model to describe our time dependent glucose traces. Future investigations are described to identify the existence of multiple conformers and to measure their rates of interconversion via HSQC-NMR.

### 3.1 Introduction

Human pancreatic glucokinase (GK, Hexokinase IV) catalyzes the phosphorylation of glucose in the pancreatic β-cells as the first step of glycolysis (35, 43, 44). Physiologically, glucokinase was found to have a critical role in maintaining glucose homeostasis, thus it is often referred to as the body’s glucose sensor (42). Regulation of glucose concentrations throughout the human body is achieved via kinetic cooperativity, a unique property for a monomeric enzyme such as glucokinase. Two theoretical kinetic models have been proposed to describe the non-Michaelis-Menten kinetics of this monomeric enzyme. Each involves the existence of multiple enzyme conformational states that slowly interconvert, as schematically shown in Figure 3.1.

![Figure 3.1](image)

Figure 3.1 (A) Schematic representation of the minimal mnemonic model and ligand induced slow transition model (B) proposed to explain the kinetic cooperativity of the human glucokinase. Figure adapted from Neet et al. (71).

The mnemonic model was developed by Ricard, Meunier and Buc (57) based upon an early concept introduced by Rabin (58) and with the name borrowed from Whitehead (111). The mnemonic model was experimentally tested using wheat germ hexokinase type LI as a case
The emphasis of the mnemonical concept is that the enzyme relaxes relatively slowly from the state after catalysis to another state, remembering the ligand-bound conformation for a certain period after the ligand has been released. After the first turnover, the enzyme conformation (E) is able to bind glucose faster and form EG so that another catalytic cycle can occur, rather than relaxing to the thermodynamically favored form (E') Figure 3.1 (A). Cooperativity in the steady-state velocity is a result of a fast catalytic cycle involving the E, EG, E* G*ATP, E*G6P*ADP, and E*ADP and a slower catalytic cycle formed by the contribution of E, E', EG, E*G*ATP, E*G6P*ADP, E* ADP. Interconversion between E and E' is slower than catalysis, producing non-equilibration and thus kinetic cooperativity. Both enzyme species E and E' form the same intermediate binary complex and thus only that conformation is able to undergo catalysis. Simulation studies done by Richard suggest that positive cooperativity may arise when the rate of forward glucose binding to E is greater than the rate of glucose binding to E' (57).

Historically, studies on rat liver enzyme provided the first experimental evidence for glucokinase kinetic cooperativity. Most of these studies were done under steady-state conditions in presence of different types of sugar inhibitors and constituted the basis for the mnemonical mechanism (51, 62-64, 111, 112).

More recently, the elucidation of the crystal structure of human hepatic glucokinase in the absence of any ligand (termed the super-open state) as well as in the closed state, with glucose and a synthetic allosteric activator (N-Thiazol-2-yl-2-amino-4-fluoro-5-(1 methylimidazol-2-yl) thiobenzamide) present, allowed mapping of the different GK conformations in the two catalytic cycles and extrapolation of a kinetic mechanism (46). At low glucose concentrations, most of the enzyme will be trapped in the ‘slow’ cycle, which involves the inactive super-open conformation (E’ from Figure 3.1 A). As glucose concentrations increase, the enzyme exists preferentially as the high affinity state (open state, E) and thus enters the ‘fast’ catalytic cycle that involves the two active conformations of GK; the open (E) and closed states (EG). These structural studies revealed the existence of an inactive conformation of GK in the absence of glucose; the high degree of conformational changes associated with the formation of the closed state added support for the mnemonical model. Heredia presented additional experimental evidence of a mnemonical mechanism by modeling the transient kinetics of glucose binding to a kinetic model described by a two-step reversible binding (66). They were able to monitor a fast phase with a linear dependence on glucose concentration and a slow phase, which showed a hyperbolic dependence on glucose concentration. The non-cooperative substrate 2-deoxyglucose displayed the same biphasic
behavior as the cooperative substrate glucose. In their simulated model, the relaxation step that led to the thermodynamically favored conformation occurs outside of the catalytic cycle at 1 s⁻¹.

The second, more general theoretical model of monomeric glucokinase cooperativity is termed the ligand induced slow transition model and was proposed by Neet et al. (67) to explain the negative cooperativity present at low MgATP concentrations, as well as the inhibition by palmitoyl-CoA (67-71). The LIST mechanism differs from the mnemonic model by the assumption that at least two enzyme forms bind glucose with different affinities, leading to two binary complexes that can undergo two catalytic cycles. Both catalytic cycles contribute to the steady-state velocity and their flux maintains the non-equilibrium between species, thus producing cooperativity. The slow conformational changes (E ⇔ E’, EG ⇔ E’G) represent the molecular basis for the kinetic cooperativity. Binding of glucose is fast, induces a conformational change, and shifts the equilibrium between E and E’ as well as EG and E’G. The steady-state velocity comes from the contribution of two catalytic cycles formed by two distinct states of the enzyme, namely EG and E’G. The observation of a burst during the pre-incubation of rat liver glucokinase with glucose in high glycerol and its reversal to produce a lag upon glucose removal constituted the basis for glucose mediated conformational transitions that produce kinetic cooperativity (70). A subsequent study using fluorescence spectroscopy demonstrated the existence of a slow transition with a half-life of about 30 seconds (71).

Inhibition studies done by Cardenas later suggested that suppression of cooperativity is achieved via the ability of inhibitors to trap the more active conformation (113). Kim et al., who conducted their own transient state kinetics of different ligand binding to glucokinase, provided additional support for the LIST mechanism. This group modeled their results to a kinetic mechanism describing the existence of two catalytic competent species in a pre-existent equilibrium. Antoine et al. (73) also postulated that glucokinase exists in multiple conformations in the absence of any ligand.

Random addition of substrates is another possible mechanism capable of generating kinetic cooperativity apart from the mnemonic and LIST models. This mechanism was introduced to explain isotope studies of the glucokinase reaction measured at equilibrium (114). Pettersson suggested that glucose binding to differentially liganded enzyme complexes could account for the kinetic cooperativity (115). These hypotheses were ruled out by product inhibition studies done in presence of the non-cooperative substrate 2-deoxy-D-glucose, and by the observation of protection against inactivation by 5,5-dithiobis-(2-nitrobenzoic acid), which proved the binding of 2-deoxy-D-glucose as the first substrate and the release of MgADP as the last product (116).
Since the proposal of these theoretical models to explain the unique sigmoidal behavior of glucokinase, a wealth of experimental data has become available in support of one or the other. Steady-state kinetics allowed the identification of a transient and ruled out other possible explanations for cooperativity such random addition of substrates or artifacts. Currently, structural data strongly support the mnemonic model to describe the sigmoidal kinetics of human glucokinase (46), while the transient state kinetic studies remain contradictory (72, 73).

Figure 3.2 Crystal structures of human pancreatic glucokinase depicting the location of the tryptophan residues responsible for generating the intrinsic fluorescence. Figure was created with Pymol (Delano Scientific and pdb entries 1V4T (blue) and 1V4S (grey)) (148). Trp-257, Trp-167, and Trp-99 are shown as yellow spheres; glucose is represented in black and the allosteric activator molecule in orange. Note that Trp-167 is not resolved in the 1V4T structure.

In an attempt to determine which of the theoretical models best describes the observed kinetic cooperativity, we decided to determine the kinetics of glucose binding to human pancreatic glucokinase using pre-steady-state methods. Three tryptophan residues, Trp-99, Trp-167, Trp-257 found in the amino-acid sequence of GK are responsible for generating the average intrinsic fluorescence. Figure 3.2 depicts the positions of tryptophans in the super-open state and their solvent accessibility upon glucose and allosteric activator binding. Trp-167 is located on a flexible loop in the super-open conformation, and its electron density is not well defined in the structure solved by Kamata et al. (46).

In this chapter we took advantage of the intrinsic fluorescence of GK, produced by the exposure of the three tryptophans to different environments, to describe the transient state kinetics of glucose binding. A series of control experiments accompanied our kinetics to
establish the authenticity of the glucose binding curves. A model to describe the glucose binding curves in the range of 3-96 mM was generated by fitting the experimental data to different kinetic mechanisms.

3.2 Materials and Methods

3.2.1. Protein Expression and Purification

Recombinant human pancreatic glucokinase was produced as an N-terminal hexahistidine tagged polypeptide in the glucokinase-deficient Escherichia coli strain BM5340(DE3). Bacterial cultures were inoculated to an initial OD
_600 nm of 0.01 and were grown at 37°C in Luria-Bertani broth supplemented with ampicillin (150 μg/mL), kanamycin (40 μg/mL) and chloramphenicol (25 μg/mL). When the OD
_600 reached 0.85, IPTG (1 mM) was added to induce gene expression and the temperature was reduced to 20°C, where growth was continued for 20 hours. Cells were harvested by centrifugation at 8,000g and 10 g of wet cell pellet was resuspended in 30 ml of buffer A containing HEPES (50 mM, pH 7.6), KCl (50 mM), imidazole (40 mM), dithiothreitol (5 mM) and glycerol (30% w/v). Cells were lysed using a French Press and subjected to centrifugation at 25,000g at 4°C for 1 hour. The supernatant was immediately loaded onto two 5 mL HisTrap Fast Flow Affinity Column (GE Healthcare) previously equilibrated in buffer A. Following loading, the column was washed with 10 column volumes of buffer A followed by 5 columns of buffer A containing 55 mM imidazole. Glucokinase was eluted with buffer A containing 250 mM imidazole and the enzyme was dialyzed overnight at 4°C against 1L of buffer containing HEPES (50 mM, pH 7.6), KCl (50 mM) and dithiothreitol (10 mM). To eliminate aggregates from of our enzyme preparation prior to transient state kinetic experiments, we purified a sample of recombinant glucokinase using an additional size exclusion chromatography step following affinity column purification. For these experiments, dialyzed glucokinase was injected onto a Superose 6 10/300 gel filtration column (Amersham-Pharmacia) pre-equilibrated in a buffer containing HEPES (50 mM, pH 7.6), KCl (50 mM) and TCEP (4 mM). The gel filtration column was run at a flow rate of 0.02 mL/min and the fractions that contained the highest A
_280 readings were pooled and retained for further analysis. Protein obtained following size-exclusion chromatography was devoid of high molecular weight aggregates detected by analytical ultracentrifugation in samples of affinity column purified enzyme. However, we detected no difference in the steady-state or transient state kinetic data.
when the size-exclusion chromatography step was included. Thus, to maximize the enzyme concentration and the signal-to-noise ratio of transient fluorescence data, affinity column purified recombinant glucokinase that was judged to be > 95% pure based on SDS-PAGE analysis was used in transient binding experiments.

### 3.2.2 Steady-State Assays
Glucokinase activity was measured spectrophotometrically at 340 nm by coupling the production of ADP to the oxidation of NADH via the combined action of pyruvate kinase and lactate dehydrogenase. Assays were conducted at 25°C in reaction mixtures containing HEPES (250 mM, pH 7.6), KCl (50 mM), NADH (0.25 mM), dithiothreitol (5 mM), pyruvate kinase (15 units), lactate dehydrogenase (15 units), ATP (0.1-50 mM), MgCl$_2$ (1.1-51 mM), and glucose (0.05-100 mM). Data were fitted to the Hill equation or the Michaelis-Menten equation depending upon the substrate under investigation. Assays were initiated by the addition of ATP and were conducted in duplicate for each substrate concentration. The kinetic constants reported herein are the average of data obtained from at least two independent preparations of enzyme.

### 3.3.3 Equilibrium Binding Experiments
For glucose binding assays, proteins were purified as described above and were dialyzed against HEPES (50 mM, pH 7.6), NaCl (50 mM), DTT (10 mM) and glycerol (5 %) prior to data collection. Binding affinities were determined by monitoring the change in fluorescence at 335 nm that occurred in the presence of varying glucose concentrations (0.010-100 mM) following excitation of glucokinase (10μM) at 280 nm using a 5 nm slit width. Enzyme and glucose were mixed in 0.5 mL, 1.0 cm pathlength cuvettes in a buffer containing sodium phosphate (5 mM, pH 7.6), KCl (25 mM), and DTT (10 mM). Reaction mixtures were allowed to equilibrate for several minutes at 25°C prior to measuring fluorescence emission intensity. Equilibrium binding experiments were performed on a Cary Eclipse Fluorescence Spectrometer housed in the Physical Biochemistry Laboratory in the Institute for Molecular Biophysics. Data were collected in duplicate, averaged, and fitted to the following equation:

$$
\Delta F = \frac{\Delta F_{\text{max}*[glucose]}}{K_d + [glucose]}
$$

### 3.3.4 Transient State Kinetic Experiments
Transient state binding data were collected on an Applied Photophysics SX20 stopped-flow spectrometer housed in the Physical Biophysics Laboratory in the Institute for Molecular Biophysics at Florida State University. Sample syringes were maintained at 25°C by a circulating water bath. Human pancreatic glucokinase from ~20 g of wet pellet was purified on four Ni²⁺ columns and this protein was dialyzed for 12 hours against 50 mM HEPES, 50 mM KCl, 10 mM DTT, pH 7.6. This procedure gave an average of 39-52 μM GK with a $k_{cat}$ of 30 s⁻¹. Variable glucose concentration stocks (2X) were made by diluting 2M glucose into the final dialysis buffer. All of the solutions were 0.2 μm filtered prior to the experiment. The intrinsic fluorescence of human glucokinase was excited at 280 nm and the emission spectrum was monitored using a 320 nm cut-off filter attached to the spectrometer. A minimum of 5 transient fluorescence traces were obtained at each glucose concentration and the multiple traces were averaged prior to analysis using the curve-fitting program of the instrument. A variety of mixing pressures were investigated to maximize reproducibility and to minimize the loss in enzyme activity that was observed at high pressure. Based on these control experiments an optimal mixing pressure of 30 psi was chosen. Each fluorescence trace was collected for a total of 250 seconds using a log scale to obtain the fast phase kinetics. At time frames above 250 seconds a drop in protein fluorescence signal was observed, presumably due to photobleaching of the sample. A variety of enzyme concentrations (1-52 μM) were investigated to maximize the fluorescence signal. To verify the existence of multiple exponentials the following experiments were conducted:

A) Enzyme was pre-incubated with glucose and mixed with buffer in the stopped-flow instrument to measure the dissociation rates. These experiments showed the same number of exponentials as the binding experiments, indicating that the number of transitions associated with glucose binding is authentic. The hexa-histidine tag was removed via insertion of a specific seven amino acids sequence, Glu-Asn-Leu-Phe-Tyr-Gln-Ser, followed by the cleavage between Gln and Ser with TEV protease. The transient state glucose binding of the cleaved enzyme showed the same number of exponents compared with the His-tagged version of the protein.

B) To assess the degree to which glycerol affects the number of observed transients, as well as their rates of interconversion, different glycerol concentration were added to the stopped-flow buffers, up to 35% in both syringes. Addition of glycerol caused a slight decrease in the rates of glucose binding without affecting the number of exponents.

C) Technical optimizations were needed to eliminate a random artifact signal observed when enzyme was initially mixed with buffer, the amplitude of which was significant enough to
interfere with the amplitude of the fastest transient observed. These optimizations included: increased number of drives in between acquisitions from two to four, decreased slit width (from 2 nm down to 0.165 nm) and longer incubation time at 25 °C prior to acquisition (from 5 min up to 40 min). Since the artifact observed was associated with the mixing pressure, we measured the activity of the enzyme after it was mixed with buffer via an ADP-coupled assay. Exposing the enzyme to high pressure (60 psi) that was routinely used to mix reagents in the stopped-flow loop caused an 80% loss in the enzymatic activity of human glucokinase. A maximum mixing pressure of 30 psi allowed the preservation of the full enzymatic activity. Since the traces were recorded on the order of minutes, photobleaching was initially observed as a decrease in the fluorescence signal. To eliminate this effect, we decreased the slit width (0.165 nm) until a straight baseline was observed up to 250 s after mixing enzyme and buffer. The kinetic rate constants and amplitudes as a function of ligand concentration were determined from the time dependent change in fluorescence intensity implementing a single, double, triple and quadruple exponential equation of the form: 

\[ I(t) = \sum_{i=1}^{n} (A_i * e^{-k_{obsi} t}) + C \]

where \( I(t) \) is the intensity of fluorescent signal at time \( t \), \( n \) is the number of exponentials \( A_i \) is the amplitude of the \( i^{th} \) exponential, \( k_{obsi} \) is the observed rate constant for \( i^{th} \) exponential, and \( C \) is a constant corresponding to the asymptotic signal limit.

### 3.3 Results

#### 3.3.1 Glucose Binding Affinity is Very Close to the Kinetic Term, \( K_{0.5 \text{ glucose}} \)

Equilibrium binding of glucose to human pancreatic glucokinase was determined by measuring the change in intrinsic fluorescence upon glucose binding.
Figure 3.3 Hyperbolic dependence of GK intrinsic fluorescence upon glucose addition. The change in fluorescence was normalized by dividing the absolute fluorescence value at each glucose concentration by the fluorescence of GK in the absence of any glucose.

GK intrinsic fluorescence follows a hyperbolic dependence as a function of glucose concentration with no signs of cooperativity in glucose binding. The equilibrium dissociation constant for glucose was found to be 5 mM, a value characteristic of a weak binding ligand. Interestingly, this thermodynamic constant was not significantly different from the kinetic term $K_{0.5\ glucose}$ (8.1 mM). Since one is a measure of binding affinity (thermodynamic term) and the other is a kinetic term, a different set of microscopic rate constants contribute to these terms, making it difficult to establish a link between these values. Based on the $K_D$ value, we chose glucose concentrations for transient state binding kinetic experiments that span a range of 0.1 to 10$K_D$.

3.3.2 Glucose Binding to Human Pancreatic Glucokinase is Best Described by Four Exponentials for Glucose Concentrations Higher than 3 mM.

To explore the mechanism of kinetic cooperativity in human pancreatic glucokinase we have conducted transient state kinetic glucose binding experiments. Different glucose concentrations ranging from 750 $\mu$M to 96 mM were mixed with 39 $\mu$M glucokinase using the stopped flow SX-20 from Applied Photophysics. The glucose binding curves were monitored for at least 200 seconds for glucose concentrations up to 96 mM. In our attempt to fit the glucose binding data to an analytical solution, we have tested different equations. Using the curve-fitting program supplied with the SX-20 stopped-flow instrument, we have tested different sums of exponentials and their justification is presented in the paragraphs below.
Figure 3.4 Attempt to fit the 96 mM glucose binding to human glucokinase curve to a double exponential fit. (A) Superposition of the experimental data (black trace) with a double exponential fit $Y = A_1 e^{-k_{1obs}t} + A_2 e^{-k_{2obs}t} + C$ (red trace). The inset shows the poor fit in the fast regime. (B) Residuals associated with the poor fit.

Double exponential fits are shown in Figure 3.4 to emphasize the poor fit as seen in the fast time scale (inset), as well as in the residual scale. The non-symmetric distribution of the residuals, as well as the large y scale (-0.06 to 0.02), suggests that a double exponential fit is far from ideal. A comparison of the experimental data (black trace in Figure 3.4 A) with the theoretical double exponential fit (the red trace), clearly reflects the poorness of the fit. Thus higher order exponentials were tested.

The addition of an extra exponential term improved the glucose binding fits, but failed to give the best solution, as seen in Figure 3.5. Superposition of the experimental data and the fit to the analytical solution show a poor correlation in the fast and intermediate time regimes.
Figure 3.5 Attempt to fit the 96 mM glucose binding to human glucokinase curve to a triple exponential fit. (A) Superposition of the experimental data with a triple exponential fit described by the equation $Y = A_1 e^{-k_{1obs}t} + A_2 e^{-k_{2obs}t} + A_3 e^{-k_{3obs}t} + C$. (B) Residuals associated with the poor fit. (C) Zoomed areas where the experimental data are not superimposable with the fit.

Finally, quadruple exponentials gave the lowest residuals and the best superposition with the experimental data (Figure 3.6), as justified by the visual inspection of the fit (red trace) and the experimental data (black trace), as well as the symmetric distribution of the residuals and their low y scale 0.006, -0.006.
Figure 3.6 Superposition of the 96 mM glucose binding curve to a quadruple exponential fit. The fit is described by the equation $Y = A_1 e^{-k_{obs1} t} + A_2 e^{-k_{obs2} t} + A_3 e^{-k_{obs3} t} + A_4 e^{-k_{obs4} t} + C$. (A) The inset shows the fast regime region. (B) Residuals associated with the best fit.

The observed microscopic rate constants show complex behavior as a function of increased glucose concentration. The fast phase, with a maximum half-life of 14 milliseconds ($k_{obs1}$), displays a hyperbolic dependence on glucose concentration. The second phase ($k_{obs2}$) increases hyperbolically with glucose concentrations up to 12 mM, where saturation is achieved. The third phase ($k_{obs3}$) shows an initial decrease up to 3 mM glucose followed by a hyperbolic dependence on glucose concentration, reaching saturation around 80 mM. The value of $k_{obs4}$, with a half-life of 26 seconds, decreases with increasing glucose concentrations (Figure 3.7).
Figure 3.7. The dependence of the observed microscopic rate constants on glucose concentration. The values for $k_{obs}$ were obtained by fitting the data to an analytical solution comprised of the sum of four exponentials.
Figure 3.8. Representative time traces for the 0.35 mM-96mM range of glucose concentrations. (A) Normalized GK (39 μM) fluorescence changes upon glucose binding, monitored up to 250 seconds. (B) Zoomed traces to show the fast regime, where a burst is monitored with a rate and amplitude, which increases with increased glucose concentrations. (C) Normalized fluorescence of GK (39 μM) after dilution (1:1) in 50 mM HEPES, pH 7.6, 50 mM KCl, 10 mM DTT.
3.3.3 Link Between the Kinetic Cooperativity and the Number of Transients in the Binding Traces

To test if the number of kinetic transitions observed upon binding of the cooperative sugar is linked with the steady-state kinetic behavior, we investigated the binding of 2-deoxy D-glucose, a non-cooperative substrate. When 250 mM 2-deoxy-glucose was mixed with 40 μM glucokinase, an increase in GK fluorescence was observed similar to the glucose binding curves (Figure 3.8).

![Stopped-flow binding trace of 40 μM human glucokinase with 250 mM 2-deoxy-D-glucose overlaid with a quadruple exponential fit. Y= A_1*e^{-k_{obs}^{-1}} + A_2*e^{-k_{obs}^{-2}} + A_3*e^{-k_{obs}^{-3}} + A_4*e^{-k_{obs}^{-4}} + C (upper figure). The inset shows the fast regime region. Residuals associated with the best fit are shown in the lower figure.](image)

Figure 3.9 Stopped-flow binding trace of 40 μM human glucokinase with 250 mM 2-deoxy-D-glucose overlaid with a quadruple exponential fit. $Y= A_1*e^{-k_{obs}^{-1}} + A_2*e^{-k_{obs}^{-2}} + A_3*e^{-k_{obs}^{-3}} + A_4*e^{-k_{obs}^{-4}} + C$ (upper figure). The inset shows the fast regime region. Residuals associated with the best fit are shown in the lower figure.

3.3.4 Simulation of Different Binding Mechanisms

The pro-Kineticist II program from Applied Photophysics was used to fit the binding data over a range of glucose concentrations from 3 mM to 96 mM. This program allows the global fitting of all glucose concentrations to a mechanism, as well as the extraction of individual microscopic rate constants. Different binding mechanisms were tested:
Mechanism II, III, and IV failed to give a good fit (sum of squares was 25.5 after 200 iterations, 3.16 after 200 iterations and 48 after 200 iterations, respectively). Mechanism I (Figure 3.11) provided the closest fit to our kinetic data. In this mechanism, we have assumed the existence of two species in the absence of any ligand, E and E'. [E] was initially considered to represent 1% of the total enzyme concentration (39 μM) and subsequently changed to 5%, 10%, 20% and 50% of the total enzyme concentration. The assumption of 20% of E gave the closest agreement between the calculated equilibrium constant, 6.4, and the assumed one, 4. A variation of this mechanism, where the steps $E_1G \leftrightarrow E'_1G$ and $E'_2G \leftrightarrow E_2G$ are missing was also considered. The sum of squares for the mechanism I decreased from 1.4 to 0.9797 when these steps were included, suggesting their importance. We did not include binding data below 3 mM glucose in our fit, because this was the minimum glucose concentration at which we could successfully fit the data to four exponentials.
From these microscopic rate constants we can calculate $K_D$ values for the high affinity state $E$ to be 613 $\mu$M and 113 mM for the low affinity state, $E'$ respectively. The second order rate constants for the initial encounter of enzyme and glucose are lower than expected for a collision encounter event ($10^{-4}$-$10^{-8}$ M$^{-1}$ s$^{-1}$). However, the isomerization steps are on the order of $10^{-2}$-$10^{-3}$ s$^{-1}$ as predicted by the LIST model.

### 3.4 Discussion of the Glucose Binding Model Proposed

Glucokinase is a monomer under physiological conditions (52), yet displays a sigmoidal kinetic response to increases in cellular glucose concentrations. This cooperative feature is essential to this enzyme’s role as the principal sensor of blood glucose concentrations in the human body (42). To explain the kinetic cooperativity responsible for the ability of human glucokinase to regulate glucose concentration in the blood stream, two theoretical models have been postulated. The mnemonical model, schematically represented in Figure 3.1A, involves only one enzyme species in the catalytic step (EG), whereas the LIST mechanism (Figure 3.1B) predicts the contribution of two conformations (EG and E’G) to catalysis, which give rise to cooperativity. Schematic representations of the two models do not include the possible formation of extra binary complexes after the formation of the collision complex. Most of the evidence available prior to 2007 suggested that human glucokinase adheres to the mnemonical model. However, two recent reports have suggested the existence of two conformers of human...
glucokinase in the absence of glucose, which interconvert slowly, as postulated by the LIST mechanism (72, 73).

To investigate which of these models best describes the kinetic cooperativity in human glucokinase, we conducted our own transient state kinetic studies. Different glucose concentrations were added to 39 μM human glucokinase in the stopped-flow apparatus at a constant temperature of 25 °C. At high glucose concentrations (above 3 mM) stopped-flow traces are best described by an analytical solution comprised by four exponents. Glucose concentrations below 3 mM could be fitted to lower order exponentials. A fast transient is observed with a maximum half-life of 14 milliseconds, whose amplitude and rate are increased upon increasing glucose concentration (Figure 3.7 and Figure 3.8). The slowest transition observed in our studies has a half-life of about 26 seconds, very similar to the slow transition (30 seconds) found by Neet et al., (71). The magnitude of the fluorescence change in the transient state binding traces corresponds to the equilibrium binding experiment (1.55 versus 1.7), thus verifying that during the transient state measurements we achieved maximum saturation (Figure 3.3 A compared with Figure 3.8).
To validate the existence of four exponents, we preincubated 100 mM glucose with 20 μM glucokinase and the dissociation rates were measured after a two-fold dilution. Glucose dissociation curves were best described by an analytical solution comprised of four exponents, similar to the glucose binding curves (Figure 3.11). The existence of a minimum of four distinguishable kinetic steps described by four independent exponentials in either forward or the reverse direction of glucose binding to glucokinase proves the authenticity of these steps. Usually, plots of the observed rate constants as a function of ligand concentration give valuable information about the number of steps in a mechanism, the reversibility of the steps and allow extraction of the microscopic rate constants. The complexity of the data did not allow us to extract any meaningful information via fitting to an analytical solution. Although our attempts are shown in Figure 3.7 we have chosen to fit the glucose binding curves to different mechanisms. This approach allowed us to test different mechanistic variations including the mnemonic and LIST models. The existence of four exponentials required that at least four kinetically distinguishable steps are needed to fit our experimental glucose binding curves as seen from the sum of squares values.
Addition of extra binary complexes after the initial collision complex decreased the sum of squares, indicating their necessity in the mechanism. Qualitatively we have found that we need at least two conformations of enzyme in the unliganded state to characterize the glucose binding curves. Thus a minimum model has to include at least two conformations of the enzyme and an isomerization step after the initial binding step. Next we sought to make a quantitative estimation of mechanism I by fitting the experimental binding curves and extracting microscopic rate constants.

To validate our fit we have compared the equilibrium constant $K_{eq}$ between E and E' assumed at the beginning of the fit with the calculated $K_{eq, \text{sim}}$ from the microscopic rate constants ($k_3$ and $k_{-3}$). Twenty percent [E] gives an equilibrium constant of 4. This equilibrium constant was verified by calculating $K_{eq, \text{sim}}$ to be 6.4.

Our transient state kinetic results are consistent with the existence of multiple transitions detectable by intrinsic fluorescence upon glucose binding to human pancreatic glucokinase. At least 200 s are needed to achieve completion of glucose binding, similar to the observations made by Neet et al. 20 years ago (71). The high order of exponentials needed to fit the data to an analytical solution gives observed microscopic rate constants which have complex behavior with increased glucose concentrations. Thus, Pro-K II was used to fit the stopped flow data to various mechanisms. Different mechanistic variations were tested including the mnemonic and LIST models. Our simulation results are consistent with the existence of at least two enzyme species capable of binding glucose and isomerizing to other binary states. They also suggest that multiple conformations of human glucokinase might exist in the absence of glucose and that the binding of glucose to two possible conformers is followed by at least one slow isomerization step.

These observations are similar to ones presented in a very recent report by Antoine et al. (73) who suggested that glucokinase samples multiple conformations between the super-open and closed states. The existence of multiple GK conformers in solution in the absence of any ligand was inferred to be necessary for the regulatory mechanism. Different binding partners would regulate its activity in vivo by preferentially binding to one of the states and thus perturbing the equilibrium between conformers. For example, they suggested that binding of the open glucokinase conformation to its inhibitor (glucokinase regulatory protein, GKRP) at low glucose concentrations would shift the pre-existing equilibrium between multiple glucokinase conformations, thus depopulating the closed, active conformers. However, in contrast to their conclusion we suggest that glucose binding data only required the existence of a mechanism comprised of two glucokinase conformers in absence of glucose, and suggested the existence
of four other interconverting binary complexes. The slowest rates are the ones corresponding to interconversion between different conformations $E \Leftrightarrow E'$, $E_1G \Leftrightarrow E'_1G$ and $E_2G \Leftrightarrow E'_2G$.

The major weakness of our model is that it relies on the assumption that two conformation of GK exist in the absence of the ligand glucose. Although this is not an unreasonable assumption, since it was made based upon the fact that none of the binding mechanism involving only one GK conformation at 0 glucose concentration gave a reasonable fit, we would like to experimentally observe these GK conformations. The initial collision complexes are formed at $1541 \, \text{M}^{-1} \, \text{s}^{-1}$ and $200 \, \text{M}^{-1} \, \text{s}^{-1}$ respectively, at least 10-fold lower than expected from a substrate-enzyme encounter event. Refinement of this initial model will be pursued via additions of more experimental information. For instance, the dissociation kinetics will be measured for the same enzyme and glucose concentrations and the experiment will be added along with the binding curves to refine the microscopic rate constants.

Our future aims are to experimentally detect the number of conformations in the absence of glucose via other methods and measure their rates of interconversion. It will be valuable to be able to trap one of the conformers, so that we could quantitatively estimate the equilibrium constant.
CHAPTER 4

INVESTIGATION OF CONFORMATIONAL HETEROGENEITY OF HUMAN PANKREATIC GLUCOKINASE VIA HSQC-NMR

ABSTRACT

Human pancreatic glucokinase exists as a monomer in solution in the presence of physiological substrates and inhibitors. Two different mechanisms (mnemonic and LIST) have been postulated to explain glucokinase kinetic cooperativity, each of which involves the existence of multiple enzyme conformations that interconvert slower than catalysis. Transient state kinetic studies, described by our group in the previous chapter, reveal multiple transitions associated with glucose binding and were modeled via the LIST model. However the complexity of the transient state kinetic data suggested that additional experimental evidence is necessary to support our model. Preliminary HSQC experiments of the unliganded state show five regions of broad peaks, which could be characteristic of an aggregated or of a partially unfolded protein. Broadening in the HSQC can also arise from multiple conformations interconverting slowly. Introducing an extra purification step eliminated the aggregation, as confirmed by analytical ultracentrifugation, but did not resolved the peaks seen in HSQC even at concentrations as low as 120 μM. These preliminary results suggest the possibility that slowly interconverting conformers could cause peak broadening in HSQC. Calculated hydrodynamic parameters of the super-open and closed structures predict a ten percent increase in sedimentation coefficient upon glucose and synthetic allosteric activator binding. However, the experimentally determined values for sedimentation coefficients determined in the presence and absence of glucose do not differ significantly. Either analytical ultracentrifugation does not provide enough resolution to distinguish different human glucokinase conformers, or formation of the more compact form of GK requires the presence of a bound activator.
4.1 Introduction

Human pancreatic glucokinase exists as a monomer in solution in the presence of physiological substrates and inhibitors (52). The steady-state velocity of the glucokinase-catalyzed reaction is not hyperbolic, but is rather described by a Hill equation with an exponent of 1.7-1.8. Other cooperative enzymes are known to have a number of subunits that is at least as large as the Hill coefficient. Thus, long before establishing the kinetic nature of glucokinase cooperativity, it was speculated that GK might be a dimeric enzyme. Moreover, hexokinases I-III, isozymes of glucokinase, are known to undergo dimerization at high enzyme concentrations. Because the non-Michaelis Menten behavior is a rare characteristic of a monomeric enzyme, initial studies on the rat liver enzyme were directed towards establishing the monomeric nature of glucokinase (52, 53). Gel filtration studies by Cardenas et al. determined the apparent molecular weight of rat liver glucokinase (52) under reacting conditions, and confirmed the existence of the enzyme as a monomer in the presence of substrates and products, as well as in the presence of the inhibitor N-acetylglucosamine, whose phosphorylation is described by classical hyperbolic kinetics. These studies eliminated the possibility of GK oligomerization as an explanation for cooperative kinetics. Nevertheless, Parry et al. (112) suggested the separation of a “light” and a “heavy” component of rat liver glucokinase whose kinetic characteristics are sigmoidal and Michaelis-Menten, respectively. Once the monomeric nature of the enzyme was fully established, new models had to be proposed to explain the kinetic cooperativity of this monomeric enzyme. The mnemonic model, developed by Richard, Meunier and Buc (57) and first applied on wheat germ hexokinase, assumes the existence of one thermodynamically favored conformation in absence of the ligand, and one catalytically competent conformation. The ligand induced slow transition model (LIST) proposed by Neet et al. (67) suggests the existence of two glucokinase conformations in the absence of glucose, which are slowly interconverting and both able to undergo catalysis. Evidence to support the mnemonic model comes from the structural studies (46), as well as transient kinetic studies (66). Recently, Kim et al. (72) and Antoine et al. (73) suggested the existence of multiple states of human glucokinase that slowly interconvert in the absence of any ligand, adding support to the theoretical model proposed by Neet et al. (67). Transient state kinetic studies described by our group in the previous chapter also reveal multiple transitions associated with glucose binding, and were modeled based on the LIST model. However the complexity of the transient
state binding kinetic data suggested that additional experimental evidence is necessary to support our model.

In this chapter, we pursue experimental validation of our model using analytical ultracentrifugation (AUC) and HSQC-NMR. Crystal structures indicate that the super-open state of glucokinase adopts a more elongated conformation, while in the closed state (upon ligand binding), the glucokinase conformation is more globular. This observation suggested that addition of glucose to glucokinase would change the shape of the conformation and thus its hydrodynamic properties. The extent to which AUC can be used to characterize GK in solution will be tested.

Biophysical studies that have been used to characterize wild type glucokinase and its interaction with substrates, inhibitors, and regulatory proteins include differential scanning calorimetry, isothermal titration calorimetry, far UV circular dichroism, and fluorescence. No NMR studies exist up to date to describe GK dynamics. To identify different conformers of glucokinase in solution in the absence or presence of glucose we have conducted preliminary H-15N HSQC experiments.

The second aim of this chapter is to investigate the composition of the sample using analytical ultracentrifugation and gel filtration. More specifically to quantify the degree of sample aggregation we used sedimentation velocity, and to estimate the apparent molecular weight of the aggregates we used gel filtration chromatography.

### 4.2 Materials and Methods

#### 4.2.1 Protein Expression and Purification

Human glucokinase was expressed and purified on a Ni^{2+}-NTA affinity column following the method described in Chapter 3.2.1. Fractions containing the highest A_{280 nm} from the Ni^{2+} column were pooled and dialyzed against 50 mM HEPES, pH 7.6, 50 mM KCl, 5% glycerol, 10 mM DTT for 4 hours. Enzyme was further concentrated by centrifugation at 1700g and 0 °C, using Amicon Ultracentrifugal devices with a 10,000 MWCO from Millipore, until concentrations between 0.5 mM and 1 mM were reached. 100 μL of concentrated human glucokinase was injected into a Superose 6 10/300 gel filtration column (Amersham-Pharmacia) previously equilibrated with 50 mM HEPES, pH 7.2, 50 mM KCl and 4 mM TCEP. Human glucokinase was run overnight at 0.02 mL/min at 4 °C. Conalbumin (Mol. weight 75000), Ribonuclease A (Mol.
weight 13700), Aprotinin (6500) were used as standards. A calibration curve was created from these standards to extract the apparent molecular weight of glucokinase and its aggregates, by plotting the partition coefficient ($K_{av}$) versus the logarithm of molecular weight. The partition coefficient was calculated using the formula: $K_{av} = (V_e - V_v) / (V_{tot} - V_v)$ where $V_e$, $V_v$ and $V_{tot}$ are the elution, void and total volumes, respectively. $V_v$ was measured using Blue Dextran ($2 \times 10^6$) and $V_{tot}$ was measured using DNP-glycine (MW=241). Fractions with the highest $A_{280}$ were pooled and immediately analyzed via either analytical ultracentrifugation or NMR. Aliquots of the enzyme were assayed after concentration, before performing AUC/NMR and after each experiment to ensure that the enzyme retained full activity.

For NMR experiments human glucokinase ($pBGM101$-$glk$) was expressed in BL21 (DE3) cells. Bacterial cultures were inoculated to an initial $OD_{600\text{nm}}$ of 0.05 and were grown at 37°C in minimal media containing 50 mM Na$_2$HPO$_4$, 20 mM KH$_2$PO$_4$, 10 mM NaCl, 0.1 mM CaCl$_2$, 1 g/L $^{15}$NH$_4$Cl, 1 mM MgSO$_4$, Thiamine HCl (25 mg/L), 0.5 % (w/v) glycerol and ampicillin (150 μg/mL). When the $OD_{600\text{nm}}$ reached 0.85, 1 mM IPTG was added to induce gene expression and the growth was continued for 12 more hours at 37°C.

Protein was purified on a Ni$^{2+}$-NTA affinity column as described in section 4.2.1 and concentrated as described above. HSQC-NMR spectra were obtained by Dr. Fengli Zhang, a researcher in Dr. Bruschweiler’s group at the National High Magnetic Laboratory.

### 4.2.2 Analytical Ultracentrifugation

To determine the sedimentation coefficient of human glucokinase in the absence of glucose, enzyme concentrations of 6.5 μM, 16.3 μM and 31.3 μM were used. 390 μL was loaded in the sample cell and 400 μL of water was included as reference. Twelve mm double-sector Epon charcoal filled cells with quartz windows were loaded into an AN60Ti four-hole rotor. Cells were tightened at 125 psi, aligned and the temperature equilibrated at 20°C for at least 30 min before acquisition. Data were collected on a Beckman Coulter Proteome Lab XL-1 Analytical Ultracentrifuge in the Physical Biochemistry Laboratory. Velocity experiments were set at different rotor speeds, 20000 rpm, 30000 rpm, 50000 rpm and 60000 rpm, to optimize the resolution of sedimentation coefficient data. To test if AUC provides the resolution necessary to distinguish different conformations in solution, based on their frictional ratio or the sedimentation coefficient, we have run the wild-type glucokinase in presence of 50 and 100 mM glucose.

### 4.2.3 MALDI-TOF Experiments
The molecular mass of peak 1, peak 2, peak 3, and peak 4 obtained after size-exclusion chromatography of human glucokinase were verified using Matrix Assisted Laser Desorption/Ionization Time-of-Flight mass spectrometry. Sinapinic acid at 10 mg/mL dissolved in a mixture of 50% acetonitrile and 0.1% trifluoroacetic acid was used as a matrix. Peaks 1-4 collected after size-exclusion chromatography were concentrated using Nanosep 3K Omega centrifugal devices from Pall Corporation. Peaks 1-4 were either diluted at least 5-fold into 0.1% trifluoroacetic acid (TFA) or ZipTip purified before spotting onto the MALDI plate. 1 μL of matrix was mixed with 1 μL of the sample and the mixture was spotted onto a clean MALDI target plate. To calibrate the mass spectrometer, Bovine Serum Albumin (66400 Da), Aldolase (39211 Da) and Aprotinin (6500 Da) were used.

### 4.3 Results

#### 4.3.1 Preliminary Heteronuclear Single Quantum Coherence (HSQC) Experiments for Wild-Type Human Glucokinase

2D H-15N HSQC NMR provides a fingerprint of the protein backbone amides. Changes in both chemical shift and intensity can be observed depending on the environment of the backbone functional groups.

Table 4.1: Activity assays for GK before and after the HSQC experiments

<table>
<thead>
<tr>
<th>sample concentration</th>
<th>$k_{cat}$ (s$^{-1}$) before HSQC</th>
<th>$k_{cat}$ (s$^{-1}$) after HSQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6 mM</td>
<td>36</td>
<td>25</td>
</tr>
<tr>
<td>120 μM</td>
<td>47</td>
<td>42</td>
</tr>
</tbody>
</table>
HSQC experiments are intended to identify different conformations of human glucokinase in solution in the absence of glucose, based upon the number of peaks produced from the chemical shifts of the peptide bond $^{15}\text{N}-\text{H}$ groups. A typical HSQC spectrum of a single conformation of GK should have 464 peaks that are dispersed across the 100-140 ppm scale for $^{15}\text{N}$ and 6-11 ppm for H. Our results show five overlapping regions of peaks corresponding to: 6.5-7 ppm (H) and 110-114 ppm (N); 7.2-7.8 ppm (H) and 110-114 ppm (N); 8 - 8.5 ppm (H) and 106-112 ppm (N); 7-7.5 ppm (H) and 124-126 ppm (N); 7.8-8.5 ppm (H) and 116-126 ppm (N) (Figure 4.1 A).

To test the hypothesis that aggregation might cause the overlapping peaks observed in the HSQC spectrum, we ran a GK sample purified on a Ni$^{2+}$-NTA affinity column, followed by Superose 6 10/300 gel filtration column (Amersham-Pharmacia), at a concentration that is 13-fold lower than the one shown in Figure 4.1A. The HSQC spectrum of 120 μM GK (Figure 4.1 B) looks very similar to the concentrated sample, suggesting that the lack of spectral dispersion is not caused by aggregation. Control experiments using activity assays (Table 4.1) and sedimentation velocity confirm the stability and monomeric nature of the sample under the conditions tested.
Figure 4.2 Sedimentation velocity of 24 μM $^{15}$N-GK conducted from the recovered sample (120 μM) after the HSQC NMR experiment. Extrapolation plot (A) shows the existence of the sample as a single component of 3.6 S. 2DSA spectrum analyses (B) confirms the existence of a single molecular weight species and no aggregation. RMSD for the fit is 0.00579.
Sedimentation velocity experiments done on the same sample recovered after NMR confirms the existence of single specie with an \( f/f_0 \) of 1.4 and sedimentation coefficient of 3.65 S. No aggregation is observed, as indicated by the lack of heterogeneity in the extrapolated values for the sedimentation coefficient (Figure 4.2 A).

### 4.3.2 Sample Composition and Shape Analysis via Analytical Ultracentrifugation

Sedimentation velocity experiments performed in an analytical ultracentrifuge provide results that can characterize the hydrodynamic properties of biological macromolecules, such as sedimentation, diffusion- and frictional parameters, as well as the molecular weight. From the enhanced van Holde-Weischet (vHW) analysis (117, 118) method (UltraScan (119)) the sedimentation coefficient distribution, corrected for diffusion, is obtained, which allows the computation of a *particle size distribution*. The vHW method is a model-independent method and relies on the different time dependence of the diffusion and sedimentation processes. Since diffusion is a process that takes place with the square root of time while sedimentation is a function of the first order with respect to time, van Holde and Weischet in 1979 realized that these two transport processes can be separated. The procedure they developed (117) is intended to eliminate the broadening due to diffusion and to extract a sedimentation coefficient corrected for diffusion via extrapolation of the apparent \( s^* \), corresponding to each boundary fraction, to infinite time. The next step is to fit the simulated finite element solutions to the experimental data by the 2-dimensional spectrum analysis (2DSA) procedure (120, 121) implemented in UltraScan (119). This method allows determination of the sedimentation and frictional ratio distributions simultaneously, as well as the subtraction of time invariant noise known to cause artifacts in analysis. To refine the noise-subtracted results and eliminate false positives due to overestimating the parameters during the 2DSA routine, we used the Genetic Algorithm (122).

To determine the molecular mass, sedimentation coefficient and the shape of human glucokinase analytical ultracentrifugation experiments were conducted. If different conformations of human glucokinase exist in solution and display significantly different sedimentation coefficients or diffusion coefficients, we could distinguish these using analytical ultracentrifugation sedimentation velocity experiments. To verify the existence of aggregates and their successful removal via the addition of a size-exclusion step in the purification procedure, we have used sedimentation velocity experiments.
Analytical ultracentrifugation experiments performed on a sample of human glucokinase purified only via Ni\textsuperscript{2+}-NTA affinity chromatography at 40,000 rpm and 20 °C, showed the presence of a fast sedimenting component that pelleted faster than the rest of the sample. This observation was a strong indicator of aggregation.

Figure 4.3. Snapshot of the van Holde and Weischet analysis of human glucokinase after the Ni\textsuperscript{2+}-NTA column showing a fast moving component sedimenting faster than the rest of the sample. Note that in this sedimentation coefficient distribution, the time invariant noise was not substracted.
Figure 4.4. Extrapolation plot of 26 μM GK (A) described in Figure 4.3 and the corresponding distribution plot (B) of the sedimentation coefficient across the boundary fraction, after the time invariant noise was substracted.
The heterogeneity caused by an aggregated sample is seen even after the subtraction of the time invariant noise, which is known to cause artifacts in the analysis. Extrapolation of the apparent sedimentation coefficient at infinite time does not give a constant value for the sedimentation coefficient. Rather, a range of sedimentation coefficients from 2.7-6 S is observed (Figure 4.4 A). This distribution suggests that a fraction of the sample (20 percent) is aggregated and sediments with higher coefficients. The aggregation is visible also after two dimensional spectrum analyses (2DSA) as reflected by solutions with f/f₀ at 4 and a wide range of s values (2-10) (Figure 4.5, A).
Figure 4.5 Finite element solutions (pairs of $s$ and $f/f_0$) for the 26 μM GK displayed in UltraScan (Genetic Algorithm Initialization module) (A). Superposition of the model generated via 2DSA and the experimental sedimentation traces (B). RMSD for the fit was 0.0087. The darkest spots represent the space with the
The highest concentration of the solute, while blue and white spot reflect the space with the least concentration.

To estimate the apparent molecular weights of the aggregated species and to eliminate these from our sample we have used Size-Exclusion Chromatography. The calibration of the Superose 6 10/300 gel filtration column (Amersham-Pharmacia) was done using Conalbumin (75000) Ribonuclease A (13700), Lysozyme (14 400) and Aprotinin (6500) (Figure 4.6).

![Figure 4.6 Calibration of the Superose 6 10/300 gel filtration column. Calibration was done using Conalbumin (75000), Ribonuclease A (13700), Lysozyme (14 400) and Aprotinin (6500) as standards. Void volume was measured using Blue Dextran (approximate Mw = 2 x 10^6). Total volume was measured with the chromogenic amino acid DNP-glycine (Mw = 241). The curve was generated by plotting the distribution coefficient $K_{av}$ of each protein against the log of its molecular mass.](image)

The human glucokinase sample elutes from the Superose 6 10/300 gel filtration column as four peaks with distinguished elution volumes. We have assigned peak 1 and peak 2 as higher order aggregates based on their apparent molecular weight, as well as their low $k_{cat}$ values (Table 4.2 and Figure 4.7). Peak 3 had an apparent molecular weight of 64 kDa, and the highest $k_{cat}$ value (33 s⁻¹).
Figure 4.7 Elution profile of human glucokinase after an overnight run at 0.02 mL/min on a Superose 6 10/300 column. Each peak was tested for activity and identity on 16.5 % SDS-PAGE gel. The turnover rates are displayed on the chromatogram.

Based on the 16.5 % SDS-PAGE gel peak 4 is formed by leakage of peak 3 and another impurity. The kinetic parameters of peak 3 were very similar to the sample purified after the first affinity column, ($K_{0.5, \text{glucose}}$ of 5.6 mM, $k_{\text{cat}}$ of 33 s$^{-1}$ and a Hill coefficient of 1.56). Absorbance scans of peak 4 revealed a $\lambda_{\text{max}}$ at 260 nm, which is characteristic of nucleic acids. This peak was considered an impurity and was not investigated further.

<table>
<thead>
<tr>
<th>Elution Volume (ml)</th>
<th>$K_{av}$</th>
<th>log (Mol. weight)</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>peak 1</td>
<td>13.7</td>
<td>0.32</td>
<td>5.2</td>
</tr>
<tr>
<td>peak 2</td>
<td>14.4</td>
<td>0.35</td>
<td>5.0</td>
</tr>
<tr>
<td>peak 3</td>
<td>15.3</td>
<td>0.39</td>
<td>4.8</td>
</tr>
<tr>
<td>peak 4</td>
<td>18.3</td>
<td>0.52</td>
<td>4.0</td>
</tr>
</tbody>
</table>

To verify the molecular weight of the peaks eluted and eliminate the possibility that other contaminants might be present within peaks 1, 2, 3, and 4, we used MALDI-TOF mass spectrometry.
Figure 4.8. MALDI-TOF spectra of peaks 1, 2, 3 and 4 with the y-axis aligned (m/z), indicating the molecular mass of the peaks.

The experimental mass of the peaks determined via MALDI-TOF (Figure 4.8) is in close agreement with the calculated mass based on the amino acid sequence (53128 Da) of GK. Peak 4 elutes along with another impurity of apparent mass 1190 Da. A 16.5 % SDS-PAGE gel of fractions between peak 3 and 4 show leakage of peak 3 into peak 4. Thus, we believe that the mass determined via MALDI-TOF corresponds to the fraction of peak 3 that leaks, and that of peak 3 with the 1190 Da impurity.

Sedimentation velocity experiments conducted upon human glucokinase purified using just an affinity column revealed a fast moving component of about 20 % of the boundary fraction (Figures 4.3 Figure 4.4). We have assigned this behavior to correspond to aggregation. Size exclusion chromatography confirms the presence of aggregates as peak 1 and 2. AUC analysis
to test if different GK conformations exist in solution will be conducted after the addition of size exclusion chromatography as an extra purification step.

Figure 4.9 and 4.10 show the sedimentation velocity experiment of 31 μM human glucokinase in the absence of glucose. The changes in the hydrodynamic parameters, if any, upon 50 mM and 100 mM glucose addition are seen in Figures 4.11, 4.12; and Figures 4.14, 4.13, respectively.
Figure 4.9 Sedimentation velocity (50,000 rpm, 20°C) of 31 μM wild-type human glucokinase in the absence of glucose. Van Holde-Weischet extrapolation plot (upper). Distribution of sedimentation coefficient across the boundary fraction (lower).
Figure 4.10 Sedimentation velocity (50,000 rpm, 20°C) of 31 μM wild-type human glucokinase in the absence of glucose. Pseudo-3D plots of calculated f/f₀ versus sedimentation coefficient indicating the regions of higher solute concentration (upper). Superposition of the experimental data (yellow) and the simulated trends (red) after the two-dimensional spectrum analysis (2DSA) (lower). RMSD for the fit is 0.00847.
Figure 4.11 Sedimentation velocity (50,000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 50 mM glucose. Van Holde-Weischet extrapolation plot (upper). The distribution of sedimentation coefficient across the boundary fraction (lower).
Figure 4.12 Sedimentation velocity (50,000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 50 mM glucose after 2DSA. Pseudo-3D plots of calculated $f/f_0$ versus sedimentation coefficient indicating the regions of higher solute concentration (upper). Superposition of the experimental data (yellow) and the simulated trends (red) after the two-dimensional spectrum analysis (2DSA) (lower). RMSD for the fit is 0.00864.
Figure 4.13 Sedimentation velocity (50,000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 100 mM glucose. Van Holde-Weischet extrapolation plot (upper). The distribution of sedimentation coefficient across the boundary fraction (lower).
Figure 4.14 Sedimentation velocity (50,000 rpm, 20°C) of 31 μM wild-type human glucokinase in the presence of 100 mM glucose after 2DSA. Pseudo-3D plots of calculated f/f₀ versus sedimentation coefficient indicating the regions of higher solute concentration (upper). Superposition of the experimental data (yellow) and the simulated trends (red) after the two-dimensional spectrum analysis (2DSA) (lower). RMSD for the fit is 0.00865.
From the van Holde and Weischet analysis (117) we can estimate the sedimentation coefficient distribution and extract an average sedimentation coefficient (Table 4.3). The straight line corresponding to $s_{(20,w)}$ across the fraction boundary shown in the second panel of Figure 4.9 reflects the existence of a principal component in the sample. Subsequent 2DSA analysis suggests the presence (below 5% of the total sample concentration) of a very globular component of 3000 Da having a sedimentation coefficient of 1 S and an $f/f_0$ of 1 (Figure 4.10).

Table 4.3: Hydrodynamic parameter dependence upon the angular velocity

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>(van Holde-Weischet method)</th>
<th>Two-dimensional spectrum analysis (2DSA)</th>
<th>Genetic Algorithm (GA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$s_{(20,w)}$ (S)</td>
<td>Mol. weight (Da)</td>
<td>$s_{(20,w)}$ (S)</td>
</tr>
<tr>
<td>20 krpm</td>
<td>26 μM wt-GK</td>
<td>4.13</td>
<td>79585</td>
</tr>
<tr>
<td></td>
<td>26 μM wt-GK in presence of 50 mM glucose</td>
<td>4.14</td>
<td>68189</td>
</tr>
<tr>
<td>30 krpm</td>
<td>26 μM wt-GK</td>
<td>3.55</td>
<td>63856</td>
</tr>
<tr>
<td></td>
<td>26 μM wt-GK in presence of 50 mM glucose</td>
<td>3.62</td>
<td>62596</td>
</tr>
<tr>
<td>50 krpm</td>
<td>31 μM wt-GK</td>
<td>3.39</td>
<td>46003</td>
</tr>
<tr>
<td></td>
<td>31 μM wt-GK in presence of 50 mM glucose</td>
<td>3.39</td>
<td>50866</td>
</tr>
<tr>
<td>60 krpm</td>
<td>31 μM wt-hGK</td>
<td>3.51</td>
<td>54185</td>
</tr>
<tr>
<td></td>
<td>31 μM wt-hGK in presence of 50 mM glucose</td>
<td>3.52</td>
<td>65187</td>
</tr>
</tbody>
</table>
Since the concentration of this component was very low, we have assigned it as an impurity. To emphasize the quality of fits from which the data in Table 4.3 is gathered, we have shown the superposition of the experimental data with the simulated curves using the 2DSA method in the last panel. RMSD values after 2DSA acceptance were less than 0.009.

Speed dependent sedimentation velocity experiments were conducted to optimize the resolution of the sedimentation coefficient. The changes in the hydrodynamic parameters between the possible human glucokinase conformers are not expected to be large. Thus the accuracy in determining the sedimentation and diffusion coefficients is critical. Increasing angular speed is expected to provide increased resolution in the determination of the sedimentation coefficient due to the low contribution of diffusion (123). We have tested different rotor speeds from 20,000 rpm up to 60,000 rpm. Wild-type enzyme concentrations in the range of 26-31 μM were run at each speed in the absence of the ligand and in the presence of 50 mM glucose. As expected, at low speed, the variation of the sedimentation coefficient is higher (Table 4.3) and thus the determined molecular weights are far from the calculated value of 53128 Da based on amino acid composition. In absence of the ligand, the experimentally measured sedimentation and diffusion coefficients vary in the range of 3.4 – 4.2 S and 4.6-6.1 x 10^{-7} cm^2/s, respectively, depending upon variations in the angular velocity. The shape of human glucokinase is given by the f/f_0 ratio which varies from 1.45-1.8 with decreasing rotor speed. A speed of 50,000 rpm allowed accurate determination of the wild-type molecular weight 52035 Da (after Genetic Algorithm was performed). The parameters extracted from the 50 krpm experiment are: a s_{20,W} of 3.4 Svedberg, a D_{20,w} of 5.91 x 10^{-7} cm^2/s and a f/f_0 of 1.45. The changes in the hydrodynamic properties of the wild-type enzyme upon addition of 50 mM and 100 mM glucose are within the range observed in the absence of ligand (Figures 4.9, 4.10, compared to Figures 4.11, 4.12, 4.13 and 4.14). Thus, using AUC, we did not observe drastic variations in the hydrodynamic properties upon glucose binding. To compare our experimental values for the sedimentation and diffusion coefficients with the expected constants, we have used the SOMO bead modeler module included in the UltraScan software (124) and the pdb entries 1V4S and 1V4T of glucokinase.

SOLution MOdeler, developed by the Rocco and Byron labs, (125-127) is based upon replacing the atoms by beads of the same volume, which are then grouped together to create a reduced representation of the molecule. The volume of hydrating water that is theoretically bound to the atoms can then be added to each bead. If the beads are overlapping, the area of overlap is further reduced without major changes in the original surface created by the bead model. Hydrodynamic parameters are calculated using the Garcia de la Torre-Bloomfield rigid
body approach (128-131). This method was used extensively for a wide range of molecular weights (124, 132) and it provides a better treatment of the water of hydration and it maintains the initial relationship between the beads and the residues.

Table 4.4: Calculated hydrodynamic parameters of wild-type GK in the super-open and closed conformations

<table>
<thead>
<tr>
<th>pdb entry</th>
<th>Calculated Molecular weight (Da)</th>
<th>Calculated $s_{(20,w)}$ (S)</th>
<th>Calculated $D_{(20,w)}$ cm$^2$/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1V4T (super-open)</td>
<td>47 645</td>
<td>3.75</td>
<td>6.99 x 10$^{-7}$</td>
</tr>
<tr>
<td>1V4S (closed)</td>
<td>50 173</td>
<td>4.12</td>
<td>7.30 x 10$^{-7}$</td>
</tr>
</tbody>
</table>

Figure 4.15 Models generated by SOMO (125-127) to calculate the hydrodynamic parameters of human glucokinase based on their X-ray diffraction structures. (A) PDB entry 1V4T was used to generate the model. (B) PDB entry 1V4S was used to create the model.

The calculated sedimentation and diffusion coefficients for the super-open structure are in close agreement with the values determined experimentally at 50 krpm (Table 4.3 and 4.4). Although, the change in the sedimentation coefficient is expected to be distinguishable based on the X-ray diffraction data, we were not able to measure significant changes upon glucose binding. The crystal structure includes the synthetic activator molecule ($N$-Thiazol-2-yl-2-amino-4- fluoro-5-(1-methylimidazol-2-yl) thiobenzamide) bound in addition to glucose, which might suggest that glucose by itself does not induce significant changes in the hydrodynamic properties of the enzyme.
4.4 Discussion and Future Directions

In Chapter 3 we have proposed a kinetic model for glucose binding to human pancreatic glucokinase that involves the existence of two conformations of the enzyme in the absence of glucose and four binary complexes that interconvert slowly. Herein we attempted to distinguish these conformers using Analytical Ultracentrifugation and H-15N HSQC NMR techniques. First we tested our sample directly after Ni²⁺-NTA affinity chromatography and found that about 20% of our sample is aggregated. Next we used size exclusion chromatography to verify the existence of aggregates and to eliminate these species. Calibration of a Superose 6 10/300 gel filtration column (Amersham-Pharmacia) with Conalbumin (75000), Ribonuclease A (13700), Lysozyme (14 400) and Aprotinin (6500) confirmed the presence of aggregates with apparent molecular weights of 159,000 Da and 105,000 Da.

The hydrodynamic properties of purified glucokinase were investigated. Although AUC did not provide enough resolution to identify conformers of human glucokinase, we were able to determine averaged sedimentation, diffusion coefficients and shape determinants. The sedimentation coefficient is in the expected range when compared to other proteins (BSA, 66 440, 4.5 S, Albumin ovul; 44000 Da, 3.6 S) A frictional ratio, \( f/f_0 \) of 1.5 suggests that human pancreatic glucokinase is a globular protein, within the range of most globular proteins (\( f/f_0 \) of 1-2.5) (123).

Calculated molecular weights, sedimentation and diffusion coefficient values using SOMO module in the UltraScan, are in close agreement with the experimentally determined parameters. Although SOMO calculations predict a ten percent increase in the sedimentation coefficient upon glucose and synthetic activator binding (from 3.75 S to 4.12 S), upon glucose addition only, we were not able to see a noticeable difference (3.4 S to 3.5 S). One explanation could be that the resolution of AUC is not enough to distinguish the slight changes in the hydrodynamic properties of human glucokinase upon glucose binding. This is surprising considering that X-ray show a significant structural change upon glucose binding and allosteric activator binding. Unfortunately a crystal structure of a binary complex GK-glucose does not exist, which would allow us to accurately compare calculated versus measured hydrodynamic parameters. Another reasonable explanation would be that the conformation(s) attained upon glucose binding is not as compact as the one(s) observed in the presence of glucose and the allosteric activator.
Other biophysical techniques such as NMR will be employed in the future. Our preliminary results using HSQC suggest the existence of multiple species interconverting slowly, causing broadening of the peaks. Since aggregation of the sample could be another cause of the broaden HSQC spectrum, we have run the sample at 120 μM and tested its composition after the NMR experiment via AUC. No aggregates were observed via AUC (Figure 4.2), but the overlapping peaks were still present in the HSQC-NMR experiment (Figure 4.1). Therefore, we postulate that the poor HSQC spectrum formed by a cluster of five broad peaks with little dispersion might reflect multiple conformations of the human pancreatic glucokinase that are slowly interconverting. Future experiments, directed towards increasing the spectral dispersion of HSQC NMR, include specific NMR labeling of the C-terminal where large changes in chemical shift are expected due to significant environment changes in these regions upon glucose and activator binding. We did not experimentally test the case of partially unfolded protein and therefore that possibility is still another valid cause of poor NMR spectrum.

Addition of glucose to the NMR sample by itself did not disperse the peaks suggesting that the binary complex is not represented by a discrete conformation.
CHAPTER 5

THE ROLE OF THE C-TERMINAL ALPHA 13 HELIX TO THE ALLOSTERIC REGULATION OF HUMAN PANCREATIC GLUCOKINASE

ABSTRACT

Human glucokinase is a monomeric enzyme that displays a sigmoidal steady-state kinetic response toward increasing glucose concentrations. The allosteric regulation produced by glucose is postulated to arise from the slow interconversion of multiple enzyme conformations during the course of catalysis. Crystallographic data suggest that structural rearrangements linked to glucokinase cooperativity involve a substrate-induced repositioning of an alpha helix (α13) located at the C-terminus of the polypeptide. Here we show that removal of helix α13 abolishes cooperativity and restores Michaelis-Menten kinetics, while reducing the \( k_{\text{cat}} \) value of the wild-type enzyme by 160-fold. The impaired catalytic activity of the truncated enzyme is not rescued by the \( \text{trans} \) addition of a synthetic α13-helix peptide. Unexpectedly, the \( K_m \) glucose value of a glucokinase variant lacking α13 is equivalent to the \( K_{0.5} \) glucose value of the full-length enzyme. Glucokinase steady-state kinetics is unaffected by the elongation of α13 via the addition of a C-terminal polyalanine tail. To explore the link between cooperativity and the primary sequence of α13, we randomized seven residues within the helix core. Genetic selection experiments in a glucokinase-deficient bacterium identified a variety of hyperactive α13 variants that display lower \( K_{0.5} \) glucose values, Hill coefficients near unity, and enhanced equilibrium binding affinities for glucose. The present results demonstrate that α13 plays an essential role in facilitating cooperativity. Our findings also establish a link between the primary amino acid sequence of helix α13 and the functional dynamics of the glucokinase scaffold that are required for allostery.
5.1 Introduction

Glucokinase (hexokinase IV) catalyzes the ATP-dependent phosphorylation of glucose in the first and rate-limiting step of glycolysis in pancreatic β-cells (35, 43, 44). The enzyme exists as a monomer under reacting conditions (52), yet displays a sigmoidal steady-state kinetic response toward increasing glucose concentrations (51,112). This unique mode of substrate-facilitated allosteric regulation is critical to the enzyme’s physiological role in maintaining glucose homeostasis in the body. Dysfunction in glucokinase catalysis and/or regulation produces two distinct disease states, maturity onset diabetes of the young (MODY) and persistent hyperinsulinemia of infancy (PHHI) (49, 50). MODY is associated with genetic lesions in glk that impair catalytic activity, while PHHI is associated with hyperactive glucokinase variants. The central role of glucokinase in modulating glucose metabolism in the human body has generated intense interest in this enzyme as a potential therapeutic target. Indeed, a number of small-molecule glucokinase activators have been recently described, at least one of which has entered phase I clinical trials as an anti-diabetic agent (133-136).

Several models have been formulated to explain the observance of kinetic cooperativity in monomeric enzymes such as glucokinase. Two prominent theories, the mnemonic mechanism (57) and the ligand-induced slow transition (LIST) model (67), postulate that cooperativity arises from a substrate-induced conformational change that occurs with a rate constant slower than the turnover number. The existence of multiple enzyme states, combined with their failure to equilibrate during the course of catalysis, produces sigmoidicity in the reaction rate profile. In support of this postulate, Neet and coworkers have observed a slow (t_{1/2} ~ 30 s) change in the intrinsic fluorescence signal of rat liver glucokinase upon exposure to glucose (71). A pair of transient state kinetic analyses of glucose binding to human glucokinase also demonstrated that substrate association promotes a conformational change that occurs outside of the catalytic cycle (66, 72). Notably, these studies disagree as to whether the enzyme samples multiple conformational states in the absence of glucose, a mechanistic detail that differentiates the LIST and mnemonic models. A very recent analysis of glucose binding kinetics suggests that human glucokinase may sample a more diverse set of conformational states than initially postulated by the LIST and mnemonic mechanisms (73).

The three-dimensional structure of human glucokinase, reported in 2004, provided the first structural insight into the conformational changes believed to be linked to the allosteric
properties of this enzyme (Figure 5.1) (46). Kamata and coworkers successfully determined the 2.3 Å resolution crystal structure of human hepatic glucokinase in complex with glucose and a synthetic activator. These investigators also determined the structure of unliganded glucokinase at 3.4 Å resolution. The results of these studies revealed a dramatic change in enzyme structure that occurs upon glucose association. When glucose and an allosteric activator bind to the human enzyme, the smaller C-terminal domain undergoes a 99° rigid body rotation toward the N-terminal domain, which remains largely stationary. As a result of this conformational change, the enzyme adopts a more compact structure. An important feature of the glucose-induced transition is the position of helix α13. This secondary structural element, which lies at the C-terminus of the glucokinase polypeptide, moves from a solvent exposed, external orientation in the unliganded structure to an internal, sequestered position in the glucose bound state (Figure 1). The authors speculate that the ability to inhibit the release of this helix from the more compact structure may provide an explanation for the activity of known allosteric effectors. Subsequent molecular dynamics simulations support a model in which the release of helix α13 is the final step in the slow conformational transition from the closed to the super-open conformation (137).

Figure 5.1. The crystal structure of unliganded human glucokinase (left) and glucokinase in complex (right) with glucose (yellow) and an allosteric activator (not shown). Shown in red is helix α13, which moves from a solvent-exposed position in the unliganded structure to a sequestered position upon glucose binding. Images were created with PyMOL (148) and PDB entries 1V4T and 1V4S.
Several activating glk mutations map to helix α13. These include two naturally occurring substitutions, V455M and A456V, identified in patients suffering from PHHI (138, 139) and two variants, S453A and A460R, created in vitro (140, 141). In general, these single site variants display decreased cooperativity and lower glucose $K_{0.5}$ values compared to the wild-type enzyme; consistent with the hypothesis that helix α13 participates in the conformational transitions that regulate glucokinase activity in vivo. Several small molecule allosteric activators also make contact with residues located within α13, demonstrating that importance of this structural feature in therapeutic design. An additional function of α13 has been recently proposed by Flatmark and coworkers, who demonstrated that this secondary structural element constitutes part of an ubiquitin interacting motif (142). Thus, helix α13 appears to participate in the post-translational regulation of glucokinase action by mediating proteasomal degradation.

In this chapter, we investigate the extent to which helix α13 contributes to the unique kinetic characteristics of human glucokinase. We show that cooperativity can be eliminated by removing α13, or by optimizing its sequence for maximal in vivo activity. Combinatorial substitutions in the helix generate polypeptides with equilibrium glucose binding affinities that mirror the high glucose binding affinities of other non-cooperative human glucokinase isozymes. Together, these results illuminate the critical role played by α13 in facilitating cooperativity and establish a link between the primary sequence of this structural element and the dynamics of glucokinase conformational rearrangements.

5.2 Materials and Methods

5.2.1 Site-Directed Mutagenesis
The Asp205Ala variant of human pancreatic glucokinase was constructed using the Quikchange® site-directed mutagenesis strategy (Stratagene) using oligonucleotides purchased from Integrated DNA Technologies. Similarly, the Quikchange® mutagenesis protocol was used to insert two consecutive amber stop codons after residue 442 in the glucokinase coding sequence. DNA sequencing using primers in both the forward and reverse directions confirmed the successful creation of site-specific variants.

5.2.2 Protein Expression and Purification
Wild-type and variant human pancreatic glucokinases were produced as N-terminal hexa-histidine tagged polypeptides in glucokinase-deficient E. coli K-12 strain BM5340(DE3) (26). Bacterial cultures were inoculated to an initial OD$_{600}$ nm of 0.01 and were grown at 37°C in Luria-
Bertani broth supplemented with ampicillin (150 μg/mL), kanamycin (40 μg/mL) and chloramphenicol (25 μg/mL). When the OD$_{600nm}$ reached 0.85, IPTG (1 mM) was added to induce gene expression and the temperature was reduced to 20°C, where growth was continued for 20 hours. Cells were harvested by centrifugation at 8,000g and 5 g of wet cell pellet was resuspended in 17 mL of buffer A containing HEPES (50 mM, pH 7.6), KCl (50 mM), imidazole (25 mM) and glycerol (30% w/v). Cells were lysed using a French Press and subjected to centrifugation at 25000g at 4°C for 1 h. The supernatant was immediately loaded onto a 5 mL HisTrap Fast Flow Affinity Column (GE Healthcare) previously equilibrated in buffer A. Following loading, the column was washed with 10 column volumes of buffer A followed by 5 columns of buffer A containing 55 mM imidazole. Glucokinase was eluted with buffer A containing 250 mM imidazole and the enzyme was dialyzed overnight at 4°C against 1L of buffer containing HEPES (50 mM, pH 7.6), KCl (50 mM), glycerol (5% w/v) and dithiothreitol (10 mM). Due to the low expression levels of the truncated variant, an additional purification step was added to eliminate contaminants. Following affinity column purification, the truncated glucokinase variant was injected onto a Superose 6 10/300 gel filtration column (Amersham-Pharmacia) pre-equilibrated in a buffer containing HEPES (50 mM, pH 7.6), KCl (50 mM) and dithiothreitol (10 mM). The gel filtration column was run at a flow rate of 0.02 mL/min and the fractions that contained the highest A$_{280nm}$ readings were pooled and retained for further analysis.

5.2.3 Kinetic assays

The glucokinase activity of individual enzymes was measured spectrophotometrically at 340 nm by coupling the production of ADP to the oxidation of NADH via the combined action of pyruvate kinase and lactate dehydrogenase. Assays were conducted at 25°C in reaction mixtures containing HEPES (250 mM, pH 7.6), KCl (50 mM), NADH (0.25 mM), dithiothreitol (5 mM), pyruvate kinase (15 units), lactate dehydrogenase (15 units), ATP (0.1-50 mM), MgCl$_2$ (1.1-51 mM), and glucose (0.05-100 mM). Initial rate data from the first 10% of reaction progress curves were fitted to the Hill equation or the Michaelis-Menten equation, depending upon the substrate under investigation and the extent of cooperativity detected in individual enzymes. Assays involving variable glucose concentrations were conducted at a saturating concentration of ATP, and assays involving variable ATP concentrations were conducted at a saturating concentration of glucose. Assays were initiated by the addition of ATP and were conducted in duplicate for each substrate concentration. The kinetic constants reported herein are the average of data.
obtained from at least two independent preparations of each enzyme. For the truncated variant, glucokinase activity was additionally measured via the glucose-6-phosphate coupled assay.

5.2.4 Equilibrium Binding Assays
For glucose binding assays, proteins were purified as described above and were dialyzed against HEPES (50 mM, pH 7.6), NaCl (50 mM), DTT (10 mM) and glycerol (5 %) prior to data collection. Binding affinities were determined by monitoring the change in fluorescence at 335 nm that occurred in the presence of varying glucose concentrations (0.010-100 mM) following excitation of glucokinase (10 μM full-length; 30 μM truncated) at 280 nm using a 5 nm slit width. Enzyme and glucose were mixed in 0.5 mL, 1 cm pathlength cuvettes in a buffer containing sodium phosphate (5 mM, pH 7.0), KCl (25 mM), and DTT (10 mM). Reaction mixtures were allowed to equilibrate for several minutes at 25°C prior to measuring fluorescence emission intensity. Equilibrium binding experiments were performed on a Cary Eclipse Fluorescence Spectrometer housed in the Protein Biochemistry Laboratory in the Institute for Molecular Biophysics. Data were collected in duplicate, averaged, and fitted to the following equation:

\[ \Delta F = \frac{\Delta F_{\text{max}}^* [\text{glucose}]}{K_d + [\text{glucose}]} \]

5.2.5 Peptide rescue experiments
A thirteen amino acid peptide of sequence GSGRGAALVSAVA was synthesized and HPLC purified by Dr. Steven Flemer at the Protein Core Facility at the University of Vermont. The molecular weight of the peptide was determined to be 1116 atomic mass units via electrospray ionization mass spectrometry. The lyophilized peptide was dissolved in a final volume of 1 mL of ddH₂O to give a 10 mM stock. Despite the hydrophobic character of the constituent amino acids, the truncated peptide was readily soluble at this concentration. Addition of peptide in trans was done at the following molar ratios of purified truncated human glucokinase to peptide: 1:1, 1:10, 1:50, 1:100, 1:500, 1:1000. Assays were performed at 25°C and 35°C in a final volume of 0.5 mL. Reaction mixtures contained HEPES (250 mM, pH 7.6), KCl (50 mM), glucose (100 mM), ATP (30 mM), MgCl₂ (31 mM), dithiothreitol (10 mM), NADP+ (0.5 mM), glucose 6-phosphate dehydrogenase (5 units), truncated glucokinase (5 μM) and varying concentrations of the peptide. Glucokinase was pre-incubated with the peptide for 10 min and then glucose was added. The reaction was incubated for an additional 10 min prior to assay initiation via the addition of ATP. Pre-incubating the enzyme with glucose, followed by peptide addition did not change the measured reaction rates.
5.2.6 Helix randomization via gapped-duplex ligation

To facilitate the insertion of a randomized oligonucleotide into the wild-type human glucokinase gene, the Quikchange® site-directed mutagenesis procedure was used to install unique EagI and SphI sites within the coding sequence for the α13 helix. A library oligonucleotide, LIB, of sequence 5’-GGCCGGGCGCG(XNK)7TGTAAGAAGGCTGCATG-3’ was designed to encode seven contiguous random amino acids. The base composition of the XNK codons was: X = A(32%)/G(39%)/C(21%)/T(25%), N = A(25%)/G(25%)/C(25%)/T(25%), K = G(50%)/T(50%) (27). Duplex DNA was prepared by annealing the LIB oligonucleotide to two shorter oligonucleotides, HELIX1 and HELIX2, which were complementary to the non-random termini of the library oligonucleotide. This procedure generated 5’ and 3’ ends of the duplex DNA that were compatible with the EagI and SphI cleavage sites, respectively. The sequence of HELIX1 was 5’-CGCGCCCC-3’ and HELIX2 was 5’-CAGGCTTCTTACA-3’. Oligonucleotides were ordered from Integrated DNA Technologies and were phosphorylated at the 5’ end. Oligonucleotides LIB (1.5 x 10^{-11} mol), HELIX1 (2.8 x 10^{-10} mol) and HELIX2 (2.8 x 10^{-10} mol) were mixed together in annealing buffer containing Tris-HCl (20 mM, pH 7.0), MgCl₂ (2 mM) and NaCl (50 mM). The annealing reaction was heated at 65°C for 5 minutes and the reaction was allowed to cool to room temperature. The resulting duplex DNA contains a gap in the randomized region that was filled in by the endogenous DNA replication machinery of *E. coli* following library transformation.

Template plasmid pBGM101-hGK (10 μg) was digested at 37°C for 3 h with EagI (15 units) in 1X New England Biolabs reaction buffer 3. Following a 3 h incubation, SphI (5 units) was added and incubation was continued for 16 h at 37°C. An additional 5 units of SphI was added and incubation was continued for an additional 2 h to ensure complete digestion of template DNA. The restriction endonuclease digestion reaction was quenched by heating at 65°C for 10 min and the insert was removed by adding the reaction mixture to a Zeba spin column (Pierce) filled with Sephacryl S-500 resin (0.6 mL) (GE Healthcare). Prior to sample addition, the resin was prepared by centrifugation for 1 min at 2500g. Purified, restriction endonuclease treated vector DNA was eluted from the column by centrifugation for 2 min at 2500g. Duplex DNA and digested vector were ligated by incubation overnight at 14°C in a reaction mixture containing purified linear plasmid (22 μL), annealed gapped-duplex DNA (10 μL), 1X T4 DNA ligase reaction buffer and T4 DNA ligase (200 units) (New England Biolabs). Following ligation, sodium acetate (0.3 M pH 4.6) and three volumes of ice-cold ethanol (95% w/v) were added to the reaction. The reaction was incubated for 30 min at -20°C to promote DNA precipitation and the sample was centrifuged at 25,000g for 10 min. The supernatant was
removed and the pellet was washed with 0.1 mL ice-cold ethanol (70%). The pellet was allowed to air dry for 10 min at room temperature, resuspended in ddH$_2$O (15 μL), and desalted by passage across an AutoSeq G50 spin column (GE Healthcare).

5.2.7 Library Analysis
Electrocompetent BM5340(DE3) cells (25 μL) were mixed with the gapped-duplex DNA library (1 μL) and cells were transformed via electroporation. SOC (1 mL) was added and cells were allowed to recover for 1 h at 37°C. Serial dilutions were plated on LB agar containing ampicillin (150 μg/mL), kanamycin (40 μg/mL) and chloramphenicol (25 μg/ml), yielding a total of 2.0 x 10$^7$ transformants. Following overnight incubation at 37°C, twelve colonies from the non-selective LB agar plates were picked and inoculated into antibiotic supplemented LB medium (1 mL). Plasmid DNA was prepared from each culture using the Promega Wizard purification kit and DNA was sequenced at the sequencing facility at Florida State University. Sequence analysis confirmed that more than 95% of the clones harbored a randomized insert.

5.2.8 Genetic Selection
To identify library members with a viable helix sequence, the gapped-duplex library was transformed into electrocompetent BM5340(DE3) cells as described above. Following 1 h of recovery in SOC, cells were transferred to a sterile 2 mL Eppendorf tube and centrifuged for 5 min at 2000g. The supernatant was removed and the pellet was resuspended in M9 minimal medium (1 mL). Centrifugation was repeated at 2000g. The washing step was repeated with a second aliquot of M9 and the final pellet was resuspended in 500 μL of M9 minimal medium. Cells were plated in 50μL aliquots on M9 plates containing glucose (250 μM), IPTG (50 μM), MgCl$_2$ (2 mM), ampicillin (150 μg/mL), kanamycin (40 μg/mL) and chloramphenicol (25 μg/mL). Plates were incubated at 37°C and colony growth was monitored closely with the assistance of a handheld magnifying glass. A small number of colonies were visible within 48 h of plating, at which time colonies were marked. Marked colonies were picked following 72 h of incubation at 37°C with a 10 μL pipet tip. Colonies were inoculated into antibiotic supplemented LB medium (1 mL) and cultures were grown overnight at 37°C. Plasmid DNA was purified as described above and DNA was submitted for sequencing. From our 2 x 10$^7$-member library, we picked 40 of the fastest growing colonies, 35 of which returned high quality DNA sequencing results.
5.3 Results

5.3.1 Characterization of truncated human glucokinase

To probe the role of the C-terminal α13 helix in glucokinase catalysis and cooperativity, two stop codons were introduced at position 442 of the human pancreatic glk gene. The resulting gene product, which lacks the final 23 amino acids of the wild-type polypeptide, was produced in a glucokinase-deficient bacterial host to avoid contamination with endogenous glucokinases. Steady-state kinetic analysis of the purified enzyme indicated that the truncated variant retained a low level of glucokinase activity and displayed standard, hyperbolic Michaelis-Menten kinetics (Figure 5.2).

![Figure 5.2. Comparison of the steady-state kinetics of cooperative wild-type human glucokinase (A) and a non-cooperative truncated variant lacking helix α13 (B).](image)

The $k_{cat}$ value of the truncated variant was 0.18 s$^{-1}$, a value that is 160-fold lower than the $k_{cat}$ value of the full-length enzyme (Table 5.1). To verify that the low $k_{cat}$ value measured with this variant was authentic, an additional control experiment was performed. Using site-directed mutagenesis, the putative catalytic active site base, Asp205, was replaced with alanine within the truncated variant. Kinetic assays of the purified D205A variant revealed a further loss of activity (200-fold), demonstrating that the $k_{cat}$ value obtained with the truncated enzyme provides a true measure of this enzyme’s catalytic ability. The glucose $K_m$ value displayed by the truncated variant was 5 mM, which closely matches the $K_{0.5}$ value (8 mM) of the wild-type enzyme. In contrast, the ATP $K_m$ value of the truncated variant was elevated by 10-fold compared to the wild-type enzyme.
Table 5.1: Kinetic and thermodynamic parameters of wild-type GK, truncated GK, and the truncated D205A variant

<table>
<thead>
<tr>
<th>Parameter</th>
<th>wild-type GK</th>
<th>truncated GK</th>
<th>D205A truncated GK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>29 ± 1</td>
<td>0.18 ± 0.01</td>
<td>&lt; (9.1 ± 0.7) × 10$^{-4}$</td>
</tr>
<tr>
<td>$K_{0.5}$ glucose (M)</td>
<td>(8.1 ± 0.4) × 10$^{-3}$</td>
<td>(5.2 ± 0.5) × 10$^{-3}$</td>
<td>n.d.$^a$</td>
</tr>
<tr>
<td>$k_{\text{cat}}$ / $K_{0.5}$ glucose (M$^{-1}$ s$^{-1}$)</td>
<td>(3.6 ± 0.1) × 10$^3$</td>
<td>(3.5 ± 0.1) × 10$^1$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Hill coefficient</td>
<td>1.7 ± 0.1</td>
<td>0.9 ± 0.05</td>
<td>n.d.</td>
</tr>
<tr>
<td>$K_D$ glucose (M)</td>
<td>(2.9 ± 0.2) × 10$^{-3}$</td>
<td>(7.6 ± 2.4) × 10$^{-4}$</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

$n.d.$ - not determined

We assessed the ability of truncated glucokinase activity to be rescued via the trans addition of a thirteen amino acid peptide with a sequence that matches that of the endogenous α13 helix. Incubation of the truncated variant with increasing concentrations of the purified synthetic peptide at ratios ranging from 1:1 to 1:1000 did not reveal substantial levels of reactivation. The turnover number for glucose phosphorylation measured at 25°C showed a modest 1.6-fold increase in the presence of high concentrations of peptide (5 mM). Increasing the temperature of the incubation to 35°C did not significantly increase the $k_{\text{cat}}$ value of the truncated enzyme (Table 5.2).

Table 5.2: Effects on $k_{\text{cat}}$ upon Peptide Addition to Truncated GK at 25°C and 35°C

<table>
<thead>
<tr>
<th>Molar ratio wt-GK : peptide</th>
<th>Temperature (°C)</th>
<th>Fold increase in $k_{\text{cat}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:0</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>1:1</td>
<td>25</td>
<td>1.1</td>
</tr>
<tr>
<td>1:10</td>
<td>25</td>
<td>1.2</td>
</tr>
<tr>
<td>1:100</td>
<td>25</td>
<td>1.2</td>
</tr>
<tr>
<td>1:500</td>
<td>25</td>
<td>1.5</td>
</tr>
<tr>
<td>1:1000</td>
<td>25</td>
<td>1.6</td>
</tr>
<tr>
<td>1:500</td>
<td>35</td>
<td>1.4</td>
</tr>
</tbody>
</table>
To investigate the effect of α13 removal upon the equilibrium binding affinity for substrate glucose, the fluorescence emission spectrum of truncated glucokinase was obtained in the absence and presence of 100 mM glucose. High concentrations of enzyme (30 μM) were required in these experiments, as the emission spectrum of the truncated variant displayed a very small increase in intensity following glucose addition. A shift toward shorter wavelengths was observed in the emission spectrum of the truncated variant upon exposure to glucose (Figure 5.3). These observations contrast with the effect of glucose binding to the wild-type enzyme, which causes a significant increase in the fluorescence intensity and does not alter the \( \lambda_{\text{max}} \) of the emission spectrum.

Figure 5.3. Comparison between the fluorescence spectrum of wild-type GK and the truncated GK. (A) Fluorescence emission spectrum of wild-type human glucokinase (1 μM) in the absence (black) and presence (red) of 0.1 M glucose. (B) Fluorescence emission spectrum of truncated human glucokinase (30 μM) in the absence (black) and presence (red) of 0.1 M glucose.

5.3.2 Elongation of the α13 helix

A comparison of the crystal structure of human glucokinase in the presence and absence of bound glucose suggested that lengthening helix α13 might impede the interconversion of these two conformational states (Figure 5.4) (46). To investigate that possibility, we added 5 and 10 alanine residues to the C-terminus of the wild-type glucokinase polypeptide. Steady-state kinetic assays of the purified elongated variants revealed that neither addition significantly altered the kinetic constants of the wild-type enzyme (Table 5.3). The cooperativity of the enzyme was also unaffected by α13 elongation.
The position of helix α13 (red) and the connecting loop (blue) in the structure of (A) unliganded human glucokinase and (B) glucokinase crystallized in the presence of glucose and an allosteric activator. This comparison suggests that elongation of α13 might impede the conformational transition between these states. Images were created with PyMOL (148) and PDB entries 1V4T and 1V4S.

Table 5.3: Catalytic effects of C-terminal polyalanine addition

<table>
<thead>
<tr>
<th>Parameter</th>
<th>wild-type GK</th>
<th>GK + 5 alanine</th>
<th>GK + 10 alanine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ (s⁻¹)</td>
<td>29 ± 0.1</td>
<td>37 ± 5</td>
<td>38 ± 8</td>
</tr>
<tr>
<td>$K_{0.5 \text{glucose}}$ (M)</td>
<td>$(8.1 ± 0.4) \times 10^{-3}$</td>
<td>$(10 ± 1) \times 10^{-3}$</td>
<td>$(10.5 ± 0.5) \times 10^{-3}$</td>
</tr>
<tr>
<td>$k_{cat}/K_{0.5 \text{glucose}}$ (M⁻¹ s⁻¹)</td>
<td>$(3.6 ± 0.1) \times 10^{3}$</td>
<td>$(3.7 ± 0.1) \times 10^{3}$</td>
<td>$(3.6 ± 0.7) \times 10^{3}$</td>
</tr>
<tr>
<td>Hill coefficient</td>
<td>1.7 ± 0.1</td>
<td>1.6 ± 0.2</td>
<td>1.7 ± 0.1</td>
</tr>
<tr>
<td>$K_{m \text{ATP}}$ (M)</td>
<td>$(5.0 ± 0.1) \times 10^{-4}$</td>
<td>$(5.2 ± 0.5) \times 10^{-4}$</td>
<td>$(4.2 ± 0.1) \times 10^{-4}$</td>
</tr>
<tr>
<td>$k_{cat}/K_{m \text{ATP}}$ (M⁻¹ s⁻¹)</td>
<td>$(5.6 ± 0.1) \times 10^{4}$</td>
<td>$(6.9 ± 1.2) \times 10^{4}$</td>
<td>$(8.9 ± 0.9) \times 10^{4}$</td>
</tr>
</tbody>
</table>

5.3.3 Randomization of the α13 helix

To explore the sequence requirements of α13, we constructed a genetic library in which all twenty amino acids were allowed to occupy seven consecutive residues (450-456) within the helix core. Transformation of this library into the glucokinase-deficient BM5340(DE3) strain yielded 2 x 10⁷ members, representing 1.5% of the theoretical 1.28 x 10⁹ member library. Sequencing of clones isolated from non-selective medium confirmed the randomness of the library. To identify active α13 variants, transformed BM5340(DE3) cells were challenged for growth on glucose minimal medium. Forty of the fastest growing colonies were picked and the plasmid DNA was sequenced to determine the identity of α13 residues. High quality sequencing data was returned from thirty-five selected clones. In general, the substitutions observed in positions 450-456 were dominated by amino acids that have a statistically significant preference...
for adopting an alpha helical structure. From a total of 245 different positions observed in the thirty-five selected enzymes, 69% (168/245) were occupied by one of seven helix stabilizing amino acids (A, E, K, L, M, Q, R) (28). An additional 18% (43/245) of positions were filled with residues that lack a strong conformational preference, while 14% were occupied by residues that could be considered helix destabilizing (G, C, N, P, S, T, Y). Charged amino acids were observed in 4.4% (11/245) of selected helix residues, whereas 76% of the 245 positional variants were occupied by hydrophobic amino acid.

Table 5.4. The identity and frequency of amino acids observed at positions 450-456 following genetic selection for *E. coli* growth of a randomized α13 helix library. Previously described activating substitutions are underlined. Substitutions that are expected to stabilize the α13 helix based upon amino acid conformational preferences are denoted with an asterisk.

<table>
<thead>
<tr>
<th>Alanine 450</th>
<th>Leucine 451</th>
<th>Valine 452</th>
<th>Serine 453</th>
<th>Alanine 454</th>
<th>Valine 455</th>
<th>Alanine 456</th>
</tr>
</thead>
<tbody>
<tr>
<td>66% Ala*</td>
<td>80% Leu*</td>
<td>51% Leu*</td>
<td>74% Ala*</td>
<td>37% Met*</td>
<td>20% Arg*</td>
<td>26% Leu*</td>
</tr>
<tr>
<td>34% Gly</td>
<td>8.5% Met*</td>
<td>23% Val</td>
<td>20% Val</td>
<td>23% Ala*</td>
<td>20% Gln*</td>
<td>20% Met*</td>
</tr>
<tr>
<td>8.5% Ile</td>
<td>17% Ile</td>
<td>3% Cys</td>
<td>11% Arg*</td>
<td>17% Gly</td>
<td>17% Val</td>
<td></td>
</tr>
<tr>
<td>3% Phe</td>
<td>6% Ala*</td>
<td>3% Ser</td>
<td>11% Ser</td>
<td>11% Ala*</td>
<td>14% Ile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3% Cys</td>
<td>6% His</td>
<td>6% Asn</td>
<td>6% Ala*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6% Leu*</td>
<td>6% Cys</td>
<td>6% Thr</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3% Gln*</td>
<td>6% Met*</td>
<td>3% Gln*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3% Tyr</td>
<td>3% Asp</td>
<td>3% Glu*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3% Ile</td>
<td>3% His</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3% Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3% Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3% Tyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.5: Kinetic and thermodynamic characterization of selected α13 variants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>wild-type (ALVSAVA)</th>
<th>variant #1 (ALIAAAAV)</th>
<th>variant #2 (ALLAMAV)</th>
<th>variant #3 (ALVASRL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\text{cat}}$ (s$^{-1}$)</td>
<td>29 ± 0.1</td>
<td>17 ± 2.4</td>
<td>18 ± 0.2</td>
<td>29 ± 0.2</td>
</tr>
<tr>
<td>$K_{0.5 \text{glucose}}$ (M)</td>
<td>(8.1 ± 0.4) × 10$^{-3}$</td>
<td>(1.0 ± 0.1) × 10$^{-3}$</td>
<td>(0.8 ± 0.1) × 10$^{-3}$</td>
<td>(1.4 ± 0.1) × 10$^{-3}$</td>
</tr>
<tr>
<td>$k_{\text{cat}}/K_{0.5 \text{glucose}}$ (M$^{-1}$ s$^{-1}$)</td>
<td>(3.6 ± 0.1) × 10$^{3}$</td>
<td>(1.7 ± 0.1) × 10$^{4}$</td>
<td>(2.2 ± 0.3) × 10$^{4}$</td>
<td>(2.2 ± 0.3) × 10$^{4}$</td>
</tr>
<tr>
<td>Fold Increase $k_{\text{cat}}/K_{0.5 \text{glucose}}$</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Hill coefficient</td>
<td>1.7 ± 0.1</td>
<td>1.2 ± 0.2</td>
<td>1.1 ± 0.2</td>
<td>1.2 ± 0.1</td>
</tr>
<tr>
<td>$K_{\text{m ATP}}$ (M)</td>
<td>(5.0 ± 0.1) × 10$^{-4}$</td>
<td>(1.6 ± 0.1) × 10$^{-4}$</td>
<td>(1.6 ± 0.1) × 10$^{-4}$</td>
<td>(2.6 ± 0.1) × 10$^{-4}$</td>
</tr>
<tr>
<td>$k_{\text{cat}}/K_{\text{m ATP}}$ (M$^{-1}$s$^{-1}$)</td>
<td>(5.6 ± 0.1) × 10$^{4}$</td>
<td>(1.3 ± 0.1) × 10$^{5}$</td>
<td>(1.1 ± 0.1) × 10$^{5}$</td>
<td>(1.2 ± 0.2) × 10$^{5}$</td>
</tr>
<tr>
<td>$K_{\text{D glucose}}$ (M)</td>
<td>(2.9 ± 0.2) × 10$^{-3}$</td>
<td>(4.9 ± 1.9) × 10$^{-5}$</td>
<td>(5.3 ± 0.9) × 10$^{-5}$</td>
<td>(6.7 ± 0.2) × 10$^{-4}$</td>
</tr>
<tr>
<td>Fold decrease $K_{\text{D glucose}}$</td>
<td>1</td>
<td>60</td>
<td>55</td>
<td>4</td>
</tr>
</tbody>
</table>

The primary sequence of clones that provided high quality sequencing data is summarized in Table 5.4. Residue 450, an alanine in the wild-type enzyme, was highly conserved in the selected α13 variants. Alanine (66% occupancy) and glycine (34% occupancy) were the only permissible amino acids at this position. Residue 451 was also highly conserved. 80% of the selected α13 helix variants possessed the native leucine at this position. The occupancy of residue 452 was dominated by three hydrophobic amino acids, leucine, isoleucine and the native valine.
Leucine was the most common substitution observed at residue 452, with 51% of selected sequences possessing this amino acid. Residue 453, occupied by serine in the wild-type enzyme, was nearly always substituted in the selected variants, most often by alanine (74% occupancy) or valine (20%). Residues 454-456 were more highly variable than residues 450-453. At residue 454, methionine was observed in 37% of selectants, while the native alanine was retained in 23% of selected clones. Twelve different amino acids were observed at residue 455, with arginine occurring most frequently (20% occupancy). Notably, the wild-type valine was never observed in our selected sequences, and the known activating substitution V455M was only observed twice. Nine different amino acids were found to occupy residue 456, the final position randomized in our library. A naturally occurring A456V variant has been observed in patients suffering from PHHI. This substitution was observed in 17% of our selectants; however, other hydrophobic amino acids including leucine, isoleucine and methionine account for an additional 60% of residue 456 occupants.

To evaluate the kinetic and thermodynamic properties of α13 variants identified by our genetic selection experiments, we chose three enzymes for further characterization that best represented the most common positional substitutions observed in our selected library members. The primary sequence of the randomized region of the chosen variants was: ALIAAAV, ALLAMAV, and ALVASRL. Steady-state assays of these three variants showed that each displayed a 6-10 fold decrease in the wild-type $K_{0.5\text{ glucose}}$ value, while none possessed a substantially altered $k_{cat}$ value (Table 5). The degree of cooperativity, as measured by the Hill coefficient, is substantially decreased in each of the α13 variants and the ATP $K_m$ value is decreased by approximately 3-fold for each variant. In general, these kinetic characteristics match those observed in other activated human glucokinase variants. Optimizing the sequence of helix α13 also had a striking effect upon the equilibrium binding affinity for glucose. Compared to the wild-type enzyme, the ALIAAAV and ALLAMAV variants showed a 60-fold increase in glucose binding affinity, while the ALVASRL variant displayed a more modest 4-fold increase in affinity.

5.4 Discussion

The purpose of the present study was to investigate the contribution of the C-terminal α13 helix of human glucokinase to the kinetic characteristics of this enzyme. Here we show that loss of helix α13 produces a large decrease in the $k_{cat}$ value of the wild-type enzyme,
demonstrating the essentiality of this structural element to catalytic turnover. We also observe a complete loss of kinetic cooperativity upon removal of helix α13. The mnemonic and LIST models postulate that monomeric cooperativity in glucokinase results from a slow alteration in enzyme structure (57, 67). The reduced $k_{cat}$ value of the truncated enzyme suggests that the rate of the chemical step is sufficiently reduced such that the conformational changes associated with glucose binding and/or product release have time to reach equilibrium. This appears to be the most probable explanation for the lack of cooperativity displayed by the truncated glucokinase variant. Alternatively, elimination of helix α13 may prevent the formation of alternate enzyme conformations. Qualitatively, this postulate appears to be supported by the failure of glucose to induce a significant change in intrinsic protein fluorescence in the truncated variant. However, the latter observation can also be explained, in part, by the fact that α13 is located in close proximity to W99, a residue that contributes substantially to the overall fluorescence intensity of wild-type glucokinase (75). Thus, α13 removal likely causes a local perturbation in the solvent accessibility of W99, which may explain why the fluorescence emission spectrum of the truncated variant changes little upon glucose binding.

Cooperativity is also reduced when the primary amino acid sequence of α13 is optimized for maximal in vivo activity following genetic selection. On the basis of the results tabulated in Table 4, maximal activity appears to correlate with a core α13 sequence that is dominated by hydrophobic amino acids that possess a conformational preference for adopting a helical structure. To understand how cooperativity is lost in the sequence-optimized α13 variants, it is useful to examine the details of both the mnemonic and LIST mechanisms. In the mnemonic mechanism, the conformational change that produces steady-state cooperativity occurs at the conclusion of the catalytic event, when the enzyme “forgets” its active conformation and slowly relaxes to an alternate conformation. In this model, cooperativity can be reduced by a perturbation of the rate constant that governs the relaxation event. For example, the Hill coefficient will approach unity if the relaxation rate is enhanced such that the conformational states of the enzyme equilibrate prior to the initiation of another catalytic cycle. Alternatively, the Hill coefficient will approach unity if the relaxation rate is reduced to such an extent that the alternate enzyme conformation is rarely sampled. A similar situation facilitates a loss of cooperativity in the LIST model, in which the enzyme is postulated to exist in multiple catalytically competent conformations prior to glucose binding. An acceleration of the rate of conformational interconversions, or a perturbation in the conformational equilibrium constant in favor of a single state, is expected to produce a Hill coefficient near unity. The hyperactive α13 variants characterized in this study display a substantial increase in glucose binding affinity,
similar to those observed in other hexokinases that do not appear to sample multiple unliganded conformations. On the basis of this observation, we postulate that an optimal α13 sequence may eliminate cooperativity by altering the conformational equilibrium of human glucokinase in favor of fewer states.

As shown in Figure 5.4, α13 moves from a solvent exposed position in the absence of substrate to a sequestered position when both glucose and an allosteric activator are bound to the enzyme. To obtain the bound structure, α13 appears to insert itself behind the connecting loop, which is comprised of residues 63-72. One full turn of the α13 helix extends beyond the border established by the position of the connecting loop. Based on this observation, we postulated that extending α13 would either (a) trap the enzyme in the bound state by preventing helix release, or (b) prevent formation of this state by impeding the insertion of the helix behind the connecting loop. Either of these situations is expected to alter the kinetic characteristics of the enzyme; however the enzyme was impervious to the addition of alanine residues to the C-terminus. This finding is consistent with previous reports indicating that a fully active and cooperative enzyme can be obtained when human glucokinase is produced as a C-terminal fusion with the 25-kDa glutathione S-transferase (143). These observations are difficult to reconcile with the picture of α13-associated conformational changes that emerge from a comparison of the unliganded and glucose-bound structures. It is worth noting that the liganded structure depicted in Figure 4B includes an allosteric activator molecule not shown, which forms contacts with residues in both the helix and the connecting loop. It is likely that the presence of the activator influences the relative positions of the helix and/or the loop. Given these considerations, we speculate that the relative orientation of the connecting loop and α13 is different from those depicted in the currently available crystal structures when the catalytically relevant, glucose bound state is achieved. A full understanding of the interplay of α13 with the connecting loop at various stages of the catalytic cycle awaits the determination of the structure of glucose-bound enzyme in the absence of an activator molecule.

The ability of helix α13 to control the allosteric properties of human glucokinase appears to be a common feature of this polypeptide scaffold. Human hexokinase I, an isozyme of glucokinase, is a 100 kDa enzyme composed of two fused 50 kDa domains connected by a helix analogous to α13 (Figure 5.5) (35). The N-terminal domain of hexokinase I is inactive, but contains a binding site for product glucose 6-phosphate, an allosteric effector that alters activity in the C-terminal catalytic domain. When glucose-6-phosphate binds to the allosteric site it forms a hydrogen bond with the side chain of Ser449, which is located in the first turn of the connecting helix (144). This interaction initiates a 6° rotation of the N-terminal domain with
respect to the catalytic C-terminal domain, a conformational change that is propagated to the active site via the alteration of several interdomain contacts. Elongation of the connecting helix, or disruption of the helical structure via destabilizing substitutions, has been shown to impede allosteric communication and disrupt the ability of inorganic phosphate to relieve feedback inhibition caused by glucose-6-phosphate binding to the N-terminal allosteric domain (145-147). Although this regulatory mechanism is distinct from the α13-associated slow conformational changes that produce cooperativity in glucokinase, the role of the helix in facilitating the allosteric properties of each enzyme is functionally conserved.

Figure 5.5. The structure of human hexokinase I, an isozyme of glucokinase, depicting the location of the connecting helix (red) that facilitates communication between the allosteric N-terminal domain and the catalytic C-terminal domain (148).

The findings presented herein demonstrate the importance of helix α13 to the allosteric properties of human glucokinase. Because cooperativity is so intimately linked to both the presence and primary amino acid sequence of α13, this structural element appears to be an ideal site for the attachment of probes to report on the molecular details of enzyme functional dynamics. The placement of spectroscopic labels or NMR sensitive nuclei within α13 promises
to facilitate the direct measurement of the rates and magnitudes of structural alterations during various stages of the catalytic cycle. Such information is essential to assembling a complete kinetic, structural and mechanistic description of cooperativity in this monomeric enzyme.
Most contemporary enzymes possess the ability to recognize subtle structural differences in individual substrate molecules, thus allowing them to be highly specific for their native substrate. Substrate specificity is defined by the ratio of second-order rate constants, $k_{cat}/K_m$, that govern the transformations of two competing substrates. The evolutionary origins of substrate discrimination are of particular interest to our laboratory. The sugar kinases encompassed by the ROK superfamily offer a unique opportunity to explore the evolutionary origins of enzyme specificity. Previous work in our laboratory has identified a divergent superfamily of bacterial sugar kinases that possess ambiguous substrate specificities toward the alternate substrate, D-glucose. In chapter 2, we investigated whether common mechanism(s) could be responsible for the specialization of glucokinase activity during the early stages of evolutionary process. Our results suggest that in the ROK-family, functional divergence is achieved via evolutionary processes that include the formation of intermediates possessing relaxed substrate specificities. Our findings are in agreement with the hypothesis introduced by Jensen that evolution of new enzymatic activity has its origin in catalytic promiscuity and/or substrate ambiguity (20). Diverse superfamilies have been characterized and suggest that nature has used common binding sites and mechanistic features to catalyze analogous reactions with structurally different substrates. The ROK superfamily adds to the myriad of examples of superfamilies such as enolases and α/β-hydrolases, where structural and mechanistic conservation suggest that they evolved from a common ancestor. Enzymes like alkaline phosphatase and adenylate kinase are thought to have evolved from ancestral enzymes possessing low levels of phosphodiesterase activity and phosphoryl transfer activity, respectively, when a selective pressure for these reactions was applied.

Our broader aim was to investigate the role of induced fit and conformational changes to the substrate specificity in AlsK and NanK. Hedstrom et al., in 2001, (97) on the trypsinogen system, suggest that induced fit does not provide discrimination for two substrates. Instead the zymogen activity is controlled by a conformational equilibrium between an inactive state and a single active conformation. In contrast, a more recent study on T7 DNA polymerase (98) suggests that this enzyme adopts two different ground state conformations, one that is
committed to catalysis when proper substrate is bound and the other that possesses a fast substrate dissociation rate when a mismatch is bound.

In our attempt to determine the structural, kinetic and thermodynamic determinants of substrate specificity in the ROK superfamily we have mapped the substitutions onto the structure of a ROK protein and we have compared the steady-state kinetic parameters for different substrates. We have found two hot spots targeted for mutations during the early stages of evolutionary processes. Mechanistically, the lack of an intrinsic fluorescence signal for AlsK and NanK did not allow us to look deeper into the contribution of the microscopic rate constants to the specificity constant. Thus, in the future, a comparison of the contribution of the microscopic rate constants to the specificity constant in the glucose phosphorylation versus allose and N-acetyl-mannosamine phosphorylation reactions by AlsK and NanK promises to reveal the kinetic basis for specificity in this family of enzymes. Such experiments could be conducted via specific fluorophore labeling. Induced fit sugar kinases such as AlsK and NanK transform various carbohydrates with widely differing catalytic efficiencies. To identify the thermodynamic basis for this difference, we could determine the entropy and enthalpy of activation for the phosphorylation of a series of carbohydrate substrates whose $k_{\text{cat}}$ values are limited by a chemical step. Entropies and enthalpies of activation could be determined by measuring the temperature dependence of $k_{\text{cat}}$ for the phosphorylation of 1,5 anhydro-D-glucitol, D-glucose, D-mannose by AlsK and NanK. Control experiments to prove that chemical step is rate limiting in those reactions have to precede the temperature dependence studies.

Human glucokinase is an exception in highly evolved enzymes, because it is not highly specific for its native substrate. The physiological role of the enzyme is not to transform every substrate molecule that it encounters, but instead to maintain glucose concentration at 5 mM. To achieve this, human glucokinase, a monomeric enzyme with one active site, has evolved sigmoidal kinetic behavior. Our goal is to understand the kinetic and molecular mechanisms responsible for generating kinetic cooperativity. Two theoretical models have been postulated to explain the sigmoidal behavior of GK, and each suggests that GK must exist in two (or more) distinct conformations. The mnemonic model makes predictions about the relationship between the bimolecular binding events, while the LIST model emphasizes the isomerization step(s) as the main contributor to cooperativity. Computer simulations of steady state and transient state kinetics from various groups are available and have been summarized in the introduction.

A combination of methodologies is used to accomplish a molecular and kinetic understanding of GK kinetic cooperativity including mutagenesis, in vivo functional selections, protein expression and purification, pre-steady state kinetic characterization of glucose binding,
comparative steady-state assays, analytical ultracentrifugation, and nuclear magnetic resonance.

In Chapter 3 of this dissertation, we attempt to explain our transient state kinetics of glucose binding to GK, via computer simulations. Although other groups have used similar methods, limiting assumptions were made during their calculations. The simulations of the transient state kinetics described herein were mainly focused on fitting the observed rate constants as a function of glucose concentration. Based on the ability of Pro-KII to fit the glucose binding curves simultaneously, we generated a model that resembles the LIST predictions. We found that two species of GK are needed in the absence of glucose to obtain the best fit, and we found that the isomerization step is rate-limiting. The weakness of our model relies upon the initial assumption of the existence of two species, and the value of the equilibrium constant that describes them. Thus, we wish to be able to measure the rates of GK conformer’s interconversion and/or their concentration via HSQC-NMR. Based on our model, addition of glucose alone does not induce only one conformation. To test the hypothesis that multiple conformations are restricted to one or two upon ternary complex formation and reduce the broadening in the HSQC spectrum, we will record the HSQC spectrum of GK in the presence of glucose and Cr(III)-ATP, Al(III)-ATP or a non-hydrolyzable ATP. If spectral dispersion is seen and we observe 464 well-resolved peaks, we could rationalize it by the formation of a dead end ternary complex with a unique GK conformation.

Additional problems in working with GK arose from the presence of protein aggregation, as identified by gel filtration and analytical ultracentrifugation. Chapter 4 describes the identification and elimination of these aggregates from GK samples prior to NMR experiments. Preliminary HSQC experiments were conducted, but the presence of a poorly resolved spectrum lead to two conclusions: a) a partially unfolded protein and b) multiple GK conformers slowly interconverting. Thus, our future efforts will focus on finding conditions that provide a well-resolved spectrum. One approach is to decrease the number of peaks expected to give a signal via specific labeling of the C-terminal domain, which is known to undergo large conformational changes upon ligand binding. The identification of different GK conformations in the absence of glucose promises to shed light onto the theoretical debate between the mnemonic and LIST mechanisms. The new view of protein dynamics proposes that proteins are characterized by a pre-existing equilibrium between functional and non-functional conformers. Binding of ligands or regulators to the functional form shifts that equilibrium towards the needed conformation. Our computer simulations predict that human glucokinase exists in two conformers in the absence of glucose, two collision complexes and two additional binary
complexes. The identification of such conformers via other methods promises to allow us to investigate the role of protein dynamics in glucokinase catalysis. Using transient state kinetics we have proposed that multiple conformers exist in the absence of substrate and upon glucose binding. Our model is intended to be a qualitative estimate of the glucose binding mechanism. Knowing the number of conformers and their equilibrium constant would allow for a precise regulation of GK function via external activators, which can be directed towards perturbing their equilibrium constant.

In chapter 5, we focused on the role of the $\alpha_{13}$ helix to the kinetic cooperativity. Crystallographic data suggest that structural rearrangements linked to glucokinase cooperativity involve a substrate-induced repositioning of an alpha helix ($\alpha_{13}$) located at the C-terminus of the polypeptide. We found that removal of helix $\alpha_{13}$ abolishes cooperativity and restores Michaelis-Menten kinetics. We have also found that optimizations in the helix $\alpha_{13}$ lead to hyperactive variants displaying loss of cooperativity. Our results demonstrate that $\alpha_{13}$ plays an essential role in facilitating cooperativity. The study also establishes a link between the primary amino acid sequence of helix $\alpha_{13}$ and the functional dynamics of the glucokinase scaffold that are required for allostery. Future studies could be extended towards the placement of specific labels, such as fluorophores or NMR probes, in helix $\alpha_{13}$ with the aim of uncovering the rates of GK interconversions.

This study adds insights into the structural role of helix $\alpha_{13}$ to the attainment of conformations associated with allostery and efficient catalysis. Our conclusions are unique with respect to the image of the binary complex GK-glucose and the transitions associated with the formation of the complex. A simple comparison of the crystal structures available can be misleading since it does not reflect the true picture of GK structures in solution. This idea is sustained by the elongations studies in the alpha-13 helix which have no effect on the kinetics of the enzyme. Thus, the position of helix alpha 13 with respect to the connecting loop in the binary complex has to be different than those readily represented in the available crystal structures.

Mechanistically, a pre-existing equilibrium in the absence of glucose is needed for computer fitting of the glucose binding traces. Glucose binding does not lead to a unique binary complex. Preliminary HSQC experiments indicate the possibility of multiple conformers slowly interconverting in the absence of glucose. Since partially unfolded proteins give rise to the same HSQC pattern, additional experiments are needed to investigate this possibility. Analytical ultracentrifugation was conducted on the same NMR sample after the experiment and established that aggregation is not present in our HSQC-NMR sample. Furthermore, addition of
glucose alone to the mixture of GK conformers does not alter the average hydrodynamic properties of the unliganded population enough to be detected in the AUC experiments. Here again, comparison of expected hydrodynamic parameters is based upon the crystal structures available.

A detailed kinetic characterization of GK has to include the investigation of the chemical step, another distinction setting apart the mnemonic and LIST model. The mnemonic model predicts only one GK transition state is formed while the LIST mechanism suggests two operational catalytic cycles. Thus, determining if one or two conformations of GK undergo catalysis constitutes another future aim of my research. Chemical quench flow under single turnover conditions will be used to determine of the ratio of enzyme: product at different glucose concentrations. The relaxation step of GK conformation following catalysis and the release of products is assumed to be very slow in the mnemonical model. Preincubating the GK with the last product to be released is expected to stabilize this kinetically active conformation and to prevent the relaxation to the thermodynamically favored conformation. Pre-steady state glucose binding traces in presence of ADP (the last product to be released) are expected to be comprised of fewer transients due to the elimination of the relaxation step.
REFERENCES


40. Kang, L., Dunn-Meynell, A., Routh, V. H., Gaspers, L. D., Nagata, Y., Nishimura, T., Eiki, J.,
hypothalamic neuronal glucosensing, *Diabetes* 55, 412-420.


42. Matschinsky, F. M. (1990) Glucokinase as Glucose Sensor and Metabolic Signal Generator


43. Lowry, O. H. and Possonneau, J. V. (1964). The relationships between substrates and


44. Ferre, T., Riu, E. Bosch, F. and Valera, A. (1996) Evidence from transgenic mice that


45. Magnani, M., Bianchi, M., Casabianca, A., Stocchi, V., Daniele, A., Alturda, F., Ferrone, M.,

and Silengo, L. (1992) A recombinant human 'mini'-hexokinase is catalytically active and


allosteric regulation of the monomeric allosteric enzyme human glucokinase, *Structure* 12, 429-

439.


49. Vionnet, N. Scoffel, M., Takeda, J., Yasuda, K., Bell, G. I., Zouali, H., Lesage, S., Velho, G.,


diabetes of young, permanent neonatal diabetes and hyperinsulinemia of infancy. *Hum. Mut.* 22,

353-362.


interactions with glucose at physiological significant concentrations, *Biochem J*. 159, 7-14.


Mioara Larion was born in 1980, in Vaslui, Romania, to Stefan and Rodica Larion. She attended Mihail Kogalniceanu high school from 1994 to 1998. In 1998, she was accepted in the Medical School and Chemistry Universities in Iasi but chose to pursue a degree in Chemistry at the Al. I. Cuza University, Iasi, Romania. She graduated in June 2002, and moved to Tallahassee where she began graduate studies at Florida State University in the Department of Chemistry and Biochemistry. In summer 2003 she joined Dr. Peter Fajer group where she conducted two independent projects entitled:

1. Sensitivity of Saturation Transfer Electron Paramagnetic Resonance to the rate of Motion at X and W-band

2. Experimental Validation of MOMD-SLE Method using Staphylococcus Nuclease as a Model Protein.

She graduated with a Master of Science in the field of Chemistry in Fall 2005 and joined the laboratory of Dr. Brian Miller where she worked on: probing the evolutionary events that occur during the optimization of glucokinase activity in two members of the ROK superfamily, Alsk and NanK; investigating the conformational heterogeneity in the unliganded human glucokinase and the role of helix α13 in governing the conformational transitions responsible for generating kinetic cooperativity.