THE ENTROPIC FORCE GENERATED BY DISORDERED PROTEIN SEGMENTS

by

NICHOLAS DONALD KEUL

(Under the Direction of Zachary A. Wood)

ABSTRACT

Protein structures are inherently dynamic and can adopt a wide range of conformations. The ensemble of structural substates that a given protein can explore embodies a conformational landscape, which is essential for various functions (e.g.; binding interactions, catalysis, and regulation). How evolution shapes this conformational landscape to tune specific functions is perplexing. One of the main constraints in the evolution of proteins is the stability of the 'native' folded structure. However, some proteins function in the absence of a stable fold and the persistence (>44%) of intrinsically disordered segments in the human proteome is striking. The bulk of these intrinsically disordered segments have no known function and have been historically regarded as non-functional. Our work shows that the entropic force generated by the intrinsically disordered C-terminus (ID-tail) in human UDP-α-D-glucose-6-dehydrogenase (hUGDH) biases the conformational landscape to favor a specific substate with higher affinity for an allosteric inhibitor. The function of the ID-tail is independent of both its sequence and chemical composition. Instead, the ID-tail tunes the affinity for the allosteric inhibitor in a length-dependent fashion. This result is consistent with the folded structure incurring an entropic cost by constraining the ID-tail. Our results show that this entropic force generated by the ID-tail alters the dynamics and structure of hUGDH, which favors a substate with higher affinity for the allosteric inhibitor. This entropic

force mechanism alters the conformational landscape of the protein to tune a specific function and has no specific sequence or structural constraints. Thus, our model highlights the utility of intrinsically disordered segments in protein evolution, which may explain the frequency of intrinsic disorder in the proteome.

INDEX WORDS: protein structure, protein dynamics, conformational ensemble, entropy, intrinsic disorder, intrinsically disordered segments, intrinsically disordered regions, IDRs, entropic rectifier, allostery, udp-glucose dehydrogenase

THE ENTROPIC FORCE GENERATED BY DISORDERED PROTEIN SEGMENTS

by

NICHOLAS DONALD KEUL

B.S., Virginia Polytechnic Institute and State University, 2013

A Dissertation Submitted to Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

THE ENTROPIC FORCE GENERATED BY DISORDERED PROTEIN SEGMENTS

by

NICHOLAS DONALD KEUL

Major Professor: Zachary A. Wood

Committee: William N. Lanzilotta
James H. Prestegard
Kelly W. Moremen

Electronic Version Approved:
Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2019

DEDICATION

To my family for their endless love and support: David Keul, Julie Crifasi, Joe Crifasi, Kelly Choate, Patrick Keul, Savannah Choate, Ciera Oldham, Josh Oldham, Alice Oldham, and Megan Wallace.

To my friends for their inspiration and support these many years: Ryan Mullaney, Carey Graham, Steven Berardinelli, Weston McDonald, Renuka Kadirvelraj, Branson Perkins, Nathan Beattie, Tony Szempruch, Joey LaMattina, Tyler Boyer, Johnathan Kent, Kevin McClelland, Andrew McNulty, John Hooven, John Kim, Nitin Daniel, Brandon Burch, Alex Burch, Robert Amos, Jeronimo Silva, Tyler Carter, and Mike Byers.

ACKNOWLEDGEMENTS

With utmost respect, I first thank my mentor Dr. Zachary Wood for sculpting me into the scientist that I am today. His teaching and mentorship abilities are unparalleled, and without him none of the work presented in this dissertation would have been possible. His enthusiasm for science is contagious and he always brought the light in darker times during my graduate career (e.g.; when I spent a year as a teaching assistant or during our long year of peer review with *Nature*). I also extend my gratitude to my committee members: Dr. William N. Lanzilotta, Dr. James H. Prestegard, and Dr. Kelley W. Moremen. They have all provided me with significant guidance (career and scientifically) throughout my graduate studies, and I am forever thankful for their support these many years.

I am thankful for the friendship and guidance of Dr. Renuka Kadirvelraj. She has supported me throughout my entire graduate career. I also thank Dr. Rob Phillips, Dr. Krishnadev Oruganty, and Dr. Steve Harvey for their help, enthusiasm, and support in putting together our *Nature* publication. In addition, I am grateful for the fruitful collaboration with Dr. Michael Gross and Elizabeth Bergman for their aid in our publication in *Nature*. I thank our collaborators Dr. Natarajan Kannan and Rahil Taujale for their contributions in my other first-author publication.

I thank my friend and colleague Nathan Beattie for his endless support. He has been one of the greatest teammates I have ever had. I am also thankful for the undergraduates (Weston McDonald and Tiffany Sirmans) who have paid dividends for me as a graduate student. Finally, I am extremely grateful for both Tony Szempruch and Joey LaMattina, as they voluntarily mentored me throughout my graduate studies and have inspired me on many occasions.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES xi

INTRODUCTION 1

CHAPTER

1 LITERATURE REVIEW 5

1.1 Protein allostery 5

1.2 The role of protein dynamics in allosteric regulation 10

1.3 Intrinsic disorder 16

1.4 The atypical allosteric mechanism of hUGDH 28

2 THE ENTROPIC FORCE GENERATED BY INTRINSICALLY DISORDERED SEGMENTS TUNES PROTEIN FUNCTION 33

2.1 Abstract 34

2.2 Main Text (Introduction, Results, and Discussion) 35

2.3 Materials and Methods 42

2.4 Supplementary Information 51

2.5 Acknowledgements 52

2.6 Main Text Figures 53

2.7 Supplementary Figures and Tables 59

3 THE CONSERVATION OF ATYPICAL ALLOSTERY IN *C. ELEGANS* UDP-GLUCOSE DEHYDROGENASE 73

3.1 Abstract 74

3.2 Introduction 75

3.3 Materials and Methods 76

3.4 Results 80

3.5 Discussion 87

3.6 Acknowledgements 89

3.7 Figures 90

3.8 Tables 99

3.9 Supplementary Figures and Tables 102

4 CONCLUSIONS 104

4.1 Broad Implications 104

4.2 Future Directions (Aim 1) 107

4.3 Future Directions (Aim 2) 110

REFERENCES 112

APPENDICES

A. HYSTERESIS IN HUMAN UDP-GLUCOSE DEHYDROGENASE IS DUE TO A RESTRAINED HEXAMERIC STRUCTURE THAT FAVORS FEEDBACK INHIBITION 143

A1 Abstract 144

A2 Introduction 145

A3 Materials and Methods 146

A4 Results 150
A5 Discussion 157
A6 Figures and Tables 162
A7 References 172
B. ALLOSTERY AND HYSTERESIS ARE COUPLED IN HUMAN UDP-GLUCOSE DEHYDROGENASE 176
B1 Abstract 177
B2 Introduction 178
B3 Materials and Methods 179
B4 Results 183
B5 Discussion 189
B6 Figures and Tables 193
B7 References 206
C. HYSTERESIS AND ALLOSTERY IN HUMAN UDP-GLUCOSE DEHYDROGENASE REQUIRE A FLEXIBLE PROTEIN CORE 212
C1 Abstract 213
C2 Introduction 214
C3 Materials and Methods 216
C4 Results 221
C5 Discussion 228
C6 Figures and Tables 231
C7 Supplementary Figures and Tables 243
C8 References 246

LIST OF TABLES

Page

Table S2.1: Data collection and refinement statistics 69

Table S2.2: Kinetic parameters of all UGDH constructs 70

Table S2.3: NAD^+ kinetic parameters for UGDH 72

Table 3.1: Data Collection and Refinement Statistics for UDP-Xyl bound cUGDH 99

Table 3.2: Kinetic parameters for cUGDH and hUGDH 100

Table 3.3: Global Analysis of UDP-xylose Inhibition 101

Table S3.1: UGDH phyla and associated Genbank/Uniprot ID's that were used in the phylogenetic analysis 103

Table A1: Data Collection and Refinement Statistics 168

Table A2: pH Dependence of Hysteresis 170

Table A3: Global Fit of Data from UDP-Xylose Inhibition Studies 171

Table B1: Data Collection and Refinement Statistics for $hUDGH_{A136M}$ 201

Table B2: Effect of UDP-Xyl on Hysteresis 203

Table B3: Steady State Parameters 204

Table B4: Global Analysis of Competitive Inhibition 205

Table C1: Data collection and refinement statistics 239

Table C2: Effect of UDP-Xyl on Hysteresis 240

Table C3: Steady State Parameters 241

Table C4: Global analysis of competitive inhibition 242

Table C.S1 Packing Defects and Associated Volumes.....244

LIST OF FIGURES

Page

Figure 1.1: Cooperativity with respect to ligand binding 6

Figure 1.2: The KNF model describes cooperativity in allosteric systems 8

Figure 1.3: The MWC model describes positive cooperativity in allosteric systems.....9

Figure 1.4: A hypothetical 3D representation of the free energy landscape of a protein 11

Figure 1.5: The conformational selection/population shift model for allosteric regulation 12

Figure 1.6: Allostery originates from changes in protein dynamics.....14

Figure 1.7: The dynamic free energy landscape of proteins is important for protein function and allosteric regulation.....15

Figure 1.8: Sequence and NMR solution structure of the dynamic protein SUMO-1.....17

Figure 1.9: The 'fly-casting' mechanism of IDRs.....19

Figure 1.10: The post-translational modification of an intrinsically disordered region alters protein function by remodeling the free energy landscape.....20

Figure 1.11: Fuzzy complexes and disorder-to-order transitions 21

Figure 1.12: Fusion partners and entropic bristles.....23

Figure 1.13: Increasing the length of a disordered loop results in an exponential decrease in protein stability 26

Figure 1.14: The entropic pulling force generated by intrinsically disordered segments.....27

Figure 1.15: The atypical allosteric inhibition mechanism of hUGDH.....29

Figure 1.16: The packing defects in the protein core of hUGDH enable the conformational change of the buried allosteric switch 30
Figure 2.1: The role of the ID-tail in allosteric inhibition of UGDH 53
Figure 2.2: Structural constraints of the ID-tail 55
Figure 2.3: The entropic force of the ID-tail alters the structure of UGDH 56
Figure 2.4: The ID-tail shapes the conformational landscape of UGDH 58
Figure S2.1: The crystal structures of UGDH(FL) and UGDH(ΔID) 59
Figure S2.2: Steady-state kinetic analysis of all UGDH constructs 60
Figure S2.3: Sedimentation velocity studies of the UGDH 61
Figure S2.4: The ID-tail is conserved in vertebrates 62
Figure S2.5: Exhaustive Monte Carlo simulations constraining the ID-tail 63
Figure S2.6: The ID-tail induces global changes in the structure and dynamics of UGDH 65
Figure S2.7: Transient-state analysis of UGDH(FL) and UGDH(ΔID) 67
Figure 3.1: The allosteric transition in hUGDH 90
Figure 3.2: Structure-based sequence alignment 91
Figure 3.3: The Inhibited E^{Ω} State of cUGDH 92
Figure 3.4: UDP-Xyl binding induces formation of the E^{Ω} state 93
Figure 3.5: Structural divergence in the cUGDH allosteric switch 94
Figure 3.6: The packing defect residues (PD residues) are conserved in cUGDH 95
Figure 3.7: Kinetic analyses of cUGDH and hUGDH 96
Figure 3.8: Phylogenetic analysis of UGDH 97
Figure 3.9: Representative sequences from each phyla in the evolutionary analysis of UGDH 98
Figure S3.1: The hexamer building interface residues of cUGDH and hUGDH 102

Figure 4.1: The evolutionary mechanisms for re-modeling the energy landscape and function of proteins 106
Figure A1: Allostery and hysteresis in hUGDH 162
Figure A2: Hexamer stability and activity are pH sensitive 164
Figure A3: UDP-xylose inhibition is sensitive to pH 165
Figure A4: Crystal structure of the E* complex 166
Figure A5: Model for the activation and UDP-Xyl inhibition of the E* hexamer 167
Figure B1: hUDGH undergoes conformational changes during allostery and hysteresis 193
Figure B2: Rationale for the design of the A136M substitution 195
Figure B3: Crystal structure of $hUGDH_{A136M}$ in the open domain 196
Figure B4: A136M substitution that stabilizes the E state of hUGDH 197
Figure B5: A136M substitution that stabilizes the hexamer 198
Figure B6: A136M substitution that disrupts hysteresis and cooperativity 199
Figure C1: Allostery and hysteresis in hUGDH 231
Figure C2: RosettaVIP design of the A10L cavity filling substitution 232
Figure C3: E* state of $hUGDH_{A104L}$ 233
Figure C4: The A104L substitution decreases the flexibility of cavity C-I and surrounding residues 235
Figure C5: Kinetic analysis of native hUGDH and $hUGDH_{A104L}$ 236
Figure C6: Proposed mechanism for allostery and hysteresis in hUGDH 238
Figure C.S1: When substrate is bound $hUGDH_{A104L}$ adopts the E state 243

INTRODUCTION

Proteins are essential macromolecules and are key to all cellular activities. Proteins catalyze chemical reactions, perform mechanical work, and provide structural stability to cells.[1,2] To obtain these various functions, proteins fold into specific three-dimensional structures that depends on their precise sequence of amino acids.[1-7] This sequence-structure-function dogma was first recognized by Charles Anfinsen and has been a central tenet of structural biology for decades.[3] However, recent work has shown that the proteome contains functional proteins and polypeptide segments that do not fold into a well-defined structure.[8-31] These intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) challenge the structure-function paradigm.[10-12,22,23] In fact, it is estimated that ~44% of the human proteome contains intrinsically disordered segments that are greater than 30 amino acids in length.[32] The majority of these IDRs have not been functionally characterized, and why they are so persistent across the proteomes of all organisms is a mystifying phenomenon.[22,33-35]

Intrinsically disordered segments in proteins are defined by low sequence complexity, and they can be accurately identified from just the pattern of amino acids in any given protein sequence.[36-42] In general, IDRs are composed of amino acid stretches that are deficient in hydrophobic residues and often have a high net charge.[21,36,39,43-45] These features disable the ability of many IDRs to undergo a hydrophobic collapse to a folded state, and as a result, disordered segments are almost always found in solvent-exposed regions of proteins.[21-23,43,45-49] Because of

the increase in solvent exposure, IDRs exhibit much larger conformational flexibility compared to the well-packed core of a protein structure.[8,22,23,33,35,42,44-46,49-54]

Both IDRs and the structured portions of proteins are subjected to thermal fluctuations (k_BT).[55-62] Accordingly, protein structures are dynamic and contain segments that display both high and low flexibility.[22,23,33,46,55,56,59,63-69] The dynamic, 'native' folded-state of proteins can adopt numerous conformations that are important for function (e.g.; binding interactions, catalysis, and regulation).[55-57,62,64,66,70-75] The ensemble of structural sub-states that a given protein can adopt are in dynamic equilibrium.[55,56,72,74] Mutations in the sequence of a protein are expected to perturb the distribution of sub-states, which often results in functional changes.[68,71,74-79] Therefore, the conformational landscape is under selective pressure. How evolution tunes the ensemble of structural sub-states to optimize specific protein functions is poorly understood.[68,74,75]

The primary project presented in this dissertation originated from a serendipitous observation that the intrinsically disordered C-terminus (ID-tail) of human UDP-α-D-glucose-6-dehydrogenase (hUGDH) had a profound effect on the affinity for an allosteric inhibitor. Because many IDRs have no annotated function[22,33,36], a typical routine is to remove them from proteins for the sake of facilitating structural studies. In our lab we had two hUGDH constructs: the full-length enzyme (hUGDH-FL) and one without the 30-residue ID-tail (hUGDH-ΔID). A previous study[80] had shown that the ID-tail had no significant effect on the catalytic activity (k_{cat}) or the K_M for the substrate UDP-Glucose, which was consistent with our own work.[81] However, we found that the ID-tail significantly enhanced (>10-fold) the enzyme's affinity for the allosteric inhibitor UDP-Xylose.[81] Therefore, the goal of this project was to explore <u>how</u> the ID-tail increased the affinity for UDP-Xylose. This long venture ultimately brought us to a paradigm shifting hypothesis about protein evolution. In a publication in *Nature*, we proposed that the entropic force generated

by intrinsically disordered segments can tune protein function by re-shaping the conformational landscape.[81] This simple mechanism has no specific sequence or structural requirements, which may explain the persistence of IDRs in the human proteome. In addition, this work highlights the utility of disorder in the evolution of proteins.[81]

Chapter 1 presents a review of the literature, focusing first on protein allostery and then on the role of protein dynamics in allosteric regulation. The chapter transitions to a broad overview of intrinsic disorder in proteins, with an emphasis on the known functional mechanisms of IDRs. Specifically, this section will cover the importance of IDRs in protein stability, function, and regulation. The chapter ends on a brief overview of the atypical allosteric mechanism in hUGDH, which is relevant to both Chapters 2 and 3.

In Chapter 2, the mechanism of the ID-tail in the allosteric regulation of hUGDH is covered. This work was published in the journal *Nature* (DOI: 10.1038/s41586-018-0699-5) and presents a novel view on how proteins might have evolved to optimize specific functions. In Chapter 3, a strong case for the conservation of the allosteric mechanism in *C. elegans* UGDH (cUGDH) is presented, which is currently in peer review with the journal *ACS Omega* (Reprinted here with the permission from the managing editor at ACS; Dr. Dinesh Soares). Overall, this publication advanced our understanding of the allosteric mechanism in UGDH. In addition, this work expands a sequence motif that may be useful in the prediction of atypical allostery in divergent UGDHs. We tracked the evolution of the newly defined atypical allostery motif by collaborating with Dr. Natarajan Kannan and Rahil Taujale. Their phylogenetic analysis combined with our structural analyses lead to the identification of a potential transition point in the evolution of allostery in UGDH.

Chapter 4 revisits the implications of the discoveries presented within the dissertation. In addition, future ideas for projects that would build on this work are also detailed. Co-authorships (3 total) that are relevant to the study of the atypical allosteric mechanism in hUGDH can be found in the appendices (*A*-*C*). The other co-authorships (3 total) that were not directly relevant to my thesis work but were obtained through collaboration on other projects during my graduate studies are not included in the appendices. Instead, below are the provided citations to these works, with DOI links provided for quick access:

> *1)* Bester, S.M., Adipietro, K.A., Funk, V.L., Myslinski, J., **Keul, N. D.**, Cheung, J., Wilder, P.T., Wood, Z.A., Weber, D.J., Height, J.J. and Pegan, S.D., The structural and biochemical impact of monomerizing human acetylcholinesterase. *Protein Science* (2019). DOI: 10.1002/pro.3625
> *2)* LaMattina, J.W., Delrossi, M., Uy, K.G., **Keul, N.D.**, Nix, D.B., Neelam, A.R. and Lanzilotta, W.N., Anaerobic Heme Degradation: ChuY Is an Anaerobilin Reductase That Exhibits Kinetic Cooperativity. *Biochemistry*, 56(6), pp.845-855 (2017). DOI: 10.1021/acs.biochem.6b01099
> *3)* LaMattina, J.W., **Keul, N.D.**, Reitzer, P., Kapoor, S., Galzerani, F., Koch, D.J., Gouvea, I.E. and Lanzilotta, W.N., 1, 2-propanediol dehydration in Roseburia inulinivorans; structural basis for substrate and enantiomer selectivity. *Journal of Biological Chemistry*, pp.jbc-M116 (2016). DOI: 10.1074/jbc.M116.721142

My contribution to the above works (*1-3*) was conducting and analyzing sedimentation velocity experiments on the proteins of interest.

CHAPTER 1

LITERATURE REVIEW

1.1 Protein Allostery

Proteins fold into specific three-dimensional structures that are essential to their biological function.[2-7] The function of many proteins is regulated by the binding of an effector molecule to an allosteric site.[82] If an effector molecule binds and enhances a specific protein function, then it is called an allosteric activator. Conversely, if an effector molecule binds and diminishes a specific protein function, then it is referred to as an allosteric inhibitor.[83,84] The allosteric regulation (activation or inhibition) of proteins provides organisms with the capacity to rapidly adapt to changes in the environment.[82-85] Therefore, the biological significance of allosteric regulation cannot be overstated, and the process has been historically described as *"the second secret of life"*.[86]

When a ligand binds to one subunit of a protein and increases the affinity for ligand binding in the another subunit, the system is described as positively cooperative (Figure 1.1A,B).[1] Positive cooperativity was first observed in 1904 by Christian Bohr, where he noticed the non-hyperbolic, sigmoidal binding of oxygen to the protein hemoglobin. His work showed a sigmoidal-shaped substrate saturation curve with oxygen binding to hemoglobin, which depended on both the pH and the concentration of carbon dioxide in the solution (appropriately named, the Bohr effect).[87] In the years following Bohr's observation, many other enzymes were found to show both positive and negative cooperativity (Figure 1.1A-C).[83,84,88-94] The observation of cooperative binding in

proteins suggested that their structures are composed of subunits with interconnected properties.[88,95] Thus, when the first crystal structures of oxy- and deoxy-hemoglobin were solved by Perutz[7,96], conceptual models were proposed to explain cooperative binding, and the term 'allostery' was born.[82,83,86,89,90]

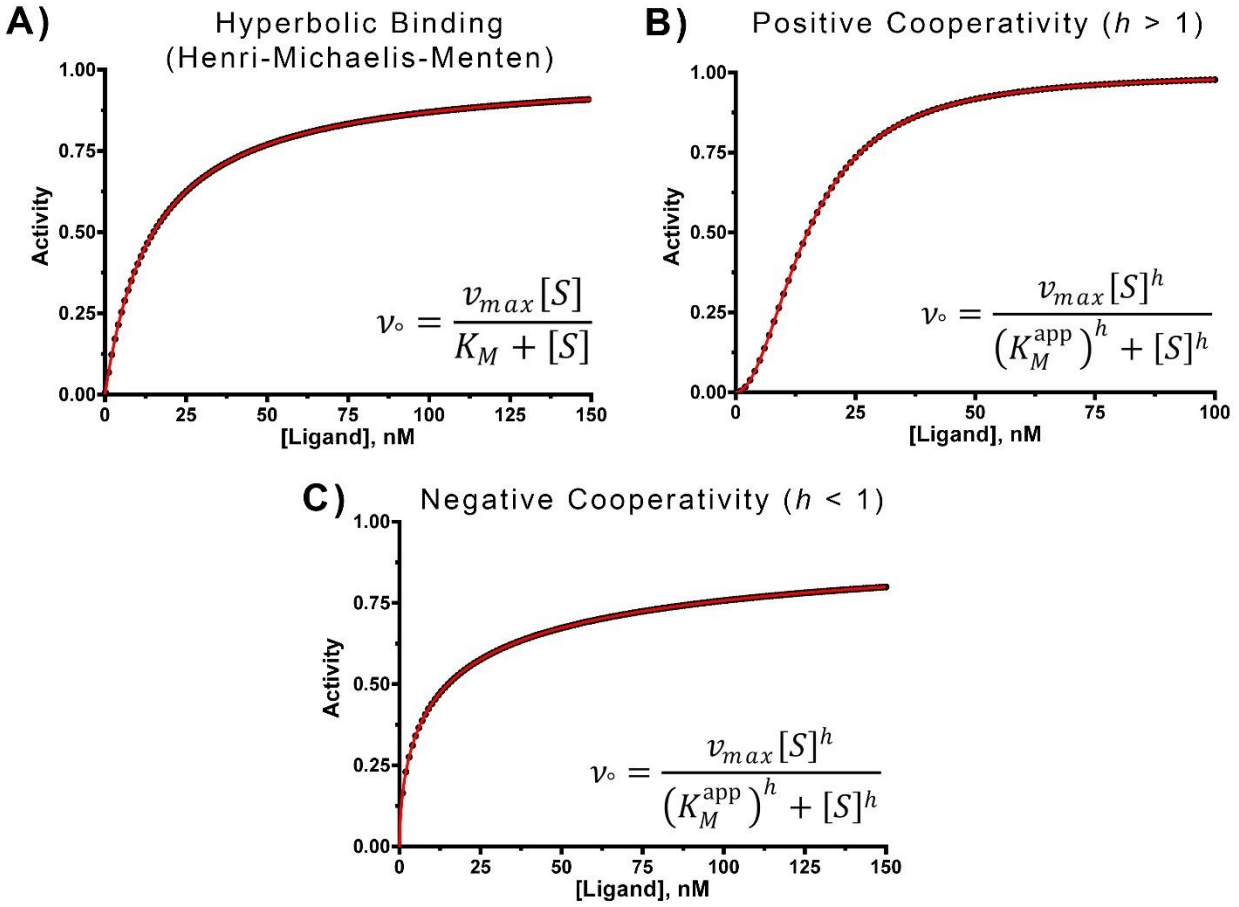


Figure 1.1: Cooperativity with respect to ligand binding. (A) Simulated data fit to the Henri-Michaelis–Menten model[97,98], which describes hyperbolic ligand binding of a substrate to an enzyme (E + S ⇌ ES → EP). **(B)** Simulated data fit to the Henri- Michaelis–Menten modified with the hill-coefficient (h)[99,100] to describe positive cooperativity (h > 1) with respect to ligand binding. The sigmoidicity of the curve increases as h increases. In this simulation: h = 2.0. **(C)** Same as in (B), except h = 0.6, which is reflected as negative cooperativity (h < 1) in the binding curve. In negatively cooperative systems, the asymptote (v_{max}) of the enzyme is more difficult to determine, as the enzyme cannot be fully saturated with ligand. All curves and data were simulated in GraphPad PRISM 6.0. Identical kinetic parameters were defined for each data set (K_M = 15; v_{max} = 1.0), except for panels B and C, which have different hill-coefficients as stated above.

The two classical models of allostery were proposed by Monod, Wyman, and Changeux (MWC)[84], and Koshland, Nemethy, and Filmer (KNF)[101].[89] The KNF (sequential or induced fit) model postulated that the unliganded structure of proteins are not optimized for ligand binding.[101,102] Instead, initially weak interactions with ligand produce a conformational change in the protein structure, which strengthens the protein:ligand interactions (Figure 1.2). In a protein made up of multiple subunits, the conformational change in one subunit produces a structural and functional change in an adjacent subunit (i.e.; cooperativity, see Figure 1.1 for examples). At its essence, this inter-subunit communication (cooperativity/allostery) originates from ligand-induced changes being transmitted through the network of interactions formed between protein subunits (Figure 1.2).[88,94,101,102] To better describe how the induced-fit model describes cooperativity, let us assume that there is an enzyme (E) that is only stable and active when four identical subunits associate to form a tetramer ($E_{Subunit\ \underline{1}} + E_{Subunit\ \underline{2}} + E_{Subunit\ \underline{3}} + E_{Subunit\ \underline{4}} \rightleftharpoons E_{Tetramer}$). In the unliganded $E_{Tetramer}$, all the subunits are in an identical conformation and initially have the same weak affinity for ligand. When one subunit binds ligand, it undergoes a conformational change that influences both the shape of the adjacent subunits and their affinity for the ligand (Figure 1.2). Therefore, as ligand concentration is increased, the affinity for ligand in other subunits is sequentially changed, which results in cooperative binding (positive or negative) (Figures 1.1A-C and 1.2).[101] In contrast to the KNF model, the MWC model hypothesizes that the subunits of a protein pre-exist in distinct conformations that have interconnected properties (e.g.; ligand binding affinity) (Figure 1.3).[84] To distinguish the MWC model, we again use the $E_{Tetramer}$ as an example (Figure 1.3). In the most simplistic case, each subunit of the unliganded $E_{Tetramer}$ can adopt one of two states (T or R). The subunits in the 'T state' have a low binding affinity for the ligand, whereas those in the 'R state' have a higher affinity for the ligand. In the $E_{Tetramer}$, all the subunits exist in a dynamic equilibrium

of T and R states. However, in the unliganded $E_{Tetramer}$, the equilibrium initially favors either the T or R state (Figure 1.3). When ligand binds to a subunit in either state, the other subunits of $E_{Tetramer}$ undergo a concerted conformational change to the high affinity, R state. In other words, ligand binding causes structural perturbations, which then increases the relative population of subunits that adopt the high affinity, R state. Therefore, in the MWC model, positive cooperativity originates from a concerted change to the equilibrium structural substates upon ligand binding (Figures 1.3).[84,92]

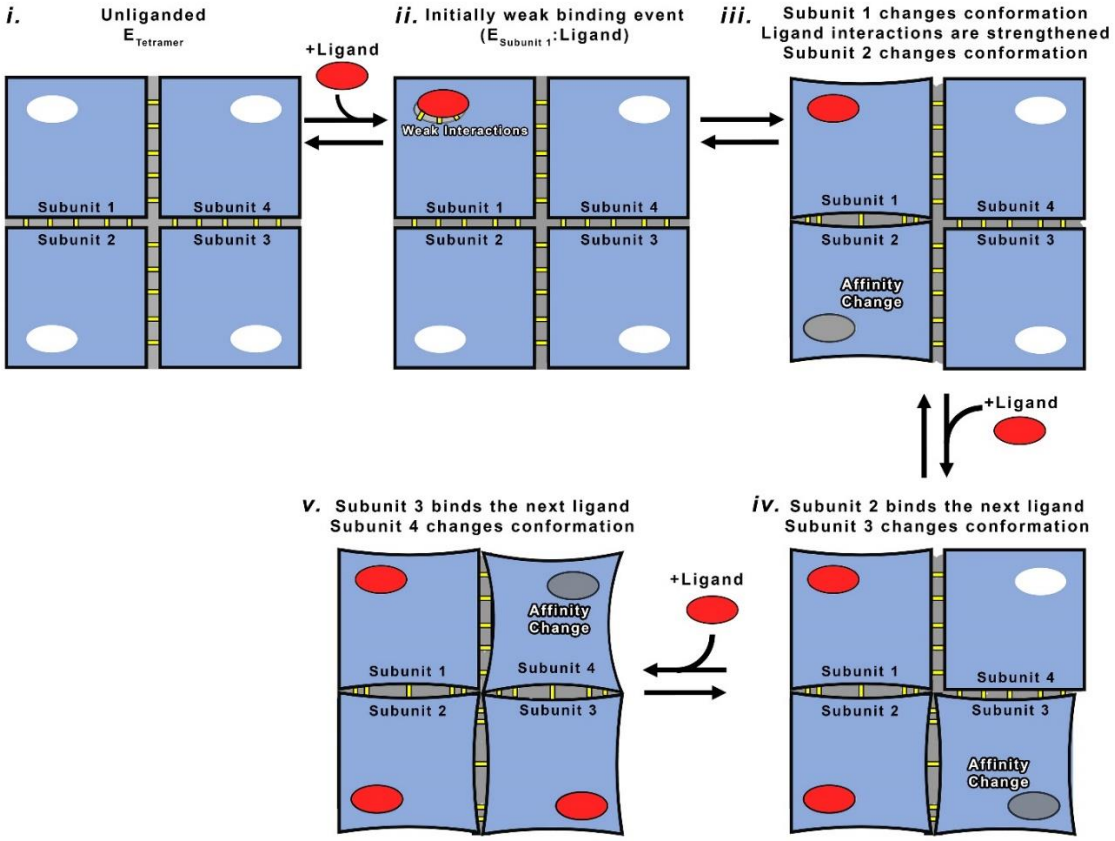


Figure 1.2: The KNF (sequential) model[94,101,102] describes cooperativity in allosteric systems. ***i)*** In the absence of ligand, the example enzyme (E) forms a tetramer ($E_{Tetramer}$) via the association of four subunits (blue squares). In the unliganded state, all the binding sites (white) in each subunit of $E_{Tetramer}$ have equal affinity for the ligand (red oval). ***ii)*** The interactions (van der Waals [grey shading] and hydrogen bonds [yellow]) between Subunit 1 and the ligand are initially weak. ***iii)*** These interactions are strengthened by a conformational change in Subunit 1. This conformational

change also alters the interactions between Subunit 1 and Subunit 2, and as a result, Subunit 2 undergoes a conformational change to a state that has a different affinity for the ligand. ***iv)*** The next ligand binds to the altered conformation of Subunit 2, which induces Subunit 3 to undergo a conformational change. In like manner, this changes the binding affinity for the ligand in Subunit 3. ***v)*** The subsequent ligand binding event with Subunit 3 produces a conformational change in Subunit 4, modifying its binding affinity for the ligand. Thus, in this model, ligand-induced conformational changes in one subunit propagates a change to the structure and function of an adjacent subunit. This leads to a sequential change in the binding affinity for the ligand at other subunits, which results in either positive or negative cooperativity (see Figure 1.1 for examples). In fact, the KNF model has historically been used to explain the rare phenomena of negative cooperativity observed in some allosteric systems.[88,91,103,104]

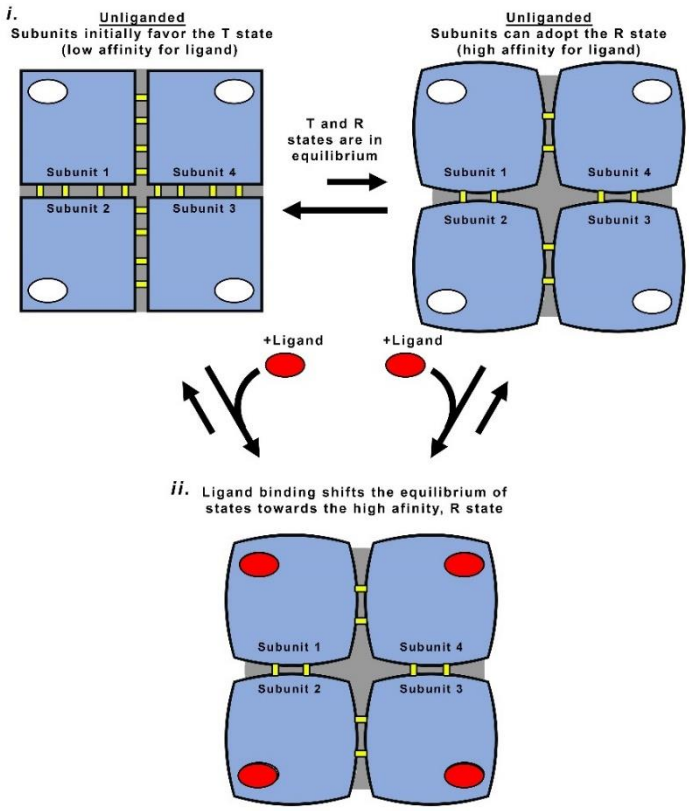


Figure 1.3: The MWC (concerted) model[82-86,89,90,105,106] describes positive cooperativity in allosteric systems. ***i)*** The subunits of unliganded $E_{Tetramer}$ are in a dynamic equilibrium of states (T and R), with the T state having a low affinity for ligand binding and the R state having a high affinity for the ligand. ***ii)*** The binding of ligand to either state shifts the equilibrium towards the high affinity, R state. In other words, as ligand concentration is increased the population of subunits that adopt the high affinity, R state also increases. Importantly, this model describes positive cooperativity in allosteric systems (for an example of positive cooperativity see Figure 1.1B).

Both the KNF and the original MWC[84] models have been successfully deployed to describe cooperativity in many allosteric systems.[74,88-91,93,94,103-105,107-113] Despite different mechanisms, both models imply that cooperativity and allostery stem from an effector-induced structural change in the protein.[89,105,109] These structure-based views of allostery were conceptual and did not initially provide insight into either the thermodynamic or mechanistic basis of allostery.[84,90,101,109]

1.2 The Role of Protein Dynamics in Allosteric Regulation

"...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

– Richard P. Feynman (*Feynman Lectures on Physics*, vol. 1, Ch. 3. 1963)

The MWC model[106] laid the groundwork for a thermodynamic explanation of allostery, which is referred to as the conformational selection model.[74,83,88-90,108,113,114] This view maintains the original idea that allostery originates from an effector molecule selectively binding to a specific structural substate, which shifts the equilibrium of conformers (Figure 1.3).[74,88,89,115] However, the modernized conformational selection model incorporates the energy landscape theory of proteins, which states that the native 'folded' structure of a protein can adopt a wide range of structural substates that have different relative free energies (Figure 1.4).[55-58,62,72,74,116-120] In this model, allostery is interpreted as a distribution of states that can be described by statistical thermodynamics.[56,74]

In solution, a protein structure fluctuates between high and low energy conformations. The exchange rate between different conformers is determined by the height of the energetic barriers that separate them (Figure 1.4).[74] Within this free energy landscape, the lowest energy conformations are sampled the most frequently. In other words, the distribution of states is

dominated by the conformers that have lowest relative free energies (Figure 1.4).[74,115,121] However, allosteric regulation has been hypothesized to originate from the binding of an effector molecule to one of the weakly populated, high energy conformers.[74] In such an event, the equilibrium between structural substates shifts towards the higher energy conformation. This is often referred to as 'population shift' (Figure 1.5A).[72,74,116-119,122,123] In reality, the binding of an effector molecule to any structural substate will reshape the energy landscape of a protein.[74,109,124] Thus, supported by numerous studies[64,74,88-90,115,125-138], the conformational selection/population shift model has become the dominant view of how proteins are allosterically regulated (Figure 1.5A).[76,139]

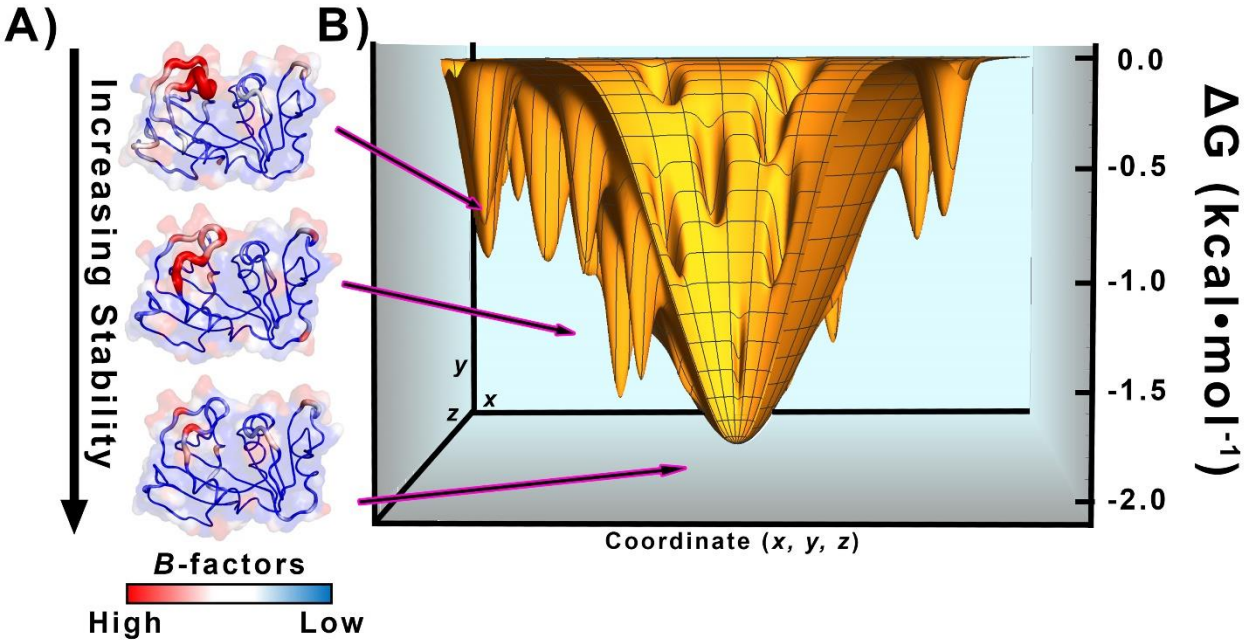


Figure 1.4: A hypothetical 3D representation of the free energy landscape of a protein. (A) Protein structures are generally dynamic and contain regions of both high and low flexibility.[29,55,56,62,65,66,70,75] The Debyre-Waller factor[140,141] (also referred to as either the temperature factor or *B*-factor) is a traditional metric used for assessing the relative fluctuation of atoms in the crystal structure of a protein. The *B*-factor metric for protein flexibility assumes that the position of every atom in a crystal structure fluctuates around an equilibrium position (i.e.; atomic positions are best represented as a Gaussian distribution). Atoms with high *B*-factors exhibit large oscillations around an average position, whereas atoms with low relative *B*-factors do not.[142-149] Protein flexibility and stability are generally assumed to be inversely correlated.[58,142,143,145,147,150] To better describe this relationship, three different isomorphic structures of the *E. coli* protein dihydrofolate reductase (ecDHFR) are used as representative

models (PDBs: 1RX1, 1RX5, and 1RX6).[64] The atoms in each structural state are colored on a *B*-factor scale. Regions of the proteins that display high flexibility (high *B*-factors) are colored red, whereas more rigid regions (low *B*-factors) are colored blue. Regions colored white are those with intermediate *B*-factor values. Moving down the stack of the different ecDHFR conformations, the overall flexibility of the protein decreases, and as a result, we assume that stability increases. Therefore, proteins are dynamic and are best represented as a structural ensemble where each substate has a different relative free energy.[56] **(B)** The conformations of ecDHFR that are important for function[64] (from panel A) are presented in a theoretical free energy landscape. Here, the *y*-axis defines the relative free energy of a specific conformation. This depends on the overall fluctuations of the atoms that make up the protein structure (i.e.; dynamics), which is is loosely defined in the free energy landscape by the coordinates *x* and *z*.[56] The most stable conformation is located at the bottom of the funnel in the free energy landscape. The conformations that have less stability (i.e.; more flexibility) are positioned higher up in the funnel.[56,64,117,151] All the possible, stable conformations occupy 'local minima' within the funnel, which are separated by energetic barriers.[56,64,65,67,72,74,117,118,151,152]

The allosteric remodeling of a protein structure traditionally invokes an observable conformational change.[84,89,109] Therefore, the original assumption behind the conformational selection/population shift model states that the essence of allostery derives from primarily changes in enthalpy (i.e.; making and breaking of bonds).[74,79,88,115,153-159] However, recent studies have

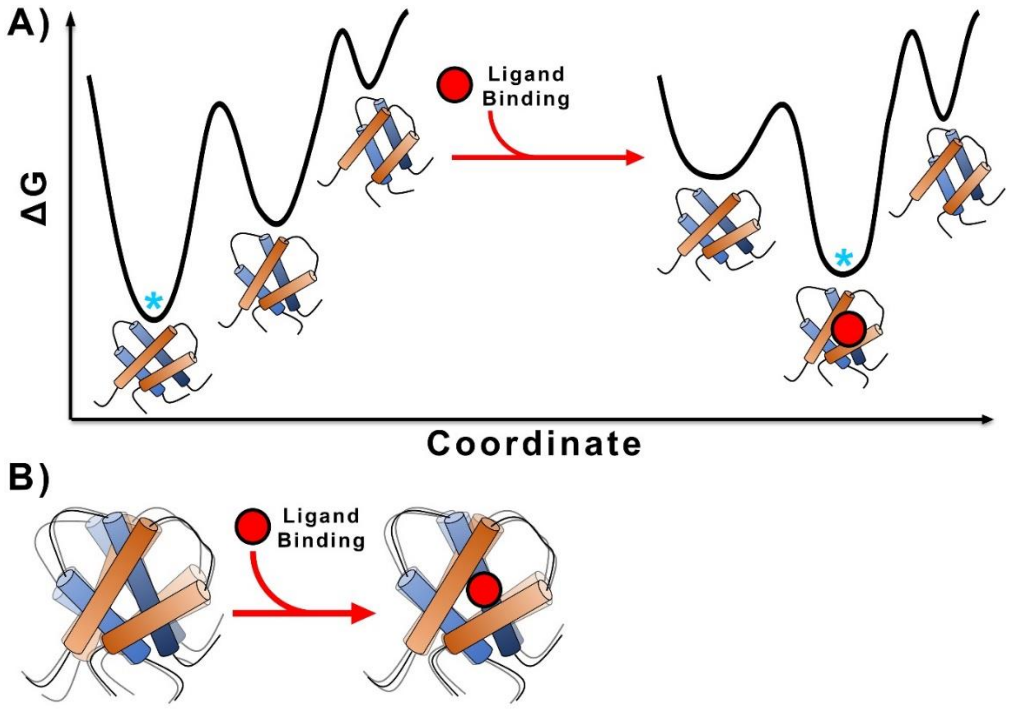


Figure 1.5: The conformational selection/population shift model for allosteric regulation. (A) A 2-dimensional free energy landscape of a protein with two identical subunits (colored blue and orange) is depicted. In the unliganded state (*left*), the protein samples three distinct conformations that have different relative free energies (indicated by the depth of the wells). These conformations are in dynamic equilibrium, but the relative distribution is dominated by the most stable conformer (indicated with a cyan star). Upon ligand (red circle) binding (*right*), the relative distribution of states is shifted to favor ligand binding. **(B)** There is accumulating evidence that allostery can occur in the absence of an observable conformational change.[55,67,79,109,115,153,155,160,161] Here, the binding of a ligand simply alters the average vibrational mode of the protein structure. The protein in the unliganded state (*left*) has greater conformational entropy than in the ligand bound state (*right*). Therefore, the absence of an observable structural change is likely due to ligand binding inducing small changes in the average fluctuations of the protein structure.

shown that allostery can occur in the absence of an observable conformational change (Figure 1.5B).[73,74,79,115,124,156,160,162,163] These studies highlight the importance of conformational entropy in allosteric mechanisms, which was first recognized by Cooper and Dryden in 1984.[153] In their early thermodynamic description of allostery, the authors suggested that the binding of an effector molecule could simply alter the average vibrational motion of a protein (Figure 1.5B).[153] Importantly, this thermodynamic model fuses into the conformational selection/population shift model, as it only emphasizes the importance of the entropic component in allosteric regulation (Figure 1.5).[88,109,115,153,154]

The thermodynamic views of allostery presented thus far do not describe how allosteric signals propagate through the structure of a protein at an atomistic level of detail. Nonetheless, the advancement of techniques such as nuclear magnetic resonance (NMR) and molecular dynamics (MD) has begun to reveal the importance of the *'jigglings and wigglings of atoms'* in the mechanisms of allosteric regulation.[55,56,65,70,88,109,115,156,162-164] Driven by thermal fluctuations in solution, the peptide backbone and side chain atoms of a protein vibrate (referred to as 'internal dynamics').[56,62,149] Recent evidence suggests that allostery originates from effector-induced changes to the internal dynamics of a protein (Figure 1.6).[55,70,72,79,88,116,155,156,160] Therefore, in

allosteric systems that do not exhibit a distinct structural change, effector-induced alterations to the internal dynamics of a protein might facilitate inter-subunit communication.[55,109,151] This model is referred to as 'dynamic allostery' (Figure 1.6).[55,74,109,151,159]

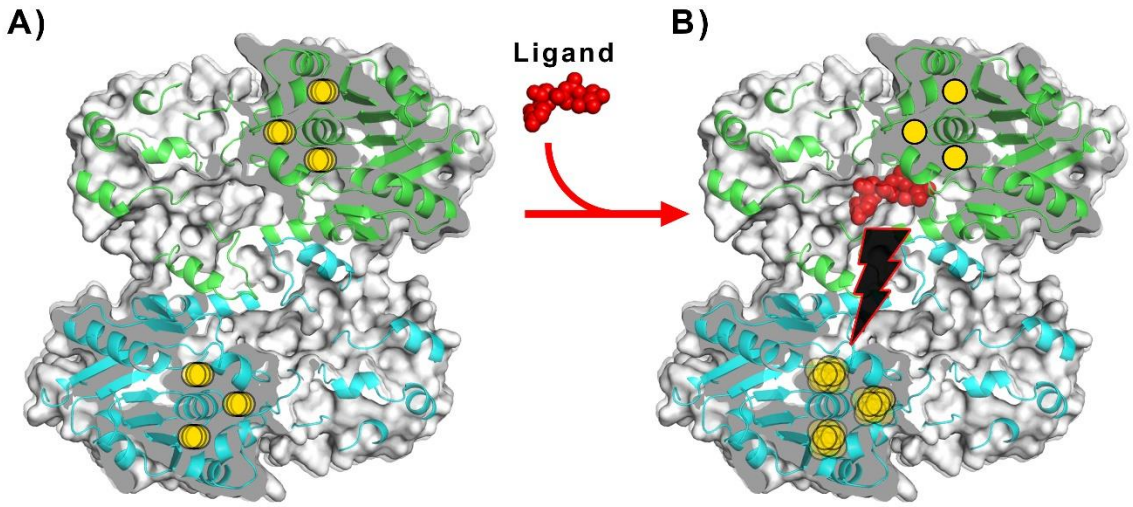


Figure 1.6: Allostery originates from changes in protein dynamics.[67,109,115,151,154,155,158,161,165] **(A)** In the unliganded state, an allosteric dimer protein forms from the association of two identical subunits (green and cyan) that are related by a two-fold rotation axis. The entire surface of the dimer is colored white, and areas where the surface is cut-away to depict the interior of the protein are colored grey. The yellow circles represent residues that are essential to the allosteric communication between subunits. The degree of shading for each circle depicts the relative vibrational mode of the atoms in that position. When the protein is unliganded, the atomic fluctuation of the residues in both subunits is nearly identical. **(B)** When the ligand binds to one subunit (green), the dynamics of the residues that are important for allosteric regulation are quenched. This produces a change in a network of interactions (not depicted), which transmits an allosteric signal (black lightning bolt) to the other subunit (cyan). As a result, the internal dynamics in the other subunit (cyan) are increased. This is depicted by an increase in the average position of those residues that are important for allosteric regulation (yellow circles in cyan subunit). In this dynamic mechanism of allostery, the protein does not require a distinct conformational change. Importantly, the increase or decrease in the internal dynamics of the adjacent subunit can lead to either an increase or decrease in its affinity for the ligand, respectively (i.e.; positive or negative cooperativity).[79,88]

As highlighted by the Nobel laureate Martin Karplus, the term 'dynamics' is often invoked and mis-used in thermodynamic descriptions of allostery.[65,88] By definition, dynamics is the change in motion over a defined time-scale. Both the 'slow' and 'fast' time-scale dynamics of

proteins are expected to play a role in the mechanism of allosteric regulation (Figure 1.7).[55,65,79,88,109,115,154-156,161] In fact, there is evidence that changes in the fast (pico-to-nanosecond) internal dynamics of a protein can influence its slow (micro-to-millisecond) dynamics (Figure 1.7).[55,79,88,109,115,154,160,161] To understand allostery at an atomistic level of detail, future work will need to address the interplay between the fast and slow dynamics.[55,88] Advancements in techniques such as NMR and molecular dynamics will likely provide mechanistic insight into allosteric regulation, as these techniques can examine the motions of atoms across a wide range of time-scales (pico-to-millisecond) (Figure 1.7).[55,79,88,109,155,156,161]

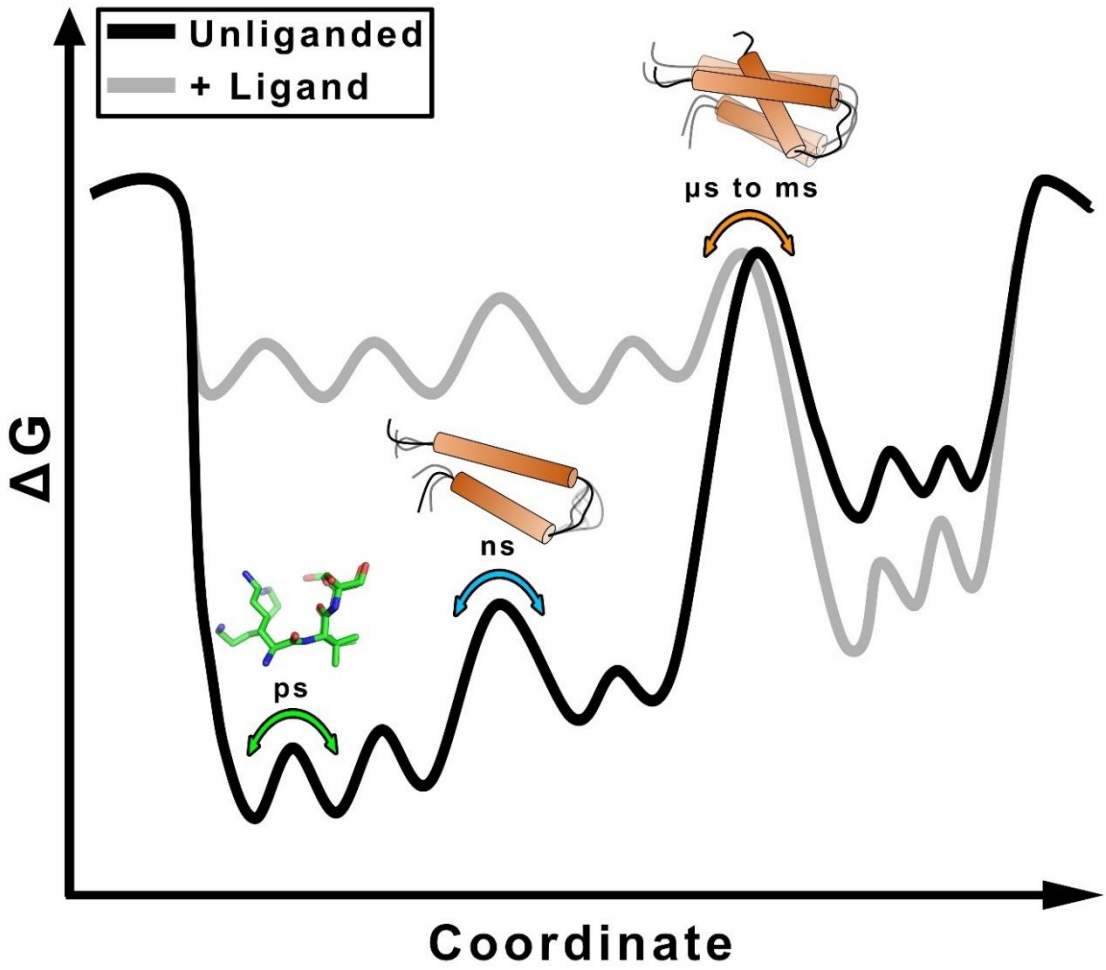


Figure 1.7: The dynamic free energy landscape of proteins is important for protein function and allosteric regulation.[55,56,65,88,109,115,154,166] Protein dynamics can be broken down into a

hierarchy of timescales, that are interrelated and important for allosteric regulation.[55,65,115,154] Depicted is a 2D free energy landscape for an unliganded protein (black wavy line), with the example timescales of dynamics (the peaks are energetic barriers) that relate the different states of a protein (wells) along a conformational coordinate. Bond vibrations (not shown) occur on the femtosecond timescale. Methyl-rotations and side-chain rotameric shifts occur on the picosecond timescale (green arrows), whereas loop motions often are detected within the nanosecond range (cyan arrows). Large protein motions (e.g.; domain motion) occur in the micro-to-millisecond timescale (orange arrows), and for some proteins, conformational changes occur in the time range of seconds (not shown). Upon binding ligand (grey wavy line) the amplitude and frequency of the proteins motion are perturbed, which reshapes the conformational landscape. Thus, both thermodynamics (e.g.; the relative free energy of conformers) and kinetics (e.g.; the rates of exchanges between conformers) are important in understanding allostery. This figure was inspired by refence[55] and completely re-created with additional detail.

In summary, allosteric regulation is manifested by effector-induced changes to the thermodynamic and kinetic (dynamic) properties of a protein. Therefore, the energy landscape theory of proteins provides a foundation for interpreting allosteric regulation. In fact, there is significant evidence that mutations to the sequence of a protein can produce or modify an allosteric function via changes to free energy landscape.[55,71,74,75,109,160,167,168] This suggests that both the stability of the folded 'native' structure and the conformational fluctuations it exhibits are under selective pressure.[68,71,72,74-79,115-118,124,131,154,162,167-173] The dynamic property of proteins likely facilitated the evolution of allosteric functions.[74,109,115,124,131,154,158,159,161,162] To highlight the important link between protein dynamics and allostery, recent work has shown that allostery is prevalent in proteins that lack a stable, folded structure.[12,15,19,20,23,24,33,155,174-179] These intrinsically disordered proteins (IDPs) display extreme conformational heterogeneity, which is important for rapidly binding multiple partners and relaying allosteric signals.[9,12,15,20,24,45,55,74,124,180]

1.3 Intrinsic Disorder

Across the proteome of all species, a large number of proteins and protein segments are predicted to be intrinsically disordered.[22,35] In fact, it is estimated that ~44% of the human proteome contains intrinsically disordered regions (IDRs) that are longer than 30 residues in

sequence.[32] These IDRs are usually devoid in bulky hydrophobic residues and often have a high net charge (Figure 1.8A).[9,13,21,36-41,43-45,181] Thus, the sequence composition inherent to IDPs and IDRs disables their ability to fold into a stable structure, and as a result, they exhibit a large degree of conformational heterogeneity compared to the well-packed core of structured proteins (Figure 1.8B).[8,12,21-23,36,44,46,48,49] The characteristic conformational heterogeneity of IDPs and IDRs presents experimental challenges for their functional characterization.[8,12,13,20,22-24]

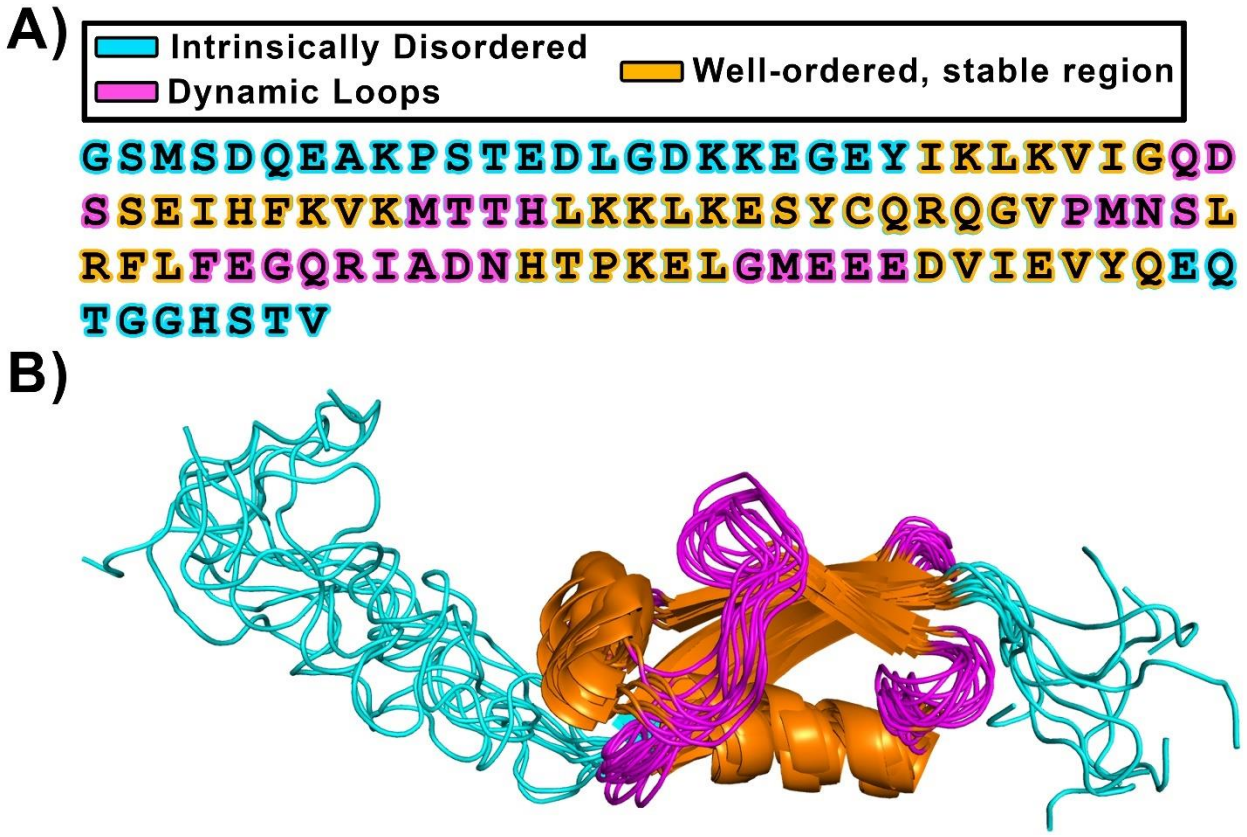


Figure 1.8: Sequence and NMR solution structure of the dynamic protein SUMO-1. (A) The N- to C-terminal sequence of SUMO-1 with residues colored by their relative degree of predicted disorder.[38] Completely disordered regions (N- and C-terminal tails) are colored cyan, dynamic internal loop regions are colored magenta, and well-ordered regions are colored orange. **(B)** The NMR solution structure of SUMO-1 (PDB: 1A5R) is depicted in several resolved states. The breadth of conformational heterogeneity exhibited by different regions in the protein is shown with coloring that is the same as in panel A. Of note, is the dynamic nature of the termini (cyan) and the internal loops (magenta), which contrasts with the average motion of the structured regions (orange).

The majority of IDRs have no known functional annotation.[22] Traditionally, IDRs have been regarded to serve the simple function of linking together secondary structural elements that make up the tertiary fold of a protein structure.[22,33] However, recent work has revealed that some IDRs have specific functions.[22,33,182] The known functions of IDRs can be broken down into two basic categories: *A)* IDRs whose function originates through an interaction or post-translational modification; *B)* IDRs whose function does not involve any interaction or modification (i.e.; entropic-based functions).[22,33] The following sections will discuss the various mechanisms associated with both of these functional categories:

A) IDRs whose function originates through an interaction or post-translational modification

Compared to the classical binding of two structured elements, there are reduced steric constraints associated with the binding of a macromolecule to a solvent-accessible disordered segment. In fact, many IDRs contain short stretches in sequence that act as recognition motifs, which facilitate the weak, transient binding to other macromolecules.[13,22,47,177,183] By positioning a recognition motif in the flexible, solvent-accessible disordered region of a protein, the rate of convergence between binding partners can be increased.[17,184] This is formally known as the 'fly-casting' mechanism[185] and has been shown to enhance the kinetics of molecular recognition events (Figure 1.9).[13,17,19,124,183,186-191] In the fly-casting mechanism, the weak, transient interactions formed between an IDR and a binding partner enables them to rapidly associate/disassociate (Figure 1.9).[17,124,185,192] This fly-casting mechanism is useful and prevalent in proteins that serve as "hubs" for the assembly of larger macromolecular complexes.[12,13,22-24,50,51,177,179,182,183,193-196] Intrinsically disordered segments also provide hub proteins with the capacity to bind a wide range of other macromolecules.[10,24,51,177,179,182,193-195,197]

It is estimated that over 70% of signaling proteins contain long intrinsically disordered segments.[24,198] Many of these IDRs contain recognition motifs that act as display sites for post-translational modification (PTM).[189,199-203] Post-translational modifications (e.g.; phosphorylation, glycosylation, and acetylation) are abundant in nature, as they expand the functional diversity of proteins.[10,22,33,124,202] In the human proteome, it is estimated that 20-50% of proteins are glycosylated[204,205] and nearly 90% of proteins are phosphorylated.[202,206-208] The prevalence of display sites for post-translational modification (PTM) in IDRs is striking.[189,199-201,203,209] This is likely due to the inherent solvent-accessibility of IDRs, which would facilitate the binding of PTM-enzymes (e.g.; kinases or *N*-glycosyltransferases).[8,22,74,200,201]

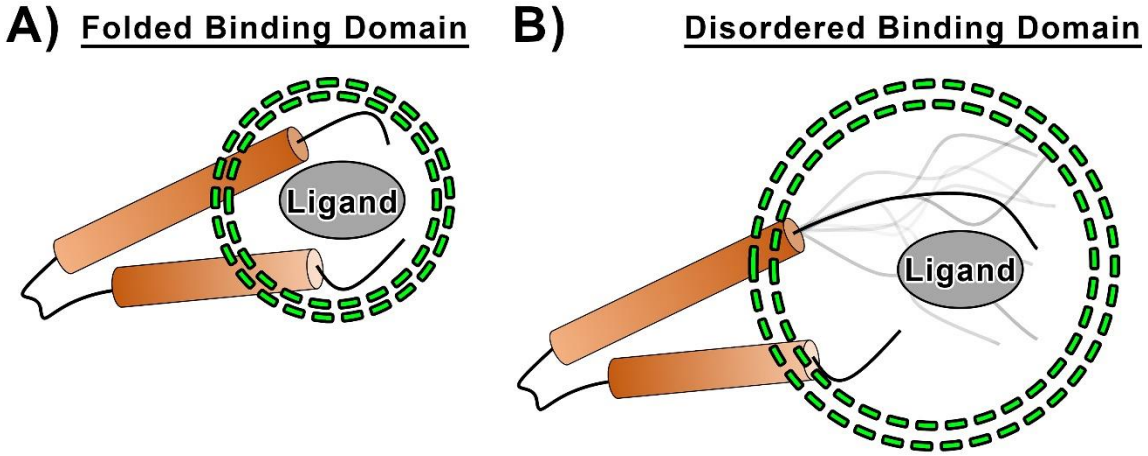


Figure 1.9: The 'fly-casting' mechanism of IDRs.[185] **(A)** The folded binding domain of a protein (orange) has a small capture radius for a ligand (green dashed circle). This yields higher specificity and tighter binding interactions with the ligand. However, the likelihood for productive collisions between the ligand and the binding pocket of the folded domain is small. **(B)** The conformational heterogeneity exhibited by disordered segments (black lines with grey shading) is useful for the weak, transient binding of a ligand. The larger radius of gyration of floppy disordered segments increases the capture radius (green dashed circle) for ligand binding, which enhances the rate of molecular recognition.[185,188] In many cases, the disordered segment undergoes a disorder-to-order transition upon ligand binding (*not depicted, but discussed below*).[17,19,185,186,188,189]

The post-translational modification of an IDR remodels the free energy landscape of either the disordered segment and/or the parent protein (Figure 1.10). This results in changes to the

function of the IDR and/or the parent protein.[13,22,33,74,124,200,206,209-211] The reversibility of post-translational modifications such as phosphorylation (e.g.; kinases and phosphatases) provides cells with the ability to regulate proteins that are important for signaling (Figure 1.10).[124,189,201,202,205,211] Therefore, the PTM of IDRs in signaling proteins allows for various cellular activities to be rapidly turned on or off.[17,22,74,124,185,186,188,192,212] In fact, some signaling proteins contain IDRs that have several phosphorylation sites which undergo reversible modification. This is useful in tuning the function of the parent protein.[213,214]

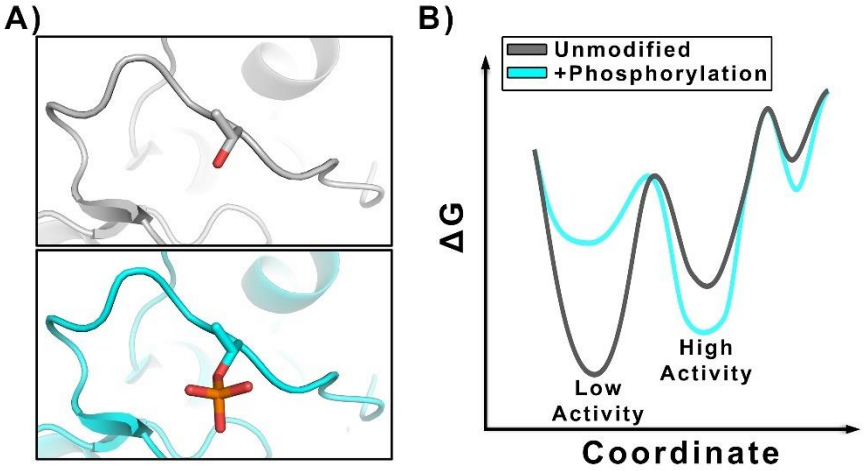


Figure 1.10: The post-translational modification of an intrinsically disordered region alters protein function by remodeling the free energy landscape. (A) A representation of a disordered region in a signaling protein (PDB: 3PY3) that is phosphorylated. The top panel shows the threonine (grey sticks) in the disordered region in the unphosphorylated state (grey), whereas the bottom panel shows the phosphorylated threonine (orange and cyan sticks). **(B)** The phosphorylation of the disordered region modifies the free energy landscape of a protein.[22,124,200,206,209-211,213] In the unmodified state, the protein (grey line) populates a substate with low activity. Upon phosphorylation of the IDR (cyan line), the population shifts to favor a higher activity state.

The interaction or modification of an unfolded polypeptide usually strengthens the inter- or intramolecular bonds in an IDR, which results in a disorder-to-order transition.[9,17,21-23,25,74,174,189,191,197,201,209,212,215-219] In rare cases, the interactions formed between an IDR and its binding partner are so weak that the polypeptide remains 'disordered' in the bound state. These

"fuzzy complexes" allow the disordered segment to maintain a significant degree of conformational heterogeneity when it is transiently bound to its binding partner (Figure 1.11A).[17,165,183,220-223] However, in most cases, a binding interaction will stabilize a disordered segment enough to produce at least some partial folding (Figure 1.11B).[10,17,19-22,25,33,44,155,174,189,194,197,201,216-218] The transition of a disordered segment to a folded or partially folded state produces perturbations to the free energy landscapes of the IDR, parent protein, and/or the binding partner.[9,17,21,23,25,174,189,197,212,216,217] Thus, the interaction or modification induced folding of an IDR results in functional alterations that often have biological significance.[8,9,13,19,20,22-24,45,124,189,200,206,209-211]

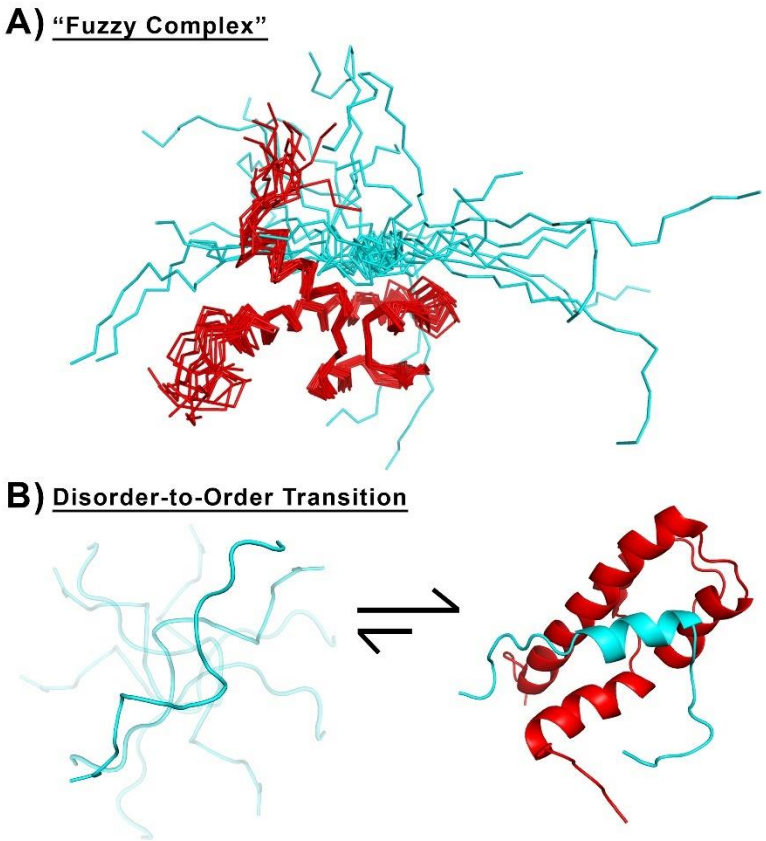


Figure 1.11: Fuzzy complexes[76,183,220-223] and disorder-to-order transitions[17,25,44,174,189,197,216]. **(A)** The multiple states from an NMR solution structure of a "fuzzy complex" (PDB: 2LPB) are depicted. Here, a well-ordered protein (red) forms weak, transient interactions with a disordered peptide (cyan). This complex is described as "fuzzy" because the disordered peptide maintains a large degree of conformational heterogeneity in the bound state. However, in most cases,

disordered regions are stabilized upon binding an effector molecule.[10,17,21,25,33,150,174,189,197,200,216,224] **(B)** The conformational freedom of a free disordered peptide (cyan) is depicted (*left*). Upon complex formation (*right*), interactions stabilize the disordered peptide, which induces local helical folding in a region of the segment. This disorder-to-order transition mechanism has been implicated in the function and regulation of many proteins.[17,25,44,72,116-118,176,183,188,189,225-232]

B) IDRs whose function does not involve any interaction or modification

Historically, disordered loop regions have been regarded to serve as connectors between units of secondary structure.[22,33] This assumption is – in part – due to the fact that disordered regions are dynamic, which results in a lack of observable electron density (i.e.; high *B*-factors) for these floppy segments in X-ray crystal structures.[22,33,174] In fact, it is estimated that ~7% of structures in the protein data bank (PDB) have resolved density for their entire sequence.[35] Thus, disordered loops are expected to be abundant within the proteomes of all organisms, yet the majority of IDRs lack functional annotation.[22] Recent work has begun to unveil the importance of IDRs in *i)* enhancing protein solubility, and *ii)* affecting protein stability.[22,33,45,233-237] These roles of IDRs are often referred to as 'entropic chain' activities.[22,45,176,238-241] The following subsections outline the different entropic-based functions of IDRs:

i) The role of IDRs in enhancing protein solubility (entropic bristles)

For the recombinant expression of most proteins, the preferred host system is *Escherichia coli*.[242] The *E. coli* host system provides many benefits for the mass production of non-host proteins, such as the ability to easily transform expression vectors containing the gene of interest and the ability to grow cells to a high density in a rapid manner.[242] However, this well-understood host system does not always yield high amounts of recombinant protein (Figure 1.12A).[242,243] To enhance production, recombinant proteins are often expressed with an attached fusion partner (e.g.; maltose-binding protein) (Figure 1.12B).[242,244-246] The fusion partners that are typically utilized in

this approach are well-folded proteins.[233,242,244-246] These well-folded fusion partners act as solubilizers, which slows the aggregation rate of the recombinant protein during expression (Figure 1.12A,B).[233,243,244,247]

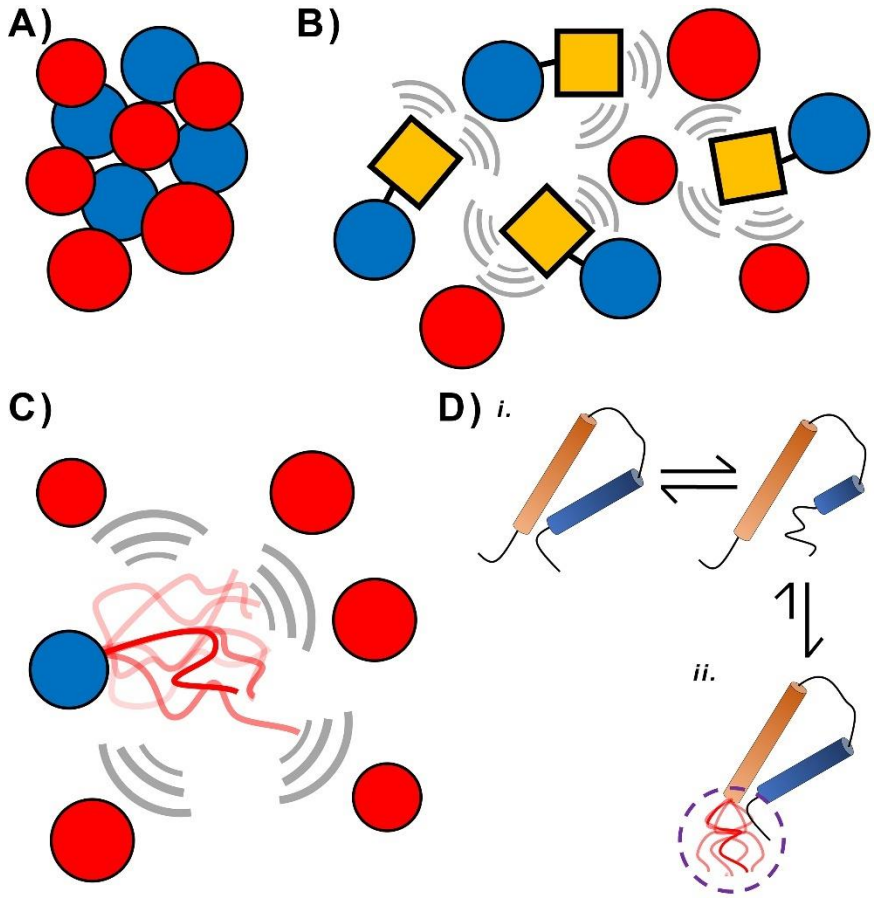


Figure 1.12: Fusion partners and entropic bristles. (A) The overexpression of a protein whose surface has an overall net positive charge (blue circles) has a propensity to interact and aggregate with proteins or other macromolecules in the cell that have an opposite, negative charge (red circles).[248-255] The overexpression of a recombinant protein can also lead to self-aggregation (*process not depicted, but discussed in panel D*).[242,245,248,250,251,253-256] **(B)** Well-folded proteins (orange boxes) expressed as fusion partners to a recombinant protein (blue circles) can decrease the propensity for aggregation by excluding volume (grey lines).[242,245,247] This excluded volume effect[250,257-264] can also increase the solubility and stability of the recombinant protein (*mechanism discussed in panel D*). **(C)** The use of intrinsically disordered segments as fusion partners (red lines) can also decrease the propensity for aggregation. These entropic bristles often have a high net charge, which helps to repel similarly charged molecules.[233] Due to a larger radius of gyration, an entropic bristle can exclude more volume than a well-folded fusion partner. Therefore, entropic bristles may provide more utility in preventing aggregation during recombinant protein expression.[233,265,266] **(D)** The excluded volume effects of a fusion partner can stabilize the folded

state of a recombinant protein. *i)* A schematic of a protein (orange and blue cylinders represented as helices) that is in equilibrium between a well-folded state (*left*) and a less stable state (*right*). Specifically, one helix (blue cylinder) undergoes local unfolding (*right*, black line). This local unfolding can also result in protein-protein interactions that lead to self-aggregation or undesired interactions with other macromolecules.[256,267] *ii)* The attachment of a fusion partner (red line) shifts the equilibrium of states to favor the well-folded, stable state of the protein through excluded volume effects.[250,251,257-264,268] In other words, the fusion partner (red line) excludes the space available for the helix (blue cylinder) to unfold (purple dashed line), which entropically stabilizes the folded state of the protein.[233,266,268] In addition to fusion partners, both glycosylation[269,270] and PEGylation[271] can also be used to entropically stabilize proteins.[268]

Recent work has shown that intrinsically disordered segments can also be used as fusion partners for improving the production of recombinant proteins.[233,266] Because of the characteristic flexibility of disordered segments, they have a much larger radius of gyration than well-folded proteins.[174,272-275] In addition, IDRs usually have a high net charge.[34,36,39,53,174,272,273] These features enable disordered segments to exclude large regions of volume and repel similarly charged molecules.[233,266,276] Therefore, the use of intrinsically disordered segments as fusion partners (often referred to as 'entropic bristles') in recombinant expression has drawn much attention in recent years (Figure 1.12C,D).[174,233,265,266,276] To avoid aggregation in the crowded cellular environment[235,248,250,251,277,278], it is likely that some proteins have evolved entropic bristles to suppress aggregation.[13,22,233,266,277] This could be one explanation for the abundance of intrinsically disordered segments across the proteomes of all organisms.[22,277]

ii) IDRs alter protein stability

In a polypeptide, the unfavorable interactions of non-polar residues with solvent drives protein folding (i.e.; hydrophobic effect). This results in the tight packing of atoms in the stable, folded structure of a protein.[152,231,279-283] However, the 'native' fold of most proteins is dynamic and can adopt various conformations that are essential for function.[29,55-58,65,66,72,74,102,116-120] To access different conformations, the subdomains that make up most protein structures will usually

rotate or bend around specific regions. These 'hinge' regions are typically located in areas of instability, where atoms in the main chain of the tertiary structure are less tightly packed (e.g.; loop regions).[33,63,66,284,285] Thus, the role of disordered loops has been implicated in facilitating structural rearrangements that are important for both protein function and regulation.[22,33,286,287]

Numerous studies have drawn interesting conclusions about the effect of both loop composition and loop length in altering protein stability.[22,33,182,286-293] The first comprehensive study on the effect of loop composition and length was conducted by Ladurner and Fersht.[288] The authors utilized the small (64 a.a.), two-state folding protein chymotrypsin inhibitor-2 (CI2), which has been extensively used as a model to study the mechanism of protein folding.[294-296] In this pioneering study, a series of insertions with different residue compositions (poly-Gln, poly-Ala, and poly-Gly) were introduced into a 3-residue surface loop of CI2. By conducting fluorescence equilibrium denaturation and re-folding studies with these mutants of CI2, the authors revealed that the protein becomes less stable as the length of the loop increases. In other words, relative to the wild-type (3 residue loop), a positive change in $\Delta\Delta G$ (decrease in stability) is observed as loop length is increased. This loss in stability does not increase linearly with increasing loop length (Figure 1.13A). Instead, as the loop length increases the destabilizing effect on the protein eventually saturates (Figure 1.13A). Thus, the authors show that the relationship between protein stability and loop length follows an exponential-like dependency (Figure 1.13A). The magnitude of the effect near saturation (10 residue insertion) was ~2.5 kcal•mol^{-1} for the poly-Gln insertion. At this length, the loop composition was changed (to poly-Gly and also poly-Ala). Despite different compositions, the overall change in stability was similar. The 10-residue poly-Gly and poly-Ala insertions destabilized the protein by ~1.7 kcal•mol^{-1} and ~2 kcal•mol^{-1}, respectively. Similar studies[289-292] on loop length (and composition) with other proteins have revealed nearly

identical changes in protein stability due to loop elongation (Figure 1.13B). Importantly, all these studies show that increasing loop length results in an exponential decrease in protein stability (Figure 1.13A,B).[288-292]

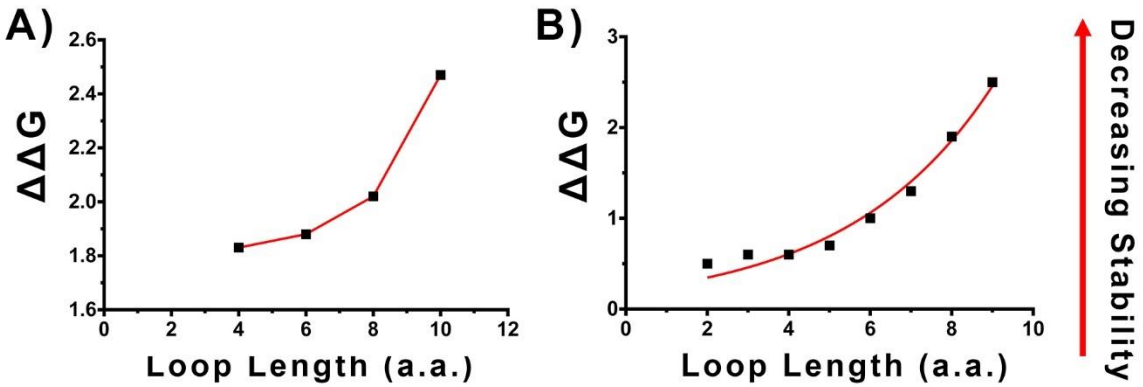


Figure 1.13: Increasing the length of a disordered loop results in an exponential decrease in protein stability. (A) Re-plotted data from Ladurner and Fersht (Table 1)[288] showing the exponential-like decrease in protein stability upon loop elongation in CI2. Unfortunately, due to a lack of data points, the data could not be adequately modeled with an exponential fit. However, points are connected by lines to reflect the exponential-like dependency of protein stability on loop length. **(B)** Re-plotted data from Nagi and Regan (Table 1)[289], which more rigorously studied the effect of loop elongation on protein stability. The data clearly shows that as loop length is increased, there is an exponential decrease in protein stability.

The entropic cost associated with confining a polymer also exhibits an exponential dependency on the length of the polymer.[261,268,297-299] The magnitude of the entropic cost depends on the geometry in which the polymer is confined.[298] For a disordered segment confined within the structure of its parent protein, the geometry of confinement is difficult to define.[288] Nonetheless, there appears to be a link between the length of a disordered loop and protein stability, which likely originates from the entropic cost of confinement.[288-290] In other words, the protein structure incurs an entropic penalty by confining the conformational freedom of a disordered loop. As the length of the disordered loop is increased, the entropic penalty of confinement increases exponentially. At long lengths, the loop begins to escape the confined space, and as a result, the

entropic cost associated with confinement starts to diminish. Therefore, the exponential effect of loop elongation on protein stability eventually approaches an asymptote (Figure 1.13B).

Recent simulations[297,300-302] have demonstrated that tethering a polymer to a surface results in a net 'entropic pulling force' at the point of attachment (Figure 1.14). In fact, tethering intrinsically disordered proteins to the surface of a membrane can exert a strong enough pulling force to distort the shape of the membrane.[303] This entropic force generated by intrinsically disordered segments is also expected to be involved in the mechanism of protein translation[304-309] and in altering protein dynamics and function[81].

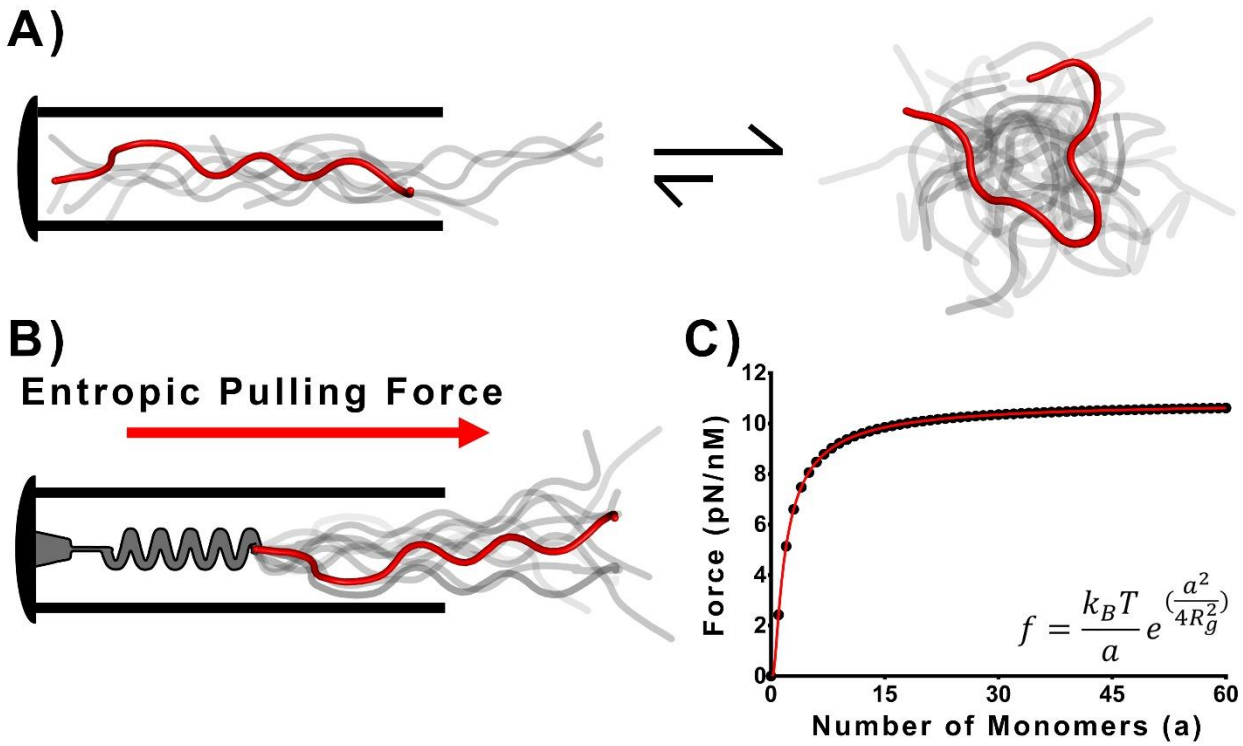


Figure 1.14: The entropic pulling force generated by intrinsically disordered segments. (A) The conformational freedom of a polymer (red and grey shading) is reduced when it is confined within a non-interacting tube (*left*). As a result, the polymer favors being in solution where it can access more degrees of freedom (*right*). In other words, entropy drives the polymer out of the tube. **(B)** A polymer is confined within a tube and tethered to a spring. Because entropy favors the polymer leaving the tube, the polymer will exert a pulling force at the point of attachment on the spring. This is referred to as an 'entropic pulling force'. **(C)** Modeling the entropic pulling force (*f)* exerted by a disordered polymer that is tethered to a surface. Driven by the unfavorable entropy

associated with confinement and thermal fluctuations (k_BT), a constrained polymer exerts a pulling force (f) at the point of attachment. This entropic pulling force increases exponentially as the number of monomer units (a) increases. At longer lengths the polymer becomes less confined by the surface, so the net entropic pulling force begins to saturate. Tethering a disordered polymer within a confined geometry is predicted to produce strong pulling forces (~10-15 pN)[300,301] that are biologically relevant. For example, unfoldases/translocases exert ~10-15 pN of force on folded proteins that are destined degradation or translocation, respectively.[310-316] In addition, tethering an intrinsically disordered polypeptide in a confined space can produce forces strong enough to: *i)* distort the curvature of membranes[303]; *ii)* drive the movement of a nascent polypeptide out of the ribosome exit tunnel[304-309], and *iii)* alter protein dynamics and function.[81]

1.4 The atypical allosteric mechanism of hUGDH

Human UDP-α-D-glucose dehydrogenase (hUGDH) uses two NAD^+ molecules to oxidize UDP-α-D-Glucose (UDP-Glc) to UDP-Glucuronic acid (UDP-GlcA).[80,317] The activity of the enzyme is regulated by an atypical allosteric inhibition mechanism, whereby the downstream metabolite UDP-α-D-Xylose (UDP-Xyl) competes for the active sites of hUGDH.[81,318-323] Formed from the association of three dimers, the hexameric enzyme can adopt three distinct states (Figure 1.15A,B).[81,318,320-325] The unliganded E* state of the hexamer is distinguished by the 'open' conformation of the NAD^+-binding domains (Figure 1.15B). Upon substrate binding, the E* state slowly isomerizes to the active E state (Figure 1.15B). This slow isomerization can be observed as a lag (hysteresis) in progress curves.[81,318,320,323] When the allosteric feedback inhibitor UDP-Xyl binds to the active site of hUGDH, the enzyme undergoes a conformational change to the horseshoe-shaped, inhibited E^{Ω} state (Figure 1.15B).[81,318,320-323] Compared to the active E state, the inhibited E^{Ω} state has a lower affinity for UDP-Glc, but a higher affinity for UDP-Xyl.[318,323] Therefore, the allosteric transition ($E \rightleftharpoons E^{\Omega}$) is observed as positive cooperativity in substrate saturation curves that contain UDP-Xyl.[81,318,320,322,323]

The communication between the subunits of the hexamer is facilitated by the conformational change of the buried Thr131-loop/α6-helix (allosteric switch) (Figure 1.16).[81,318,320,322,323] The allosteric switch links the active site to the hexamer building interfaces of

the enzyme (Figure 1.15B). The binding of the inhibitor UDP-Xyl to the active site of hUGDH elicits a conformational change in the allosteric switch, which increases the affinity between adjoining subunits (Figure 1.15B). This produces the horseshoe-shaped hexamer, where enzyme is in the inhibited E^{Ω} state and the subunits have a higher affinity for UDP-Xyl.[81,318,320-323] The conformational change of the buried allosteric switch requires flexibility in the core of the protein.[318,321-323] Specifically, cavities and declivities (packing defects) in the protein core enable the ligand-induced movement of the allosteric switch between the three distinct conformations (E, E*, and E^{Ω}) (Figure 1.16).[318,321,323]

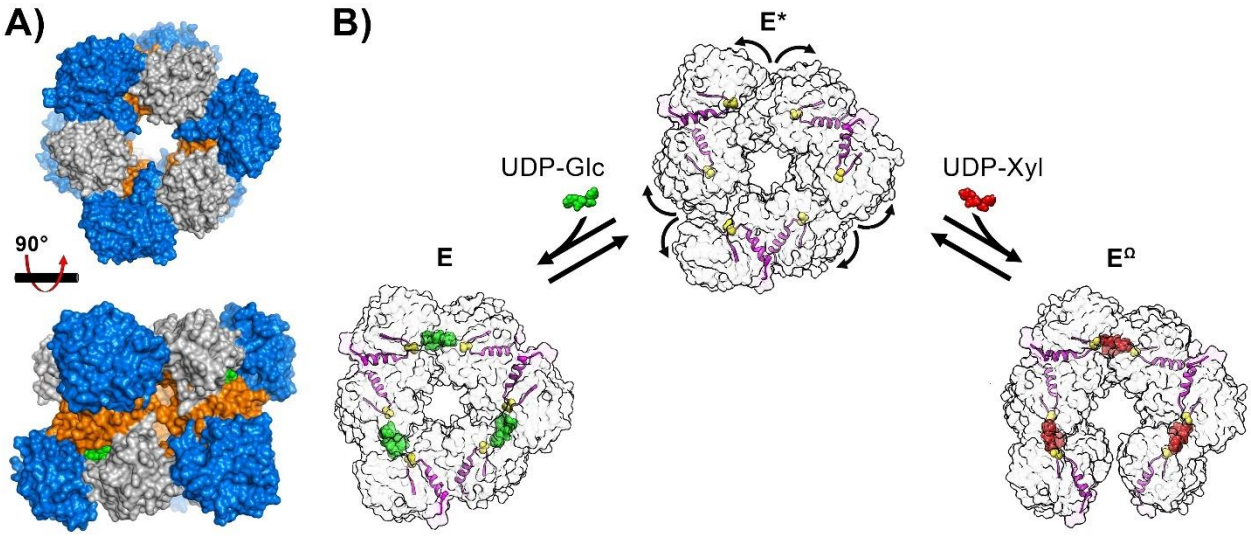


Figure 1.15: The atypical allosteric inhibition mechanism of hUGDH. (A) Top and side views of the hUGDH hexameric complex (trimer of dimers). The domains of each monomer subunit are colored: NAD^+ binding domain (blue), nucleotide-sugar binding domain (grey), and dimerization domain (orange). The active sites are depicted with UDP-Glc bound (green spheres). **(B)** In the unliganded state, the enzyme adopts the inactive E* state (*middle*) that is distinguished by the 'open' conformation adopted by the NAD^+ binding domains (black curved arrows).[320] Upon binding the substrate UDP-Glc (green spheres), the enzyme slowly isomerizes to the active E state (*left*). The allosteric feedback inhibitor UDP-Xyl (red spheres) competes for the active sites of hUGDH, and upon binding induces a conformational change in the enzyme to produce a horseshoe-shaped hexameric complex. This inhibited E^{Ω} state has an increased affinity for UDP-Xyl compared to the active E state.[318,323] The allosteric transition between states ($E \rightleftharpoons E^* \rightleftharpoons E^{\Omega}$) is controlled by the conformation of the allosteric switch (magenta), which links the active sites to the hexamer building interfaces. In the E^{Ω} state, the UDP-Xyl induced conformational change of the buried allosteric switch increases the affinity between adjoining subunits.[81,318,320-323]

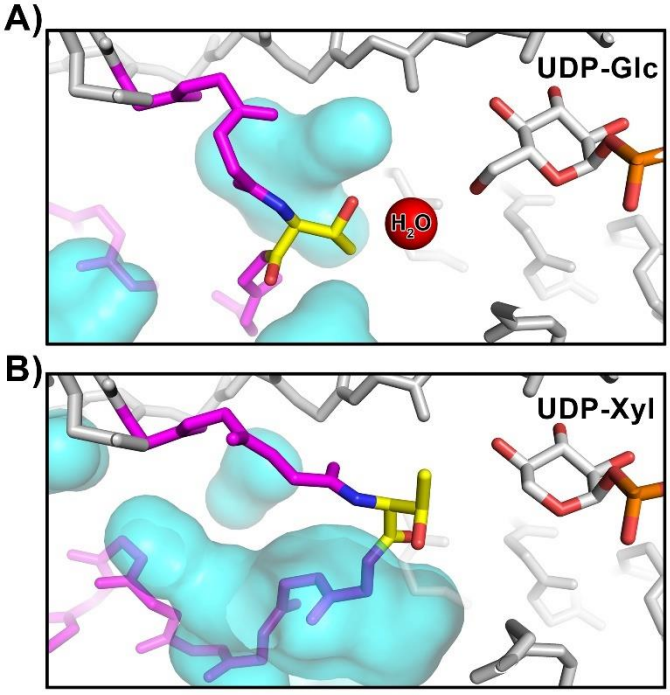


Figure 1.16: The packing defects in the protein core of hUGDH enable the conformational change of the buried allosteric switch. (A) The active E state conformation of the allosteric switch (magenta) is depicted. Here, a water molecule (red sphere) hydrogen bonds with the Thr131 (yellow) of the allosteric switch and the C5'CH_2OH of UDP-Glc. **(B)** The absence of the C5'CH_2OH in UDP-Xyl triggers the buried allosteric switch to change conformations, which produces the inhibited E^{Ω} state.[81,318,320-323] Packing defects (cyan) facilitate the conformational change of the allosteric switch in the core of the protein.[321,323]

The packing defects are essential the movement of the allosteric switch between states ($E \rightleftharpoons E^{\Omega}$) (Figure 1.16).[318,320-323] A comparison between the allosteric, human enzyme and the non-allosteric enzyme in *Streptococcus pyogenes* (spUGDH) suggested that allostery originates from a series of six large-to-small substitutions in the core of the protein near the allosteric switch.[321] Therefore, these putative packing defect residues are believed to have facilitated the evolution of atypical allostery in hUGDH. In fact, the mutation of the putative packing defect residue Ala104 to a leucine has been shown to abolish the allostery.[323] Recent work (Chapter 3) with the homolog from *Caenorhabditis elegans* (cUGDH; overall 61% sequence identity to hUGDH) supports the

fact that the packing defects in the protein core are essential for both hysteresis and allostery. This work also supports the hypothesis that these phenomena (hysteresis and allostery) are coupled in the atypical allosteric mechanism of UGDH.

The characterization of cUGDH expanded the atypical allostery sequence motif to include a co-evolving residue (Ser290 in cUGDH; Asn283 in hUGDH). By conducting a phylogenetic analysis of UGDH sequences from diverse taxonomic groups, we tracked the evolution of the atypical allostery motif. This analysis revealed a potential transition point in the evolution of atypical allostery (Ala141-to-Thr141; cUGDH numbering). The Ala-to-Thr substitution would bury a polar group in the core of the protein, which would likely require large structural re-arrangements and may prohibit the formation of the E^{Ω} complex. However, future mutational work will be needed to test the impact of this substitution on the atypical allosteric mechanism of hUGDH.

The work presented in Chapter 2 shows that the entropic force generated by the intrinsically disordered C-terminus (ID-tail) of hUGDH alters the stability and dynamics of the enzyme.[81] This shifts the equilibrium of states in hUGDH towards the inhibited E^{Ω} state, which enhances the affinity for the allosteric inhibitor UDP-Xyl. The function of the ID-tail does not depend on either its structure or its sequence. Instead, the length of the ID-tail tunes the affinity for UDP-Xyl in an exponential like fashion. This result is consistent with the protein structure constraining the conformational freedom (entropy) of the ID-tail. In other words, the protein structure incurs an entropic cost for confining the ID-tail, which alters the dynamics of the protein to favor the E^{Ω} state. Specifically, this work shows that the ID-tail enhances the dynamics of the buried allosteric switch. At an atomistic level of detail, it would be interesting to see how the entropic force generated by the ID-tail propagates an effect throughout the entire structure. Molecular dynamics

simulations would be useful in elucidating how the ID-tail perturbs the allosteric network of hUGDH. However, due to the size of hUGDH (~55 kDa per monomer) and the timescale that the allosteric transition occurs on (seconds), molecular dynamics simulations are currently impractical.

CHAPTER 2

THE ENTROPIC FORCE GENERATED BY INTRINSICALLY DISORDERED SEGMENTS TUNES PROTEIN FUNCTION

Nicholas D. Keul, Krishnadev Oruganty, Elizabeth T. Schaper Bergman, Nathaniel R. Beattie, Weston E. McDonald, Renuka Kadirvelraj, Michael L. Gross, Robert S. Phillips, Stephen C. Harvey, Zachary A. Wood. *Nature*, 563(7732), p.584 (2018). Reprinted here with permission of the publisher.

2.1 Abstract

Protein structures are dynamic and can explore a large conformational landscape.[55,56] Only some of these structural substates are important for protein function (such as ligand binding, catalysis and regulation).[72,74,75] How evolution shapes the structural ensemble to optimize a specific function is poorly understood.[74,75] One of the constraints on the evolution of proteins is the stability of the folded 'native' state. Despite this, 44% of the human proteome contains intrinsically disordered peptide segments greater than 30 residues in length[32], the majority of which have no known function.[22,33,36] Here we show that the entropic force produced by an intrinsically disordered carboxy terminus (ID-tail) shifts the conformational ensemble of human UDP-α-D-glucose-6-dehydrogenase (UGDH) towards a substate with a high affinity for an allosteric inhibitor. The function of the ID-tail does not depend on its sequence or chemical composition. Instead, the affinity enhancement can be accurately predicted based on the length of the intrinsically disordered segment, and is consistent with the entropic force generated by an unstructured peptide attached to the protein surface.[297,300-302] Our data show that the unfolded state of the ID-tail rectifies the dynamics and structure of UGDH to favour inhibitor binding. Because this entropic rectifier does not have any sequence or structural constraints, it is an easily acquired adaptation. This model implies that evolution selects for disordered segments to tune the energy landscape of proteins, which may explain the persistence of intrinsic disorder in the proteome.

2.2 Main Text (Introduction, Results, and Discussion)

Intrinsically disordered segments can exhibit complex functions such as ligand binding, scaffolding of multi-protein complexes and mediating allosteric regulation.[17,151,175,326,327] However, many intrinsically disordered segments are assumed to be nonfunctional and are often removed from proteins to facilitate structural studies. For example, the 30-residue disordered C terminus of UGDH (residues 465–494) is often removed with no apparent impact on kinetic parameters.[80] Here, we show that this C-terminal segment (called the ID-tail) plays a role in the allosteric mechanism of UGDH. UGDH catalyses the NAD^+-dependent oxidation of UDP-α-D-glucose (UDP-Glc) to UDP-α-D-glucuronic acid[80], and is regulated by the allosteric feedback inhibitor UDP-α-D-xylose (UDP-Xyl).[319,328] Three UGDH dimers associate to form an inactive hexamer (E*)[318,320-322,324] (Figure 2.1a, b). The binding of substrate induces an allosteric switch (T131-loop–α6 helix) in the E* hexamer to produce the active state (E)[318,320,321,325] (Figure 2.1a, c). The allosteric inhibitor UDP-Xyl competes with UDP-Glc for the active sites, and upon binding, triggers the allosteric switch to produce the inhibited state (E^{Ω}).[318,320-322] This inhibition mechanism is atypical in that the active site also functions as an allosteric site to control the structure and activity of the hexamer[318,320-322,324,325] (Figure 2.1a, c). The E^{Ω} state has a high affinity for UDP-Xyl and a low affinity for UDP-Glc.[318,320] Therefore, the allosteric transition of the inhibited E^{Ω} hexamer to the E state can be observed as cooperativity in substrate saturation curves.[318,320] We compared the structure and activity of full-length UGDH (UGDH(FL)) to a construct lacking the ID-tail (UGDH(ΔID)). We solved the structures of E* states of UGDH(FL) and UGDH(ΔID) in isomorphous crystal lattices, showing that there were no substantial differences (Figure S2.1a–c). UGDH(FL) and UGDH(ΔID) also have a similar catalytic rate constant (k_{cat}) and Michaelis constant (K_M) for both substrate and coenzyme, consistent with earlier

reports[80] (Table S2.2). By contrast, the allosteric response is sensitive to the ID-tail; deletion of the ID-tail reduces the affinity for UDP-Xyl by an order of magnitude (Figure 2.1d). UGDH(ΔID) still binds UDP-Glc cooperatively, indicating that the deletion of the ID-tail reduces UDP-Xyl affinity but does not prevent the formation of the E^{Ω} hexamer (Figure 2.1d, Figure S2.2a, b).

Both the ID-tail and the α6 helix of the allosteric switch are located in the hexamer-building interface between adjacent dimers, suggesting that these two elements may work together to increase the affinity for UDP-Xyl (Figure 2.1b). We used the allostery-quenching A136M substitution to determine whether the ID-tail functions independently of the allosteric switch. This substitution has been shown to lock the allosteric switch and the hexamer in the low UDP-Xyl-affinity E state.[318] Inhibition studies show no marked difference between the UDP-Xyl affinities of UGDH(FL/A136M) and UGDH(ΔID/A136M), which suggests that the ID-tail requires both a functional allosteric switch and the E^{Ω} state to enhance the affinity for UDP-Xyl (Figure 2.1d).

The location of the α6 helix in the hexamer-building interface suggests that the oligomeric structure might be important for the function of the ID-tail (Figure 2.1b). Sedimentation velocity studies show that the UGDH(ΔID) E* hexamer is slightly less stable than the UGDH(FL) E* hexamer, perhaps explaining its reduced affinity for UDP-Xyl (Figure S2.3a). We tested the role of the hexamer with the M11 interfacial loop substitution, which prevents hexamer formation and stabilizes the dimer (UGDH(FL-dimer) and UGDH(ΔID-dimer)).[320] UDP-Xyl binds to the UGDH(FL-dimer) with sevenfold higher affinity than UGDH(ΔID-dimer), demonstrating that the ID-tail does not require the hexamer to enhance the affinity for UDP-Xyl (Figure 2.1d and Supplementary Information Section 1).

The ID-tail is highly conserved in vertebrate UGDHs (Figure 2.2a). We examined the importance of primary structure in the ID-tail by randomizing the native sequence to create two

distinct ID-tails (UGDH(R1) and UGDH(R2)) (Figure 2.2b). Surprisingly, the UGDH(FL), UGDH(R1) and UGDH(R2) constructs have similar affinities for UDP-Xyl (Figure 2.2c). Next, all six prolines in the ID-tail were substituted with serine (UGDH(−Pro)) (Figure 2.2b). Because serine and proline both promote disorder[329,330], this substitution conserves the unfolded state and disrupts any possible proline-specific interactions. Analysis of UGDH(−Pro) shows that the prolines do not contribute to UDP-Xyl affinity (Figure 2.2c). Because all of the above constructs conserve the positive charge of the native ID-tail ($pI = 10.1$), we created a negatively charged ID-tail ($pI = 4.4$) using lysine-to-serine substitutions (UGDH(−Lys)) (Figure 2.2b). Despite the charge switch, there was still no substantial change in UDP-Xyl affinity (Figure 2.2c). Finally, we replaced the ID-tail with polyserine (UGDH(Ser)) without causing a marked change in UDP-Xyl affinity (Figure 2.2b, c). Therefore, the conserved primary structure is not required for UDP-Xyl affinity, but may have been selected for because of an additional, unrelated function in vivo (Figure 2.2a). The absence of sequence constraints argues against any mechanism in which the ID-tail specifically interacts with the inhibitor or the protein.

Next, we considered the possibility that the ID-tail might enhance UDP-Xyl affinity through a sequence-independent interaction involving the polypeptide backbone. Because the six prolines in the UGDH(FL), UGDH(R1) and UGDH(R2) ID-tails sample 16 unique positions throughout the sequence without altering UDP-Xyl affinity, it is unlikely that a backbone-specific interaction is important for function (Figure 2.2b, c). Nevertheless, if there was a backbone-specific interaction, a plot of affinity versus ID-tail length would reveal a discontinuity when the critical segment was removed. Inhibition studies comparing UGDH(FL), UGDH(ΔID) and three new constructs with ID-tails of varying length (UGDH(2×FL), UGDH(0.5×FL) UGDH(0.26×FL) and UGDH(0.13×FL), shown in Figure 2.2b) show that the affinity can be modelled as a simple

exponential decay (Figure 2.2d). We confirmed that this saturable effect is independent of sequence by using polyserine ID-tails of corresponding lengths (UGDH(Ser), UGDH(0.5×Ser), UGDH(0.26×Ser) and UGDH(0.13×Ser)) and similarly, using corresponding lengths of the scrambled R1 construct (UGDH(R1), UGDH(0.5×R1), UGDH(0.26×R1) and UGDH(0.13×R1)) (Figure 2.2d). It is notable that UGDH(0.13×FL), UGDH(0.13×Ser) and UGDH(0.13×R1) still enhance UDP-Xyl binding affinity; the conformations of these short, four-residue ID-tails are tightly constrained within a surface pocket, which should stabilize any weak structure (Figure 2.3a). Nevertheless, none of the E, E* and E^{Ω} UGDH(FL) crystal structures (42 unique chains) show evidence of an ordered interaction within the pocket (Figure S2.1 and refs[318,321,322,325]).

The data presented so far provide strong evidence that the high-affinity binding of UDP-Xyl is a function of the unfolded state of the ID-tail. An unstructured polymer tethered to a surface generates an entropic force at the point of attachment[300-302], which can be strong enough to distort lipid bilayers[303] and alter protein stability[297]. This force originates from the volume exclusion effects of the surface, which reduce the conformational entropy of the attached polymer (Figure 2.3b). Because the entropy of the polymer increases with distance from the surface, the entropic force converges to a maximum value as the chain length increases.[300-302] The unfavourable change in free energy produced by constraining an unstructured, non-interacting peptide ($\Delta G_{\text{constrained}}$) is:

$$\Delta G_{\text{constrained}} = -RT\, ln\left(\frac{\Omega_2}{\Omega_1}\right) \qquad \text{(equation 1)}$$

where Ω_1 is the sum of all possible states of an unconstrained peptide and Ω_2 is the subset of states constrained by the protein surface and the adjacent ID-tail (RT is the product of the molar gas constant, R, and the temperature, T). Using Monte Carlo sampling of coarse-grained, sterically allowed bins of ϕ and ψ torsion angles we calculated the fraction of constrained conformations for various ID-tail lengths (see Methods, Figure 2.3b, Figure S2.5). For this simulation, the adjacent

ID-tail was held in a fixed conformation (Figure S2.5). If the conformational entropy of the ID-tail contributes to the change in UDP-Xyl affinity, then we would expect Ω_2/Ω_1 and the affinity constant K_i to display similar behaviour with increasing tail length. Despite the simplicity of the Monte Carlo model, the simulations confirm that Ω_2/Ω_1 converges as the ID-tail length increases (Figure 2.3c).

Studies have shown that the entropic force generated by a tethered polymer can alter protein stability.[297] We carried out thermal denaturation studies of UGDH dimers (chosen to avoid complications arising from hexamer dissociation), and found that the high-affinity UGDH(FL-dimer) ($K_i = 0.17$ μM) is less stable than the low-affinity UGDH(ΔID-dimer) ($K_i = 1.23$ μM) (Figure 2.3d). The destabilizing effect of the ID-tail should also be reflected in the structure and dynamics of UGDH. To examine these changes at the peptide level, we compared the hydrogen–deuterium exchange (HDX) rates of UGDH(FL-dimer) and UGDH(ΔID-dimer) using mass spectrometry. As expected, the fragment corresponding to the ID-tail is fully exchanged in less than 120 s, which is consistent with a disordered peptide[331] (Figure S2.6a). The ID-tail increases the HDX rates of several segments in the NAD^+ binding domain, with the largest increases occurring in the allosteric switch and an adjacent peptide (Figure 2.3e–g). An increase in HDX rates for a buried peptide such as the allosteric switch and the surrounding segments indicates an increase in the overall dynamics of the domain. This is notable, because the binding of UDP-Xyl induces the allosteric switch and surrounding core residues to change conformation and repack into the high affinity E^{Ω} state[318,320] (Figure 2.1a, c). The ID-tail also decreases the HDX rates of several segments in the dimerization and sugar-binding domains, suggesting that these areas become more structured (Figure 2.3e, g). The largest decrease is observed in the α9 helix of the dimerization domain (residues 222–240). This helix is largely inaccessible to solvent in crystal

structures, which suggests that the ID-tail reduces the overall dynamics of the dimer interface (Figure 2.3g). Overall, the data show that the cost of constraining the ID-tail destabilizes a low-affinity substate, which biases the conformational ensemble towards a structurally and dynamically distinct high-affinity substate. A simple exponential fit of Ω_2/Ω_1 (Figure 2.3c) shows that the energetic cost of constraining the ID-tail converges to approximately 2.4 kcal•mol^{-1} (Equation (1)). Therefore, our simple Monte Carlo model supports the argument that entropic confinement effects generate sufficiently strong forces to explain the maximum expected gain in UDP-Xyl binding affinity of −1.45 kcal•mol^{-1} (Figures 2.2d, 2.3c, Figure S2.5). More rigorous calculations on other systems using simpler polymer models (and simpler confinement geometries) also find confinement free-energy costs of the same magnitude.[298,332]

If the ID-tail favours the dynamics associated with the repacking of the allosteric switch into the E^{Ω} state, then we would expect to see a difference in the activation (E* to E) and inhibition kinetics (E* to E^{Ω}) (Figure 2.1a). Pre-steady-state analysis of progress curves shows that the ID-tail slows the rate of activation hysteresis (E* to E) by 39% (Figure 2.4a). Next, we examined the UDP-Xyl-induced isomerization of UGDH to the EΩ state. Transient-state analysis of UDP-Xyl binding kinetics revealed a three-phase exponential decay of UGDH time-resolved tryptophan fluorescence, and the data were globally fit by computer simulation (see Methods and Figure S2.7a–e). The same kinetic model produced the best fit for both UGDH(FL) and UGDH(ΔID) and predicts UDP-Xyl affinities that are consistent with our steady-state inhibition studies (Figure S2.7):

$$E^{*} + UDX \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E^{*}{:}UDX \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} E^{\dagger}{:}UDX \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} E^{\Omega}{:}UDX \quad \text{(equation 2)}$$

Where k_n is the rate constant for reaction n. According to this model, UDP-Xyl binds to the E* state and induces two sequential isomerizations. On the basis of the allosteric model, we had expected a single isomerization from E* to the E^{Ω} state (Figure 2.1a). We call the additional transient $E^{\dagger}$; it represents an intermediate between the E* and E^{Ω} states. The ID-tail changes the kinetic parameters of each transient observed in the time-resolved fluorescence (Figure S2.7e). The largest effect of the ID-tail is a 4.4-fold enhancement of the initial UDP-Xyl binding step, corresponding to a −0.9 kcal•mol^{-1} gain in affinity (Figure 2.4b). The kinetic model predicts an overall favourable gain in binding affinity of −1.3 kcal•mol^{-1}, which agrees well with the observed gain of −1.39 kcal•mol^{-1} (Figure 2.4b, Table S2.2). The different stabilities of the corresponding UGDH(FL) and UGDH(ΔID) transients, combined with the fact that the ID-tail slows activation hysteresis and accelerates inhibition kinetics, supports our conclusion that the ID-tail alters the energy landscape to favour inhibition by UDP-Xyl (Figure 2.4c).

Collectively, our data supports a model in which the entropic force of the ID-tail rectifies the energy landscape of UGDH to favour a substate with a high affinity for UDP-Xyl. We can now interpret the exponential curve in Figure 2.2d as follows:

$$K_i(l) = K_i^{\text{unbiased}} e^{-kl} + K_i^{\text{biased}}(1 - e^{-kl}) \qquad \text{(equation 3)}$$

This implies that: *(i)* UGDH exists as an ensemble of low-affinity (K_i unbiased) and high-affinity (K_i biased) substates; *(ii)* the ID-tail functions as a length (l)-dependent entropic rectifier that shifts (with bias k) the distribution towards the high affinity substate; and *(iii)* the observed UDP-Xyl affinity results from a fractional summation of the low and high affinity substates at a given ID-tail length (Figure 2.4c). The fit to equation (3) produces a K_i biased of 0.46 ± 0.18 μM UDP-Xyl, which corresponds to a maximum favourable gain in binding energy of approximately −1.45 kcal•mol^{-1}. The lack of sequence constraints implies that the entropic force of any intrinsically

disordered segment is capable of shaping the conformational ensemble of a protein. In fact, an N-terminal hexahistidine affinity tag has been shown to alter the internal dynamics of a myoglobin.[333] Thus, the persistence of low-complexity intrinsically disordered segments in the proteome may reflect the selection for entropic rectifiers that can tune the function of a protein by shaping the native-state ensemble.

2.3 Materials and Methods

Protein expression, purification, and quantification of UGDH constructs

All UGDH coding sequences were synthesized and cloned into pET-15b vectors (Norclone). Sequences contained an N-terminal hexahistidine affinity tag adjacent to a tobacco etch virus (TEV) cleavage site. The expression and purification of UGDH constructs were conducted under identical conditions as previously described.[318,320-322,324,325] Following purification, the N-terminal hexahistidine tag was cleaved with TEV protease. An additional immobilized metal affinity column (IMAC) was used to obtain the pure, His-tag-free protein. Unless otherwise noted, all proteins were dialysed into a storage buffer (25 mM Tris pH 8.0 and 50 mM NaCl) and concentrated to ≥20 mg/ml. Proteins were quantified in dilution replicates ($n \geq 6$) using their respective molar extinction coefficients, based on their specific amino acid composition.[334]

Protein crystallization, data collection, and structure solution

To crystallize the E* conformation of UGDH(ΔID), the protein (10.4 mg/ml) was dialysed into 20 mM MES pH 5.6, 150 mM NaCl and crystallized at 20 °C using free interface diffusion in a 1.0-mm capillary containing 5 μl of 10.4 mg/ml enzyme and 200 μl of precipitant solution (100 mM MES pH 6.2, 100 mM MgCl2, and 16% PEG 3350). Crystals were cryoprotected in the precipitant solution supplemented with 18% glycerol and then plunged into liquid nitrogen. A 2.64

Å resolution dataset was collected on the 22-ID beamline (SER-CAT) at the Argonne National Laboratory using an MAR 300-mm CCD detector. The data were processed in space group C2 using XDS[335], and 5% of the data were set aside for cross-validation[336]. The crystal parameters and data collection statistics are summarized in Table S2.1. The structure was solved by molecular replacement using the PHENIX software suite[337] and human UGDH (Protein Data Bank (PDB) entry: 3TF5) as a search model. The structure was then subjected to iterative cycles of manual rebuilding using Coot[338] and automated refinement using PHENIX with both NCS restraints.[337,339] *B*-factors were refined using TLS as implemented in PHENIX. Refinement statistics[340,341] are summarized in Table S2.1.

The E^{Ω} UGDH(FL) was crystallized in the presence of 5 mM UDP-Xyl and 10 mM adenosine diphosphate at 25 °C using the hanging drop vapour diffusion method. One microlitre of protein was mixed in a 1:1 ratio with reservoir solution (0.1 M HEPES pH 7.2, 14% 1,6-hexanediol, and 10% PEG 3350). Crystals were cryoprotected in the precipitant solution supplemented with 20% glycerol and then plunged into liquid nitrogen. A 2.0 Å resolution dataset was collected on the 21-ID beamline (SER-CAT) at the Argonne National Laboratory using a MAR 300-mm CCD detector. The dataset was processed using XDS[335] and 5% of the data were set aside for cross validation[336]. The data collection statistics are listed in Table S2.1. The E^{Ω} UGDH(FL) structure was solved by molecular replacement using the PDB entry 2Q3E as a search model in PHENIX[337], and refined as described above. Refinement statistics[340,341] are summarized in Table S2.1.

Steady-state kinetics

All steady-state kinetic assays were conducted as previously described.[318,320-322,324,325] In brief, assays contained either 100 nM UGDH (FL, FL-A136M, ΔID, ΔID-A136M, R1, R2, −Pro,

−Lys, 0.13×FL, 0.26×FL, 0.5×FL, 2×FL, 0.13×Ser, 0.26×Ser, 0.5×Ser, Ser, 0.13×R1, 0.26×R1 or 0.5×R1) or 500 nM UGDH (FL-dimer, ΔID-dimer) in a standard reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA) with either saturating amounts of NAD^+ or UDP-Glc (Sigma). Substrate and enzyme were incubated separately at 25 °C for 5 min, and then reactions were initiated by rapid mixing of both solutions. Progress curves were obtained by continuously monitoring NADH production at 340 nm (molar absorptivity coefficient of 6,220 $M^{-1}\bullet cm^{-1}$) on an Agilent 8453 UV/Vis spectrometer equipped with a Peltier temperature controller (25 °C). UGDH progress curves display hysteresis, thus the observed initial velocity (v_i) represents a transient and does not satisfy steady-state conditions. To obtain steady-state initial velocities (v_{ss}), progress curves before the depletion of 10% substrate were fit to Frieden's equation[342] as in previous studies[318,320,342]:

$$P(t) = v_{ss}t - \tau(v_{ss} - v_i)\left(1 - e^{-\frac{t}{\tau}}\right) \qquad \text{(equation 4)}$$

where P is the product produced at time t, τ is the relaxation time of the lag, and the length of the lag is $e\tau$. The v_{ss} was used for determination of UGDH steady-state kinetic parameters. Data were fit using nonlinear regression analysis in PRISM (GraphPad Software).

Because the UGDH(FL-A136M), UGDH(ΔID-A136M), UGDH(FL-dimer) and UGDH(ΔID-dimer) constructs do not exhibit hysteresis, the observed initial velocity was used for the determination of steady-state parameters as previously described.[318] UDP-Glc substrate saturation curves were fit to equation (5).

$$v_\circ = \frac{k_{cat}[E_t][S]}{K_M + [S]} \qquad \text{(equation 5)}$$

where v_0 is the initial steady state velocity (vss in equation (4)), E_t and S are the enzyme and substrate concentrations, respectively. As previously reported[318,320,324], the NAD^+ saturation curves

of the UGDH hexameric enzyme display negative cooperativity and were fit to the sigmoidal rate equation (equation (6)):

$$v_\circ = \frac{k_{\text{cat}}[E_t][S]^h}{(K_{0.5})^h + [S]^h} \qquad \text{(equation 6)}$$

where $K_{0.5}$ is the half saturation point and h represents the Hill coefficient. The determination of the K_i for the allosteric inhibitor UDP-Xyl has been previously described.[318,320] In brief, data were globally fit to the model for competitive inhibition with cooperativity (equation (7)) using PRISM.

$$v_\circ = \frac{k_{\text{cat}}[E_t][S]^h}{(K_M^{\text{app}})^h + [S]^h} \text{ where } K_M^{\text{app}} = K_M\left(1 + \frac{[I]}{K_i}\right) \qquad \text{(equation 7)}$$

K_M, k_{cat}, and K_i were shared parameters in global fitting, whereas h was fit locally to each curve. The UGDH dimers (UGDH(FL-dimer) and UGDH(ΔID-dimer)) exhibited mixed inhibition with respect to both UDP-Glc and NAD^+, and were globally fit to equation (8).

$$v_\circ = \frac{(k_{\text{cat}}^{\text{app}})[E_t][S]}{(K_M^{\text{app}}) + [S]} \text{ where } k_{\text{cat}}^{\text{app}} = \frac{k_{\text{cat}}}{\left(1 + \frac{[I]}{(\alpha K_I)}\right)} \text{ and } K_M^{\text{app}} = K_M \frac{\left(1 + \frac{[I]}{K_i}\right)}{\left(1 + \frac{[I]}{\alpha K_I}\right)} \qquad \text{(equation 8)}$$

Here, K_i refers the competitive inhibition component, and αK_i gives the noncompetitive contribution. K_M, kcat, α and K_i were shared parameters for global fitting.

Sedimentation velocity

Sedimentation velocity analysis was conducted as previously described.[318,320-322,324,325] In brief, UGDH constructs were dialysed for >12 h at 4 °C into 25 mM HEPES pH 7.5 and 150 mM KCl and diluted to a final concentration of 9 μM. In ligand-bound studies, UGDH constructs were dialysed with comparable amounts of either substrate (UDP-Glc) or allosteric inhibitor (UDP-Xyl) for >24 h. Samples were loaded into cells equipped with 12-mm double-sector Epon centrepieces and quartz windows. The cells were then loaded into an An60 Ti rotor and equilibrated to 20 °C for 1 h. Sedimentation velocity data were collected at 50,000 r.p.m. in an Optima XLA analytical ultracentrifuge for 8–12 h. Data were recorded at 280 nm in radial step sizes of 0.003 cm.

SEDNTERP[343] was used to estimate the partial specific volume of all UGDH constructs, and the buffer density (1.00726 g/ml) and viscosity (0.01018 P). SEDFIT[344] was used to model and fit all data. Data were modelled as a continuous sedimentation coefficient (*c(s)*) distribution. The baseline, meniscus, frictional coefficient, and systematic time-invariant, and radial invariant noise were fit.[345] HYDROPRO[346] was used to predict s values based on crystal structures. The expected drag from the ID-tail was estimated by calculating the expected s values from crystal structures with and without modelled, energy minimized ID-tails. The data fits for all experiments can be found in Figure S2.3.

Evolutionary rate analysis

Seventy-nine UGDH sequences from vertebrates were used for analysis after removing redundancy at the organism level (only one UGDH sequence used per organism). The protein sequences were aligned using MUSCLE[347], and rates of evolution at each alignment position was calculated under the JTT model[348] using MEGA7 (log-likelihood method)[349]. The rates were normalized such that the average rate of evolution was 1.0 across the entire protein. Residue positions evolving faster than average show a rate greater than 1.0. In Figure S2.4, only the rates at alignment positions where the human UGDH did not have an indel were used.

Monte Carlo sampling

The free-energy cost of tethering an unstructured, non-interacting peptide to an impenetrable surface depends on the ratio of all constrained and unconstrained states:

$$\Delta G_{\text{constrained}} = -RT\, ln\left(\frac{\Omega_2}{\Omega_1}\right) \qquad \text{(equation 1)}$$

Where R is the gas constant, T is temperature, Ω_1 is the number of all possible states of an unconstrained, self-avoiding peptide and Ω_2 is the number of the Ω_1 states that do not conflict with the constraint imposed by the protein surface. To simplify, we used polyserine peptides, ignored

side-chain entropy and used a hard sphere potential along with 166 coarse-grained ϕ, ψ bins to calculate Ω_1 and Ω_2. Each bin represents a $10 \times 10°$ range of ϕ, ψ values of peptide conformations in the 'allowed' region of the original Ramachandran map (Figure S2.5a, b). This calculation is nontrivial for large polymers, and an exhaustive grid search of all conformations was only conducted for the 3- and 4-residue ID-tails (Figure S2.5c). We used the following Monte Carlo procedure to estimate the fraction of surface-constrained conformations (Ω_2/Ω_1) for each ID-tail. To determine the self-avoiding Ω_1 mesostates, we randomly assigned one of the 166 ϕ, ψ bins to each ϕ, ψ torsion angle in the ID-tail and then looked for steric clashes within the conformer using the 'outer limit' for atomic clashes as described in the original Ramachandran map.[350] Next, each of Ω1 mesostates was analysed for steric clashes with the surface or the adjacent ID-tail (Figure S2.5d–l). Prior to the simulation, hydrogens were added to the hexamer structure using the 'reduce' program[351], and an adjacent ID-tail was modelled in an extended conformation and fixed during the simulation (Figure S2.5d–f). The simulation was stopped when a minimum of 124,000 self-avoiding conformers were analysed and the ratio of surface-constrained conformations (Ω_2/Ω_1) reached convergence (Figure S2.5c). The convergence threshold was defined as a change in the cumulative ratio of less than 10−5 within a window of 5,000 trials. All runs reached convergence except for the 10-mer simulations, which only converged to 2 decimal places (Figure S2.5c–l). We estimated the accuracy in our Monte Carlo simulations by comparing the results to the full grid search of the 3- and 4-residue ID-tails (Figure S2.5c).

Thermodynamic shift assay

Solutions of UGDH (FL-dimer or ΔID-dimer) at 0.1 mg/ml were prepared with 5× SYPRO Orange ThermoFluor (Thermo Fisher) in the standard reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA). Samples were then briefly spun and allowed to equilibrate for 20

min. The thermal denaturation experiments were conducted in replicates ($n \geq 3$) and data were acquired using a Bio-Rad MiniOpticon Real-Time qPCR machine. A fluorescence excitation spectrum wavelength between 470–505 nm and an emission spectrum between 540–570 nm were used. The fluorescence emission for each solution was recorded every 30 s as the temperature was increased from 25 to 80 °C (ramp speed of 0.5 °C/s). Baselines were subtracted from the raw data using the buffer control experiments. The baseline, plateau and slope of the denaturation curve were fit to equation (9) to obtain the apparent T_m (melting temperature) values.[352]

$$Y = Baseline + \frac{Plateau - Baseline}{1 + 10^{\frac{T_m - X}{Slope}}} \qquad \text{(equation 9)}$$

where Y represents the fluorescence signal at temperature X.

Hydrogen–deuterium exchange–mass spectrometry

Studies have shown that hydrogen–deuterium exchange (HDX) is an appropriate probe for protein dynamics and can illuminate differences between wild-type and mutant proteins.[353,354] HDX is a powerful tool for footprinting the solvent-accessible regions of a protein[355], and was used in this study to compare structural and dynamic changes between the dimerized versions of UGDH (UGDH(FL-dimer) and UGDH(ΔID-dimer)).

Proteins were expressed and purified in the Wood laboratory as previously described.[318,320-322,324,325] Proteins were then flash-frozen and shipped overnight on dry ice to the Gross laboratory at Washington University in St. Louis for hydrogen–deuterium exchange–mass spectrometry (HDX–MS) analysis. Protein solutions (2 µl) were continuously labelled at 25 °C by adding 20 µl of 10 mM HEPES buffer containing 99.9% deuterium oxide (pD = 7.4). Samples were quenched by adding 33 µl of 8 M guanidine hydrochloride and 100 mM TCEP (final pH = 3.0) at 30 s, 1 min, 2 min, 15 min, 1 h, and 2 h time points[356,357]. One minute after quenching, samples were flash-frozen in liquid nitrogen and stored for less than 36 h at −80 °C. Control samples contained

10 mM HEPES in water rather than deuterium oxide. Each sample was thawed seconds before liquid chromatography followed by mass spectrometry (LC–MS). On-line protein digestion was performed with a custom-packed pepsin column (2 mm × 20 mm) at a flow rate of 200 µl/min in 0.1% trifluoroacetic acid. For desalting, a Zorbax Eclipse XDB-C-8 trap column (2.1 × 15mm, 3.5 µm) was used to trap peptic peptides for 3 min. Following this, peptides were separated using a Hypersil Gold C-18 analytical column (2.1 × 50 mm, 2.5 µm), 4–80% gradient of acetonitrile with 0.1% formic acid (B), and a 100 µl/min flow rate. Peptides were detected using a LTQ XL Orbitrap mass spectrometer (Thermo Fisher Scientific), with a mass resolving power of 50,000, m/z 400. Additional parameters were spray voltage of 5 kV, capillary temperature of 275 °C, capillary voltage of 49 V, and a tube lens of 163 V. All experiments were conducted in quadruplicate.

As a prelude to HDX, protein mapping was performed by identifying pepsin-digested peptides. Product-ion mass spectra were collected in the data-dependent mode, picking the six most abundant ions from selected MS/MS. Peptides were identified using Mascot (Matrix Science). Following HDX, mass spectra were analysed with HDX Examiner (Sierra Analytics). The per cent deuterium uptake was plotted against time for UGDH(FL-dimer) and UGDH(ΔID-dimer). To magnify slight, yet significant changes in uptake, the cumulative differences in HDX for UGDH(FL-dimer) versus UGDH(ΔID-dimer) were calculated. These values were plotted alongside 3 times the error propagation for all measurements of both variants for each peptide, after the data and error were normalized—divided by the number of time points considered for each data point (Figure S2.6). The propagation error for each peptide is equal to the square root of the sum of all squared standard deviation values for collective time-dependent measurements of UGDH(FL-dimer) and UGDH(ΔID-dimer). The cumulative per cent deuterium uptake was compared to 3 times the propagation error. Differences that were greater than 3 times the

propagation error were noted as regions of change affected by the presence of the ID-tail. We chose to normalize the data to be more inclusive of peptides with low intensity that are found at most time points. In a similar manner, we have excluded those peptides that have avoided detection for more than two time points.

Stopped-flow analysis of UGDH hysteresis

The allosteric activation (E* to E) of UGDH can be observed as a lag (hysteresis) in progress curves[318,320] (see Figure S2.7f for examples). The allosteric activation rates for UGDH(FL) ($n \geq 6$) and UGDH(ΔID) ($n \geq 6$) were monitored at 25 °C using an Applied Photophysics SX20 stopped-flow spectrophotometer. Enzyme solutions contained 500 nM UGDH(FL) or UGDH(ΔID) in the standard reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA). This solution was rapidly mixed with an equal volume of standard reaction buffer that contained both substrate and cofactor. The mixed solution contained 250 nM UGDH(FL) or UGDH(ΔID), with saturating amounts of both substrate and cofactor. The progress of the reaction was monitored by NADH production, with the absorbance reading at 340 nm being acquired every 10–15 ms. Progress curves were fit to equation (4) to determine the length of the lag in enzyme activation (E* to E). The mean and standard deviation of the hysteretic lags were derived from 6 or more progress curves.

Transient-state kinetics of UDP-Xyl binding

Stopped-flow fluorescence studies were conducted at 25 °C using an Applied Photophysics SX20 stopped-flow spectrophotometer with a dead time of ~1.2 ms. Syringes were loaded with 500 nM UGDH(FL) or UGDH(ΔID) and variable concentrations of UDP-Xyl, and then rapidly mixed. The change in intrinsic tryptophan fluorescence was continuously monitored using an excitation wavelength of 290 nm and an emission filter with a cut-off below 320 nm (Figure S2.7).

Fluorescence decay curves were averaged from experimental replicates ($n \geq 4$) for each concentration in the series. Raw data was corrected for the inner-filter effect using the molar absorptivity at both the excitation and emission of UDP-Xyl.[358] Data were globally fit using computer simulation with KinTek Global Kinetic Explorer program[359,360] (KinTek). Multiple input models based on the known structural states were tested, and the best fit model was determined using confidence contour analysis[361]. Microscopic rate constants and errors are reported in Figure S2.7e. Fit data and confidence contours can be found in Figure S2.7a–d.

2.4 Supplementary Information

Section 1. hUGDH Dimer [FL and ΔID] Mixed Inhibition

In contrast to hexameric hUGDH, both hUGDHFL-dimer and hUGDHΔID-dimer display mixed inhibition kinetics with UDP-Xyl (Figure S2.2). Briefly, the mixed inhibition occurs because UDP-Xyl binds strongly to the allosteric/substrate-binding site, but can also bind weakly to the coenzyme-binding site. This occurs in both the dimers and the native hexamer (see Figure S2.1e-l). However, we do not observe mixed inhibition in the native hexamer under steady state conditions because the coenzyme binds with a K_d of 30 uM[324] and our inhibition studies contain a saturating concentration of NAD^+ (6 mM). This means that the low concentration of UDP-Xyl (<5 μM) in these assays cannot outcompete NAD^+ binding in the coenzyme site. In contrast, the NAD^+ K_M of the dimers is roughly two orders of magnitude weaker than that of the hexamer, which makes it difficult to saturate the enzyme (Table S2.3). Thus, in the dimer we observe mixed inhibition, as we are able to detect both the tight binding of UDP-Xyl at the allosteric site and the much weaker binding at the coenzyme site (Tables S2.2 and S2.3). The $\alpha > 1$ shows that the mixed inhibition is dominated by UDP-Xyl competing with UDP-Glc for the allosteric site. Using NAD^+ as the

competing substrate, we confirmed that the mixed inhibition was a result of UDP-Xyl weakly binding to the coenzyme binding site in the $hUGDH_{dimers}$ (Figure S2.2 and Table S2.3).

2.5 Acknowledgements

The authors thank A. P. Karplus, B. W. Matthews, S. N. Savvides and the members of the Z.A.W. laboratory for helpful discussions. We also thank R. Wang of Norclone for producing the R1 truncation constructs. This work was supported by the NIH National Institute of General Medicine grants R01GM114298 awarded to Z.A.W. and P41GM103422 awarded to M.L.G.

2.6 Main Text Figures

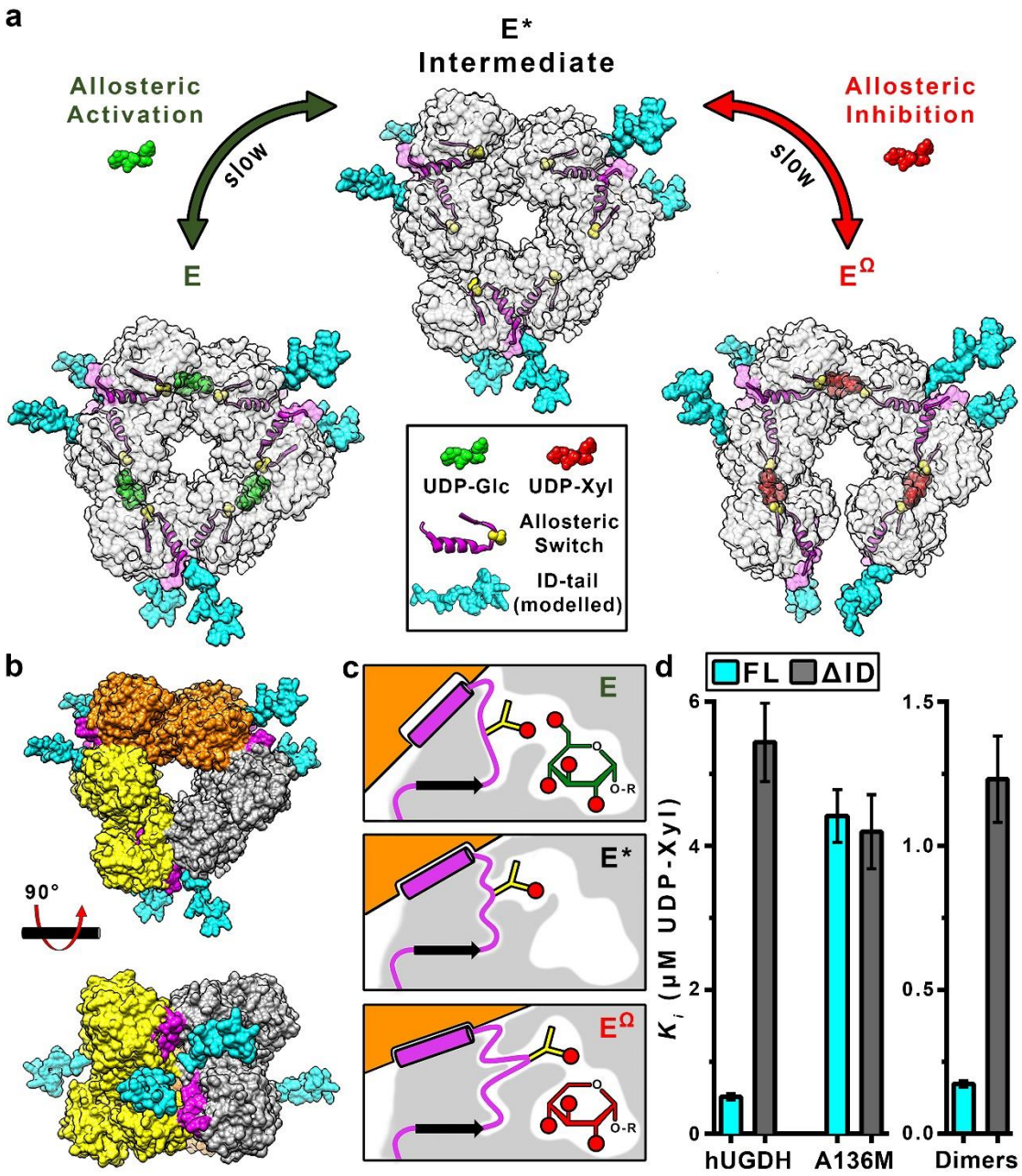


Figure 2.1: The role of the ID-tail in allosteric inhibition of UGDH. a, Unliganded UGDH forms an inactive (E*) hexamer. UDP-Glc (green) induces the Thr131-loop–α6 allosteric switch (yellow spheres and magenta ribbons and surface) to slowly isomerize into the active (E) state. UDP-Xyl (red) competes with UDP-Glc for the active site, and induces the allosteric switch to slowly isomerize into the inhibited (E^{Ω}) state. The slow isomerizations are due to the repacking of the allosteric switch in the protein core.[318,321,322,325] Because the ID-tail is disordered in the E, E* and E^{Ω} states (Figure S2.1 and refs[318,320-322]), we have modelled energy-minimized conformations

of the ID-tail (cyan) onto the structures of UGDH to depict the proximity to the active site, hexamer-building interface and the allosteric switch. **b,** Top and side view of the UGDH E* hexamer that forms from the association of three dimers (orange, grey and yellow).[318,320-322,324,325] The ID-tail of each dimer is located near two allosteric switches in the hexamer-building interface. **c,** The allosteric switch (magenta) is buried in the protein core (grey shading), which changes conformation in the E, E* and E^{Ω} states. Thr131 (yellow sticks) responds to the presence or absence of the C6′ OH in UDP-Glc (green) or UDP-Xyl (red), respectively. This response shifts the α6 helix (magenta cylinder) in the hexamer-building interface, which rotates the adjacent subunit (orange) to produce the E or E^{Ω} hexamer, as appropriate. Red circles depict hydroxyl (OH) groups. **d,** The UDP-Xyl affinity depends on the ID-tail and allostery. Data are the globally fit $K_i \pm$ s.e.m. derived from two or three independent rate curves with varying amounts of inhibitor ($n \geq 31$ independent data points; see Table S2.2 for specific values).

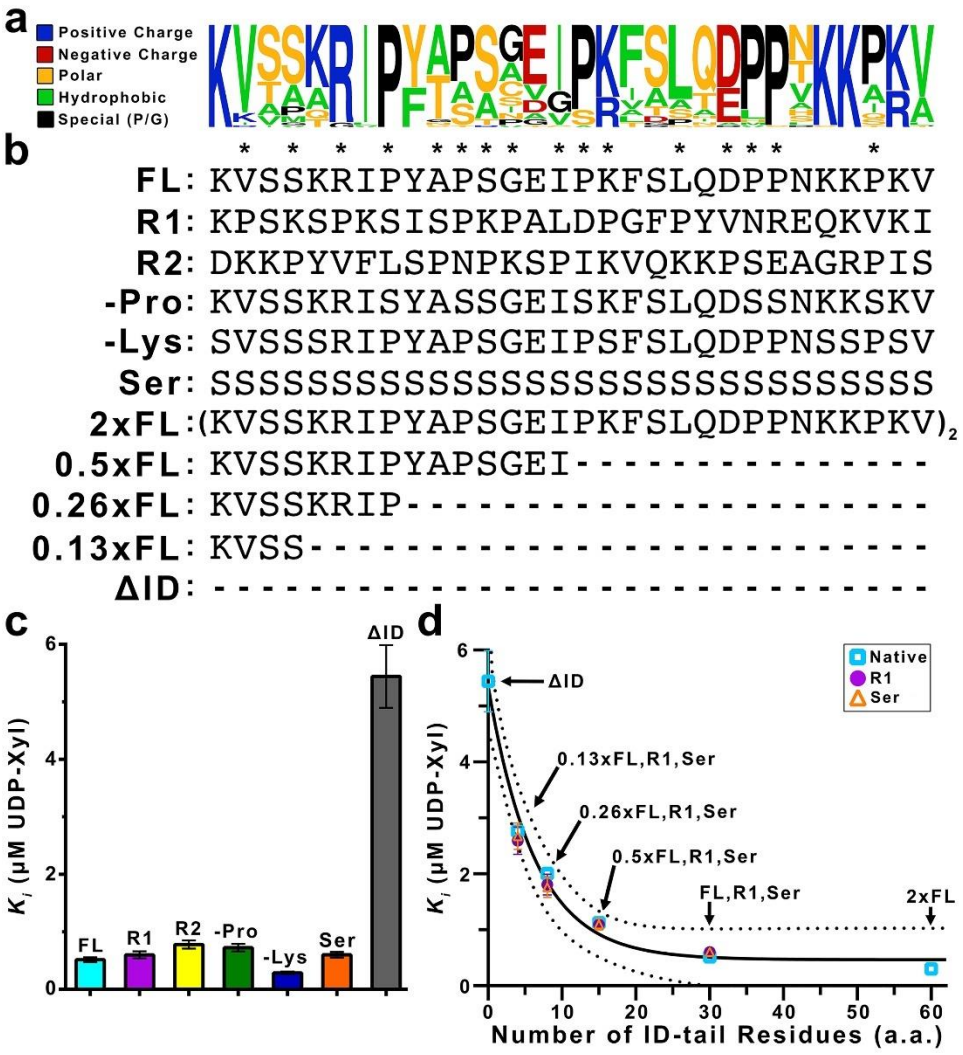


Figure 2.2: Structural constraints of the ID-tail. a, Alignments of the 30-residue ID-tail sequence (residues 465–494) from 79 vertebrate UGDHs (Figure S2.4a, b). Residues are coloured according to type, and the height of each residue represents the relative frequency. Alignments were generated using the WebLogo server (http://weblogo.berkeley.edu). **b,** Sequence modifications made to the primary structure of the ID-tail (Figure S2.4b). Asterisks indicate positions in the sequence that are sampled with a proline residue in either UGDH(FL), UGDH(R1) or UGDH(R2). **c,** UDP-Xyl affinity is independent of the ID-tail sequence. Data are the globally fit $K_i \pm$ s.e.m. derived from two or three independent rate curves with varying amounts of inhibitor. See Table S2.2 for the specific number of independent data points ($n \geq 27$). **d,** The affinity for UDP-Xyl depends on the length of the ID-tail. Data are the globally fit $K_i \pm$ s.e.m. derived from three independent rate curves with varying amounts of inhibitor ($n \geq 38$ independent data points; see Table S2.2 for specific values). For some points the s.e.m. is smaller than the data label. The data were fit to equation (3) (solid line) with 95% confidence intervals indicated (dashed lines). The fit predicts a maximum affinity of 0.46 ± 0.18 μM, corresponding to a free-energy change of -1.45 kcal•mol^{-1}.

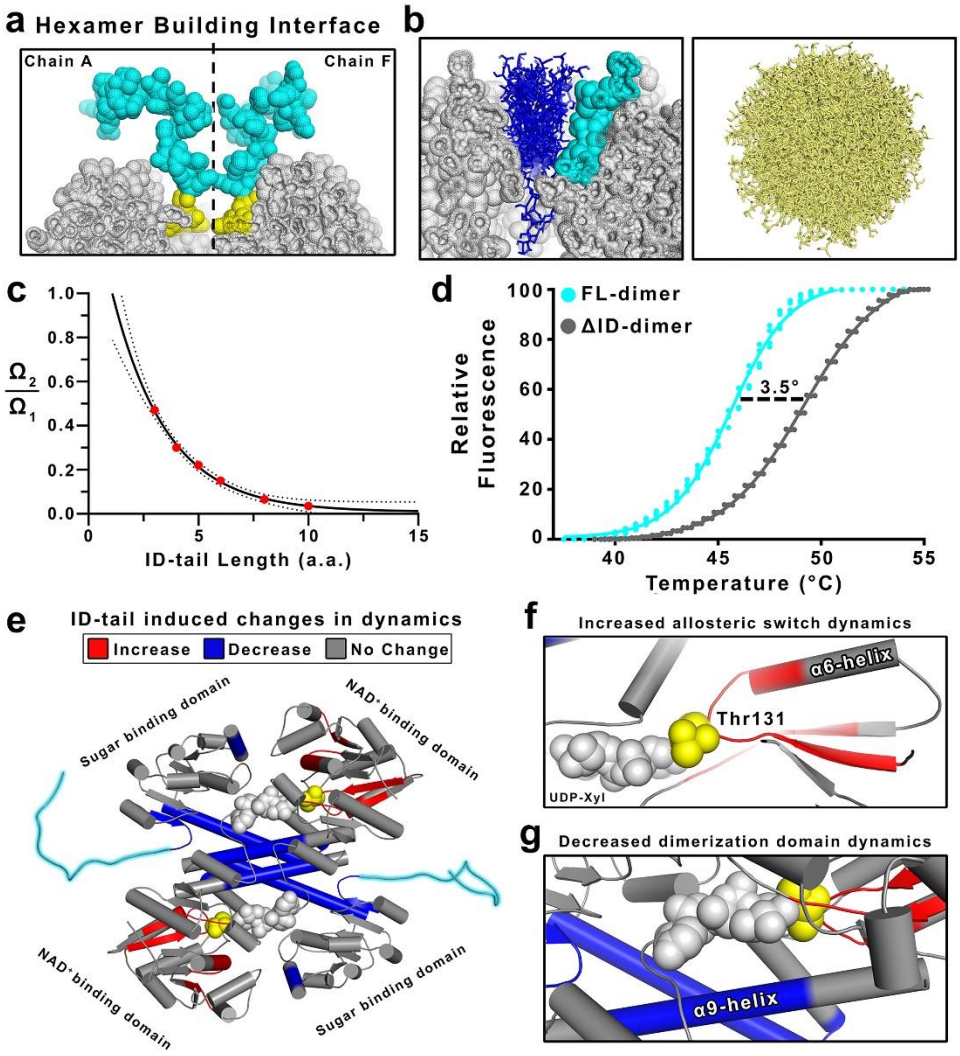


Figure 2.3: The entropic force of the ID-tail alters the structure of UGDH. a, Cut-away of the UGDH surface (grey spheres) at the hexamer-building interface (dashed lines), depicting the modelled ID-tails (cyan and yellow spheres) from adjacent subunits (grey, chains A and F). The volume-exclusion effect of the hexamer-building interface tightly constrains the conformations of the first four disordered residues (465–468) of the ID-tail (yellow). **b,** Left, a representative subset of the surface-constrained conformations of a 10-residue ID-tail (blue sticks) from Monte Carlo sampling (see Methods for details). The adjacent ID-tail is shown as cyan spheres. Right, a representative sampling (tan sticks) of accessible conformations without surface constraints (see also Figure S2.5). **c,** The fraction of constrained ID-tail conformations (Ω_2) over the possible conformations of a free ID-tail (Ω_1) exponentially converges with increasing ID-tail length. The data were fit to an exponential decay (Figure S2.5c). **d,** The ID-tail destabilizes UGDH by 3.5 °C. **e,** Comparing HDX rates of UGDH(FL-dimer) and UGDH(ΔID-dimer) shows that the ID-tail (cyan) alters the structure and dynamics of UGDH. Peptides displaying increases (red), decreases (blue) and no change (grey) in HDX rates are mapped to the structure. UDP-Xyl (grey spheres) was not used in the assay but is modelled in the active site. Thr131 of the allosteric switch is shown as yellow spheres. **f,** Close-up view of the allosteric switch (Thr131–α6 helix), which shows an increase in HDX rates. **g,** Close-up view of the of the dimerization domain, which is largely

inaccessible to solvent. Data shown in e–g were derived from the normalized cumulative per cent deuterium uptake (%D) comparing UGDH(FL-dimer) and UGDH(ΔID-dimer) (Figure S2.6).

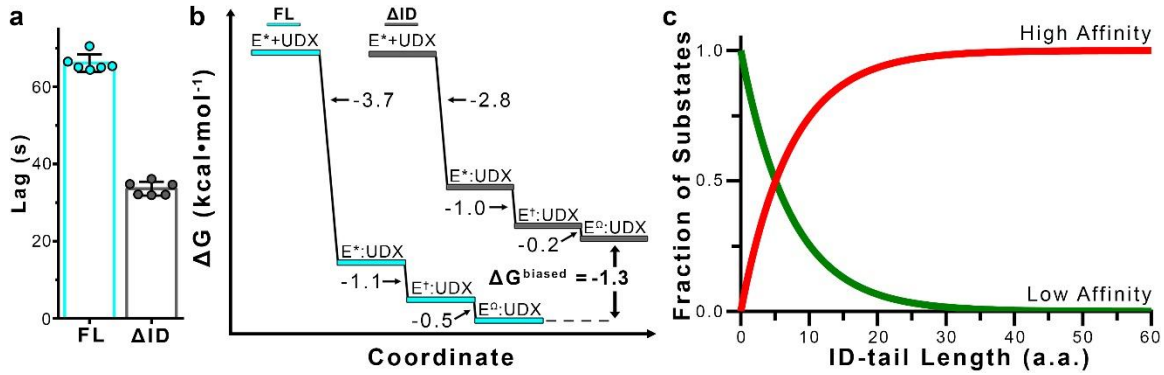


Figure 2.4: The ID-tail shapes the conformational landscape of UGDH. a, The ID-tail increases the hysteresis of allosteric activation (E* to E). Data are mean ± s.d. (n = 6 independent experiments). **b,** Free-energy plot of the transient-state kinetic model for the allosteric inhibition with UDP-Xyl (UDX) (equation (2)) of UGDH(FL) (E*FL) and UGDH(ΔID) (E*ΔID). The free energy for the initial binding step (second order) was determined using the standard state defined in units of mM, and the two isomerization steps (first order) were determined from the equilibrium constants K_1, K_2 and K_3 defined in Figure S2.7e. **c,** Plot of the high- and low-affinity fractional components of equation (3), showing that the shift in states is a function of length of the ID-tail.

2.7 Supplementary Figures and Tables

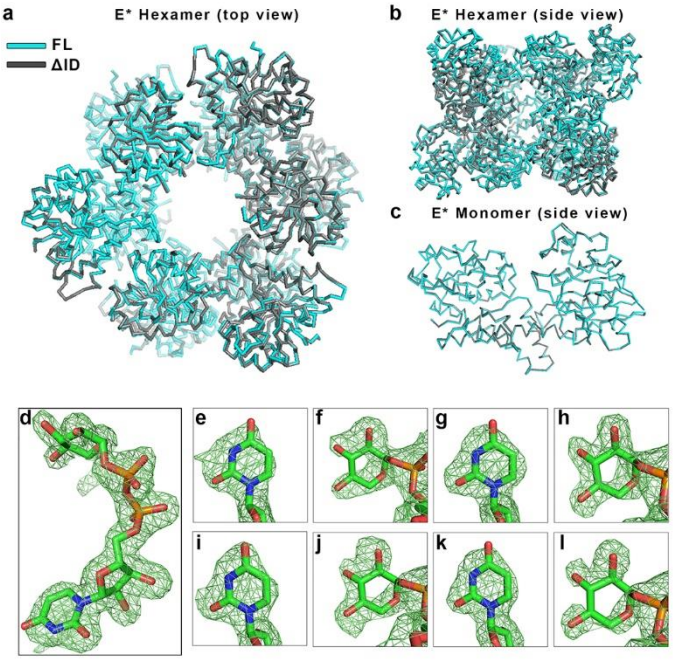


Figure S2.1: The crystal structures of UGDH(FL) and UGDH(ΔID) show no significant structural deviations, and structural evidence for UDP-Xyl binding in the NAD+ site. a–c, Structural overlay (root mean square deviation (r.m.s.d) = 0.385 Å), comparing the UGDH(FL) (cyan) and UGDH(ΔID) (grey) E* hexamers (a, b) and monomers (c). PDB entries for UGDH(FL) and UGDH(ΔID) are 4RJT and 5W4X, respectively (Table S2.1). **d,** Crystal structure of native UGDH with UDP-Xyl bound in the active site. Difference density map ($F_0 - F_c$) of UDP-Xyl (chain B) calculated at 2.0 Å resolution and contoured at 3.5σ. The map was calculated after omitting the UDP-Xyl and subjecting the model to simulated annealing. **e–l,** UDP-Xyl can also bind weakly to the NAD^+-binding site of native UGDH. Difference electron density maps ($F_0 - F_c$) were calculated as in d. The uracil and xylose in the NAD^+-binding sites were contoured at 3.5 and 3σ for chain A (e and f, respectively), chain B (g and h, respectively), chain D (i and j, respectively) and chain E (k and l, respectively). Chains C and F do not contain UDP-Xyl in the NAD^+-binding site. UDP-Xyl binding in the NAD^+ site is the source of mixed inhibition observed in the UGDH(FL-dimer) and UGDH(ΔID-dimer) constructs. (see Supplementary Information, Section 1). PDB entry: 5VR8 (this work, Table S2.1).

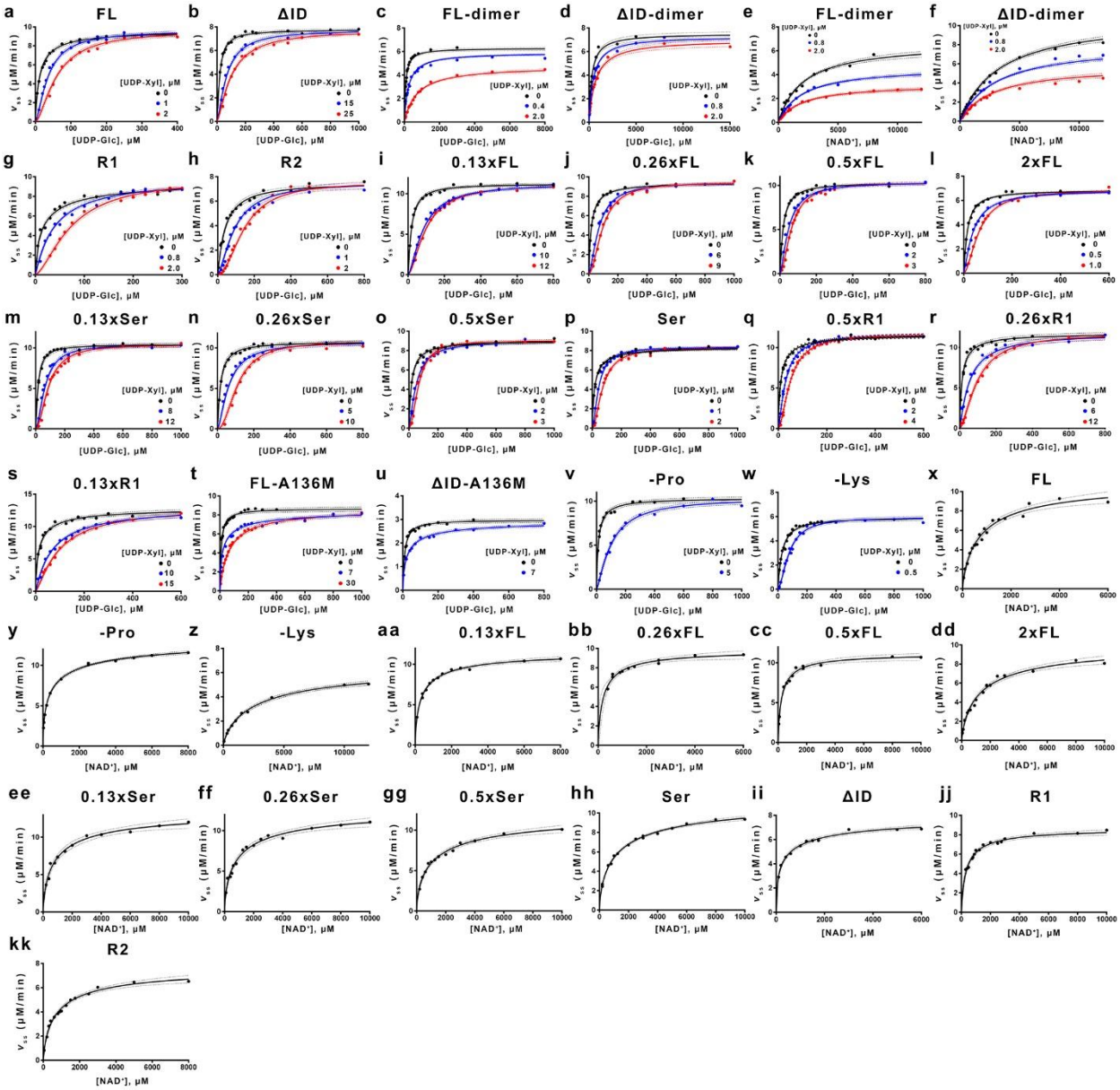
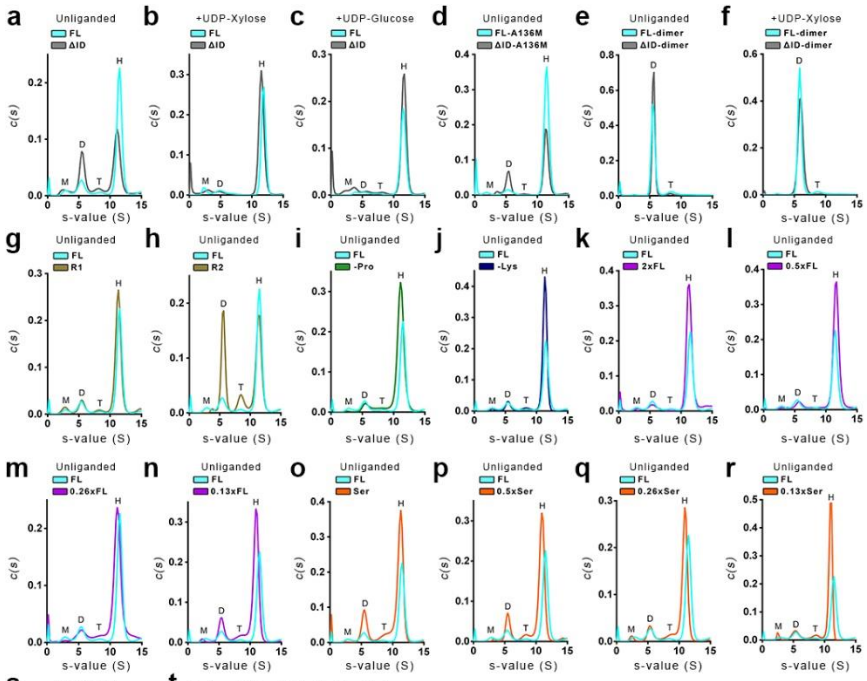


Figure S2.2: Steady-state kinetic analysis of all UGDH constructs. a–w, Inhibition studies with the allosteric inhibitor UDP-Xyl. Data from two or three independent rate curves were globally fit to equation (7) (or equation (8) for dimers c–f) using nonlinear regression ($n \geq 26$ data points). See Table S2.2 for the specific number of data points and fit parameters. Dashed lines indicate 95% confidence intervals. **x–kk,** NAD^+ substrate-saturation curves fit to equation (6) using nonlinear regression ($n \geq 10$ independent data points). See Table S2.3 for the specific number of data points used in global fitting.



t Sedimentation Velocity Results[a]

hUGDH	Ligand	M[b] S (%)	D[b] S (%)	T[b] S (%)	H[b] S (%)	Predicted S-Value	R.M.S.D[a]
FL	--------	3.0 (4.1)	5.5 (12.6)	8.5 (3.4)	11.5 (69.2)	----------	0.005113
FL	UDX	2.5 (5.5)	5.4 (8.6)	0.0 (0.0)	11.9 (70.9)	11.2[c]	0.003801
FL	UDG	0.0 (0.0)	5.7 (3.7)	0.0 (0.0)	11.5 (68.2)	11.7[c]	0.005448
ΔID	--------	2.9 (6.4)	5.6 (29.4)	8.3 (7.7)	11.1 (50.9)	----------	0.003784
ΔID	UDX	2.9 (6.0)	5.6 (5.0)	0.0 (0.0)	11.6 (78.3)	11.5[c]	0.008909
ΔID	UDG	3.3 (9.6)	5.9 (3.9)	8.3 (3.0)	11.6 (67.8)	12.0[c]	0.004379
FL-A136M	--------	0.0 (0.0)	5.4 (6.1)	0.0 (0.0)	11.5 (80.9)	----------	0.009808
ΔID-A136M	--------	3.8 (3.6)	5.4 (22.9)	8.0 (2.0)	11.4 (63.6)	----------	0.003788
FL-dimer	--------	2.5 (1.1)	5.6 (78.9)	8.8 (9.7)	0.0 (0.0)	5.6[d]	0.009960
FL-dimer	UDX	0.0 (0.0)	5.9 (88.6)	9.0 (5.8)	0.0 (0.0)	----------	0.004547
ΔID-dimer	--------	0.0 (0.0)	5.6 (90.9)	8.5 (4.0)	0.0 (0.0)	5.7[d]	0.004564
ΔID-dimer	UDX	0.0 (0.0)	6.0 (91.3)	0.0 (0.0)	0.0 (0.0)	----------	0.008205
R1	--------	2.9 (4.9)	5.6 (9.7)	8.3 (3.8)	11.3 (70.5)	----------	0.003979
R2	--------	3.9 (1.3)	5.6 (33.9)	8.6 (10.3)	11.4 (47.2)	----------	0.005785
-Pro	--------	0.0 (0.0)	5.8 (7.6)	8.0 (3.6)	11.1 (86.4)	----------	0.013871
-Lys	--------	3.0 (1.2)	5.5 (8.0)	8.4 (4.3)	11.3 (84.8)	----------	0.004136
2×FL	--------	3.1 (1.6)	5.7 (6.8)	0.0 (0.0)	11.4 (75.7)	----------	0.011126
0.5×FL	--------	3.4 (2.2)	5.8 (7.2)	8.4 (2.6)	11.7 (84.4)	----------	0.005328
0.26×FL	--------	0.0 (0.0)	5.5 (10.0)	0.0 (0.0)	11.3 (84.1)	----------	0.000772
0.13×FL	--------	2.4 (1.4)	5.5 (14.8)	0.0 (0.0)	10.9 (75.6)	----------	0.003486
Ser	--------	0.0 (0.0)	5.5 (17.4)	0.0 (0.0)	10.9 (77.7)	----------	0.004350
0.5×Ser	--------	3.4 (1.0)	5.6 (14.2)	8.3 (5.9)	10.9 (73.6)	----------	0.003647
0.26×Ser	--------	2.5 (2.4)	5.5 (10.8)	0.0 (0.0)	10.9 (77.8)	----------	0.003143
0.13×Ser	--------	2.5 (2.8)	5.5 (11.1)	8.5 (6.4)	10.9 (80.2)	----------	0.003144
K465D	--------	2.5 (4.1)	5.4 (21.7)	8.3 (2.8)	11.6 (64.5)	----------	0.008206

[a] Sedimentation values, relative distributions, and root-mean-square deviation (R.M.S.D) were derived from fit data to greater than 65 independent sedimentation profiles.
[b] M = Monomer; D = Dimer; T = Tetramer; H = Hexamer
[c,d] HYDROPRO predicted S value for the hexamer[c] and the dimers[d]

Figure S2.3: Sedimentation velocity studies of the UGDH constructs. a–s, Plots of the c(s) distributions with oligomeric species labelled as H (hexamer), T (tetramer), D (dimer) or M (monomer). The R2 mutant (panel h) shows no change in UDP-Xyl affinity (Figure 2.2c and Table S2.2), yet shows evidence of a less stable hexamer. Panel s was included to show that the hexamer in h is less stable partly owing to the K465D substitution in the UGDH(R2) construct. The K465D substitution introduces an unfavourable negative charge near E460 in the hexamer interface, which may reduce the stability. **t,** Relative distributions, s values (S) and r.m.s.d. values for all sedimentation velocity experiments.

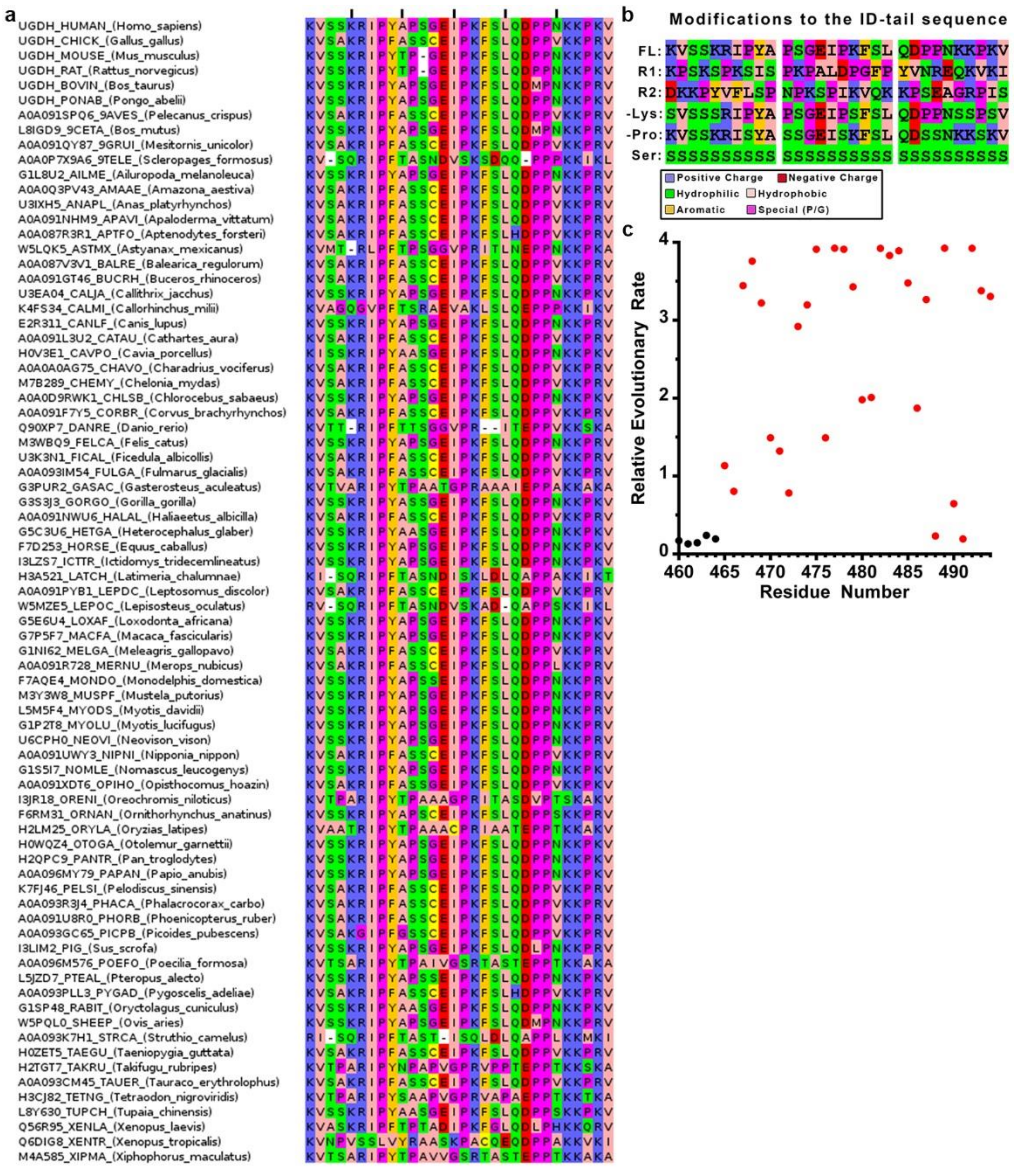
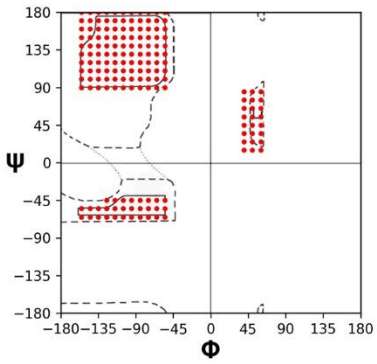


Figure S2.4: The ID-tail is conserved in vertebrates. a, ClustalO sequence alignment of all vertebrate UGDH ID-tail regions (79 total). Residues are coloured by type, where blue is positive charge (K, R, H), red is negative charge (D, E), peach is hydrophobic (A, V, L, I, M), orange is aromatic (F, W, Y), green is hydrophilic (S, T, N, Q), yellow is cysteine (C), and magenta is special (P, G). **b,** The ID-tail was extensively randomized and modified. Sequences of UGDH (FL, R1, R2, −Lys, −Pro, and Ser), aligned by position and coloured by residue type. **c,** Relative evolutionary rate of UGDH residues from the alignment of 79 vertebrate sequences. The ID-tail (red dots) begins at residue 465 and displays an approximately threefold higher rate of divergence than the folded portion of the protein (black dots). For clarity, only a small, representative segment of the folded protein is shown (residues 460–464). All rates were scaled such that the average rate is 1.0 across the entire dataset.

a



b Values for the Φ,Ψ ranges used in Monte Carlo Simulations

Allowed Regions	ϕ-range	ψ-range	Number of Conformations
I	-155° to -55°	90° to 180°	110
II	-125° to -55°	-65° to -45°	24
III	-155° to -125°	-65° to -55°	8
IV	40° to 60°	15° to 85°	24
			166 (total)

c Monte Carlo Simulation Results

Tail	Ω_2/Ω_1 (M.C.) [1]	Ω_1	Ω_2/Ω_1 (Exhaustive Search)[2]	$\Delta G_{constrained}$ (kcal•mol^{-1}) [3]
3	0.47	124075	0.52	0.446
4	0.30	124876	0.29	0.709
5	0.22	125697	-	0.905
6	0.15	124838	-	1.134
8	0.06	479271	-	1.612
10	0.03	150112	-	1.98

[1] Monte Carlo calculation (see text and methods). Ω_1 and Ω_2 are the numbers of unconstrained and surface-constrained conformations, respectively.
[2] The difference between the grid search and the MC results are a measure of accuracy of the latter
[3] *Equation 1* (see text)

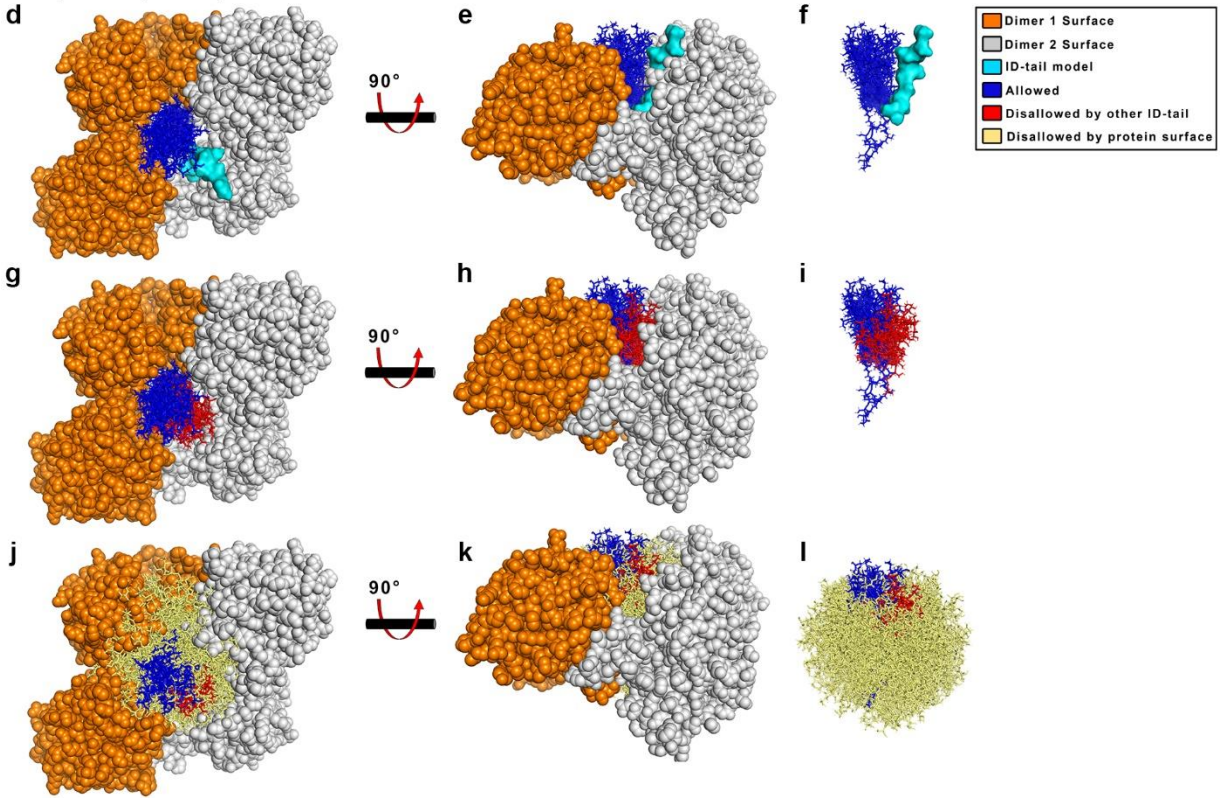


Figure S2.5: Exhaustive Monte Carlo simulations constraining the ID-tail. a, Dashed lines outline the traditional, generously allowed regions of the Ramachandran plot, whereas the red circles identify the conformations used in the Monte Carlo simulations. **b,** The ranges of ϕ and ψ

angles depicted in a. The 10 × 10° bins are centred on the first and last numbers in the range. For example, in region 1, the first ϕ, ψ bin (−155°, 90°) represents the ϕ range −155° to −145° and the ψ range 85° to 95°. **c,** Ratio of ID-tail conformations constrained (Ω_2) to the number of conformations when the ID-tail is unconstrained (Ω_1). The entropic costs of confining tails of each length were calculated using equation (1). **d-e** The results of the 10-residue ID-tail simulations, shown in a surface representing the hexamer-building interface (orange and grey dimers) with the adjacent ID-tail (cyan) that was fixed during simulations. Also depicted is a representative sampling of 20 allowed Ω_2 conformations (blue sticks) from the 4,503 identified in the Monte Carlo simulation. **f,** The same view as in e, but without the protein surface. **g–i,** Same as in d–f, but now including a sampling of 20 of the 3,002 Ω_1 conformations (red sticks) that clash with the fixed adjacent ID-tail (not depicted for clarity). **j–l**, Same as g–i, but including 750 of the 142,607 Ω_1 conformations (tan sticks) that clash with the protein surface.

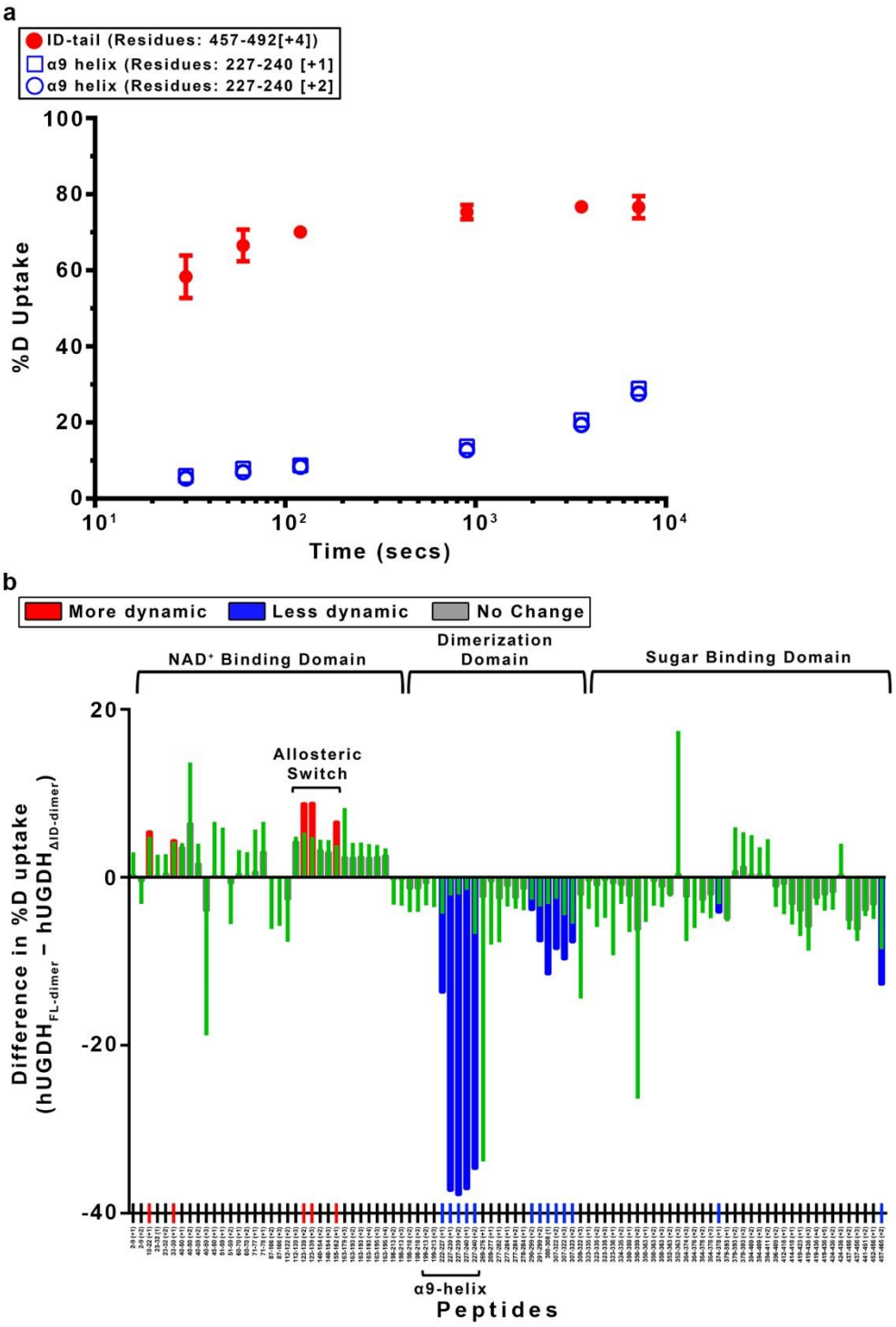
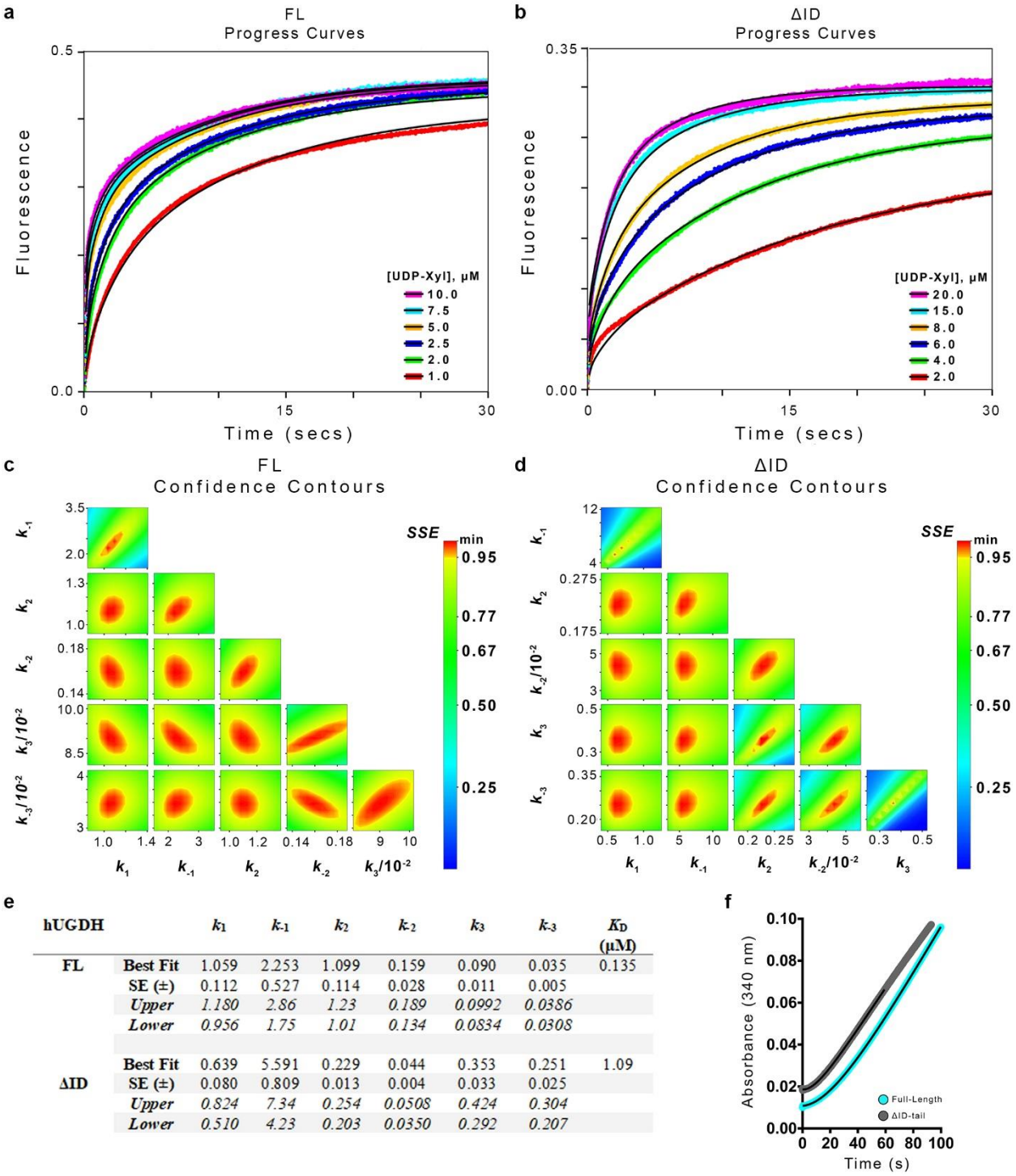


Figure S2.6: The ID-tail induces global changes in the structure and dynamics of UGDH. a, The per cent deuterium uptake of the ID-tail peptide region (residues 457–492; red closed circles) saturates rapidly, consistent with an unfolded peptide[331]. For comparison, two peptides corresponding to the well-ordered α9 helix region (open blue squares and circles) saturate slowly. Data are mean ± s.d. of independently replicated time points ($n = 4$). For some points, the standard deviation is less than the dimensions of the data symbol. **b,** The normalized cumulative changes in

the hydrogen–deuterium exchange rates (UGDH(FL-dimer) –UGDH(ΔID-dimer)). Most of the kinetics measurements consisted of six independently replicated time points (n = 4), processed to give the mean exchange (red, blue or green bars). Approximately 5% of the data displayed low signal:noise or was missing, and in those cases the means were derived from four or more time points. Results were normalized by dividing by the number of measurements. The propagation error for each peptide is equal to the square root of the sum of all squared standard deviation values for the collective measurement of UGDH(FL-dimer) and UGDH(ΔID-dimer).



e

hUGDH		k_1	k_{-1}	k_2	k_{-2}	k_3	k_{-3}	K_D (µM)
FL	Best Fit	1.059	2.253	1.099	0.159	0.090	0.035	0.135
	SE (±)	0.112	0.527	0.114	0.028	0.011	0.005	
	Upper	*1.180*	*2.86*	*1.23*	*0.189*	*0.0992*	*0.0386*	
	Lower	*0.956*	*1.75*	*1.01*	*0.134*	*0.0834*	*0.0308*	
ΔID	Best Fit	0.639	5.591	0.229	0.044	0.353	0.251	1.09
	SE (±)	0.080	0.809	0.013	0.004	0.033	0.025	
	Upper	*0.824*	*7.34*	*0.254*	*0.0508*	*0.424*	*0.304*	
	Lower	*0.510*	*4.23*	*0.203*	*0.0350*	*0.292*	*0.207*	

f

Figure S2.7: Transient-state analysis of UGDH(FL) and UGDH(ΔID). a, b, Transient-state analysis of UDP-Xyl binding kinetics using intrinsic protein fluorescence. Six independent progress curves (coloured traces) at different inhibitor concentrations were globally fit (black line) to the allosteric inhibition model (see Figure 2.4b) for UGDH(FL) and UGDH(ΔID). Each progress curve was replicated ($n \geq 4$) with similar results, and the final kinetic model was refined against the averaged progress curves (see e for fit parameters). **c, d,** Confidence contour plots

depicting how constrained each globally fit parameter is relative to the others, for all progress curves in a and b (parameters are listed in e). **e,** Table of the microscopic rate constants from global fitting of the progress curves described in a and b. The best fit and s.e.m. were obtained from global nonlinear regression based on the numerical integration of rate equations for the described model (see main text and Methods). Upper and lower limits were obtained from the confidence contour analysis. $K_d = (K_1K_2K_3)^{-1}$, where $K_n = k_n/k_{-n}$. **f,** Enzyme hysteresis is observed as a lag in progress curves. Representative progress curves (of $n = 6$ independent measurements) for both UGDH(FL) (cyan) and UGDH(ΔID) (grey) are fit to equation (4) (black line). Curves are displayed with the y axis offset for clarity. Final results for all replicate curves are displayed in Figure 2.4a.

Table S2.1: Data collection and refinement statistics

Data collection		
Protein Data Bank Entry	5W4X	5VR8
	E* $hUGDH_{\Delta ID}$	E^{Ω} $hUGDH_{FL}$
Space group	C2	$P12_11$
Unit cell dimensions a,b,c (β)	178.19, 114.07, 97.24 (116.9°)	89.08, 196.49, 111.26 (111.9°)
Completeness (%)	99.9 (91.1)[a]	93.2 (60.0)[a]
No. reflections	324,675	2,730,154
Redundancy	6.4 (6.1)	12.3 (10.3)
***I* / σ(*I*)**	21.9 (1.5)	14.9 (2.5)
$CC_{1/2}$[b]	99.9 (64.9)	99.7 (79.3)
R_{meas} (%)[c]	6.5 (122.5)	13.2 (89.3)
Refinement		
Resolution (Å)	2.65	2.00
***R*work / *R*free**	0.19 / 0.23	0.16 / 0.19
No. atoms: Protein / Ligand / Water	10887 / 33 / 36	21584 / 394 / 1097
***B*-factors (Å²): Protein / Ligand / Water**	89.9 / 97.4 / 64.3	33.2 / 27.1 / 32.3
Stereochemical Ideality		
Bond lengths (Å²)	0.004	0.008
Bond angles (°)	0.75	0.91
φ,ψ Preferred (%)[d]	98.98	97.8
φ,ψ Additionally allowed (%)	1.02	2.2
φ,ψ Disallowed region (%)	0.0	0.0

[a] Values in parenthesis are for the highest-resolution shell (2.71–2.64 and 2.0221–1.9994 for 5W4X and 5VR8, respectively).
[b] $CC_{1/2}$ is the percentage of correlation between intensities from random half-data sets.[341]
[c] R_{meas} is the redundancy-independent merging R factor.[340]

Table S2.2: Kinetic parameters of all UGDH constructs[a]

hUGDH	K_M (UDP-Glc, μM)	k_{cat}[b] (s^{-1})	K_I^{UDX} (UDP-Xyl, μM)	α_{UDG}[c]	Δ ΔG[e] (kcal•mol^{-1})	# of Data Points[h]
ΔID	17.8 ± 0.9	0.7 ± 0.01	5.44 ± 0.55	------	0.00	42
FL	12.7 ± 0.6	0.8 ± 0.01	0.52 ± 0.04	------	-1.39	38
R1	12.9 ± 1.0	0.8 ± 0.01	0.60 ± 0.06	------	-1.31	59
0.13×R1	12.8 ± 1.2	1.0 ± 0.01	2.59 ± 0.24	------	-0.44	40
0.26xR1	12.4 ± 1.0	1.0 ± 0.01	1.81 ± 0.18	------	-0.65	42
0.5xR1	11.1 ± 0.8	1.0 ± 0.01	1.09 ± 0.08	------	-0.95	47
R2	43.7 ± 3.6	0.7 ± 0.01	0.78 ± 0.07	------	-1.15	50
-Lys	30.1 ± 1.9	0.5 ± 0.01	0.29 ± 0.03	------	-1.73	39
-Pro	13.1 ± 0.9	0.9 ± 0.01	0.72 ± 0.07	------	-1.20	26
0.13×FL	18.8 ± 0.9	1.0 ± 0.01	2.76 ± 0.15	------	-0.40	46
0.26×FL	18.3 ± 0.7	0.8 ± 0.01	1.99 ± 0.12	------	-0.60	42
0.5×FL	18.8 ± 0.9	0.9 ± 0.01	1.12 ± 0.08	------	-0.94	50
2×FL	15.2 ± 0.7	0.6 ± 0.01	0.30 ± 0.02	------	-1.72	43
0.13×Ser	16.9 ± 1.0	0.9 ± 0.01	2.67 ± 0.24	------	-0.42	49
0.26×Ser	18.4 ± 1.0	0.9 ± 0.01	1.76 ± 0.18	------	-0.67	43
0.5×Ser	17.4 ± 1.3	0.8 ± 0.01	1.09 ± 0.10	------	-0.95	49
Ser	17.8 ± 1.0	0.7 ± 0.01	0.60 ± 0.05	------	-1.31	53
ΔID-dimer	286 ± 27	0.1 ± 0.01	1.23 ± 0.15[d]	22 ± 12	0.00[f]	36
FL-dimer	83.2 ± 2.2	0.1 ± 0.01	0.17 ± 0.01[d]	36 ± 5	-1.17[f]	50
ΔID-A136M	9.9 ± 0.6	0.3 ± 0.01	4.20 ± 0.51	------	0.00[g]	30
FL-A136M	8.5 ± 0.6	0.7 ± 0.01	4.41 ± 0.37	------	0.03[g]	55

[a] Kinetic parameters and associated s.e.m. for all constructs were derived from global analyses of data in Figure S2.2.
[b] One catalytic turnover of UDP-GlcA produces two molecules of NADH per cycle.

[c] α describes the mode of mixed inhibition (equation (8)). An α >1 in the UDP-Glc saturation curves shows that UDP-Xyl binds preferentially to the allosteric binding site, and secondarily to the coenzyme-binding site.
[d] Competitive K_i from the fit to the mixed inhibition (equation (8)).
[e] Change in UDP-Xyl binding free energy (kcal•mol^{-1}) of UGDH constructs relative to UGDH(ΔID): $\Delta\Delta G = R \bullet T \bullet \ln((K_i \text{ Construct})/(K_i \Delta ID))$.
[f] Change in UDP-Xyl binding free energy relative to the UGDH(ΔID-dimer).
[g] Change in UDP-Xyl binding free energy relative to the UGDH(ΔID-A136M).
[h] The number of independent data points used in global analysis (see Methods).

Table S2.3: NAD^+ kinetic parameters for UGDH[a]

hUGDH	K_M (NAD^+, mM)	$K_{0.5}$[b] (NAD^+, mM)	Hill (h)	k_{cat}[c] (s^{-1})	UDX (K_i, μM)	α_{NAD}[d]	# of Data Points[e]
FL	----------	0.8 ± 0.20	0.8 ± 0.1	0.9 ± 0.08	----------	----------	18
ΔID	----------	0.3 ± 0.06	0.6 ± 0.1	0.7 ± 0.03	----------	----------	12
FL-dimer	2.0 ± 0.26	----------	----------	0.1 ± 0.01	2.1 ± 0.4	0.9 ± 0.2	37
ΔID-dimer	3.2 ± 0.10	----------	----------	0.2 ± 0.01	3.6 ± 0.8	0.6 ± 0.2	47
R1	----------	0.4 ± 0.03	0.9 ± 0.1	0.7 ± 0.01	----------	----------	17
R2	----------	0.8 ± 0.14	0.7 ± 0.1	0.7 ± 0.01	----------	----------	15
-Lys	----------	2.9 ± 0.61	0.8 ± 0.1	0.6 ± 0.04	----------	----------	10
-Pro	----------	0.5 ± 0.06	0.6 ± 0.1	1.2 ± 0.03	----------	----------	12
0.13×FL	----------	0.4 ± 0.03	0.7 ± 0.1	1.0 ± 0.03	----------	----------	13
0.26×FL	----------	0.2 ± 0.03	0.8 ± 0.2	0.8 ± 0.05	----------	----------	11
0.5×FL	----------	0.3 ± 0.03	0.9 ± 0.1	0.9 ± 0.02	----------	----------	12
2×FL	----------	1.4 ± 0.31	0.8 ± 0.1	0.9 ± 0.01	----------	----------	18
0.13×Ser	----------	0.9 ± 0.24	0.7 ± 0.1	1.2 ± 0.10	----------	----------	12
0.26×Ser	----------	1.0 ± 0.27	0.7 ± 0.1	1.1 ± 0.09	----------	----------	15
0.5×Ser	----------	1.2 ± 0.34	0.7 ± 0.1	1.1 ± 0.09	----------	----------	13
Ser	----------	1.3 ± 0.19	0.7 ± 0.1	1.0 ± 0.04	----------	----------	15

[a] Kinetic parameters and associated s.e.m. for all constructs were derived from global analyses of data in Figure S2.2.

[b] Hexameric UGDH displays negative cooperativity with NAD^+ binding, which indicates a mix of high-affinity and low-affinity sites.[318,320-322,324,325] In previous work, we showed that the native UGDH(FL) $K_{0.5}$ of 0.8 mM NAD^+ corresponds to a mix of high-affinity and low-affinity sites with K_M of 88 μM and 1.8 mM, respectively. This is consistent with the published K_d of 30 μM for the coenzyme.[324]

[c] One catalytic turnover of UDP-GlcA produces two molecules of NADH per cycle.

[d] α describes the mode of mixed inhibition (equation (8)). An α <1 in the NAD^+ saturation curves show that UDP-Xyl binds preferentially to the allosteric binding site, and secondarily to the coenzyme-binding site.

[e] The number of independent data points used in nonlinear regression (see Methods).

CHAPTER 3

THE CONSERVATION OF ATYPICAL ALLOSTERY IN *C. ELEGANS* UDP-GLUCOSE DEHYDROGENASE

Nicholas D. Keul, Tiffany N. Hicks Sirmans, Nathaniel R. Beattie, Weston E. McDonald, Trevor M. Talmadge, Rahil Taujale, Natarajan Kannan and Zachary A. Wood. (Submitted and in peer review at ACS Omega 2019).

3.1 Abstract

Human UDP-glucose dehydrogenase (hUGDH) oxidizes UDP-glucose to UDP-glucuronic acid, an essential substrate in the phase II metabolism of drugs. The activity of the hUGDH is controlled by an atypical allosteric mechanism in which the feedback inhibitor UDP-xylose competes with substrate for the active site and triggers a buried allosteric switch to produce an inactive complex (E^{Ω}). Previous comparisons with a non-allosteric UGDH identified six large-to-small substitutions (including F109A) that produce packing defects in the protein core that provide the conformational flexibility necessary for the allosteric transition. In fact, the A109L substitution has been shown to abolish allostery in UGDH. *Caenorhabditis elegans* UGDH (cUGDH) conserves the motif with the exception of an Ala-to-Pro substitution in position 109. Here we examine the impact of the proline substitution on allostery in cUGDH and test the hypothesis that these large-to-small substitutions form a motif that can used to identify allosteric UGDHs. The crystal structures of unliganded and UDP-xylose bound cUGDH show that the A109P substitution is accommodated by a co-evolved Asn-to-Ser substitution at position 290. The allosteric transition is conserved in cUGDH, but with a significant difference in the rotation of the allosteric switch. Allostery in cUGDH was confirmed using both steady-state and sedimentation velocity studies. The enzyme also exhibits hysteresis in progress curves, and negative cooperativity with respect to NAD^+ binding. Both of these phenomena are conserved in the human enzyme, which is strong evidence that these represent fundamental features of atypical allostery in UGDH. A phylogenetic analysis of UGDH shows that the atypical allostery motif is ancient, and identifies a potential transition point in the evolution of the UGDH family.

3.2 Introduction

The nucleotide sugar UDP-α-D-glucuronic acid (UDP-GlcA) is the essential substrate of glucuronidation, a major mechanism of drug metabolism in mammals.[362-364] Some cancers have been shown to exploit glucuronidation as a means of chemotherapeutic resistance.[365-368] Limiting the availability of UDP-GlcA represents a novel strategy for combatting this drug-resistance mechanism.[369-371] UDP-GlcA is produced by UDP-α-D-glucose 6-dehydrogenase (UGDH), which catalyzes the NAD^+-dependent oxidation of UDP-α-D-glucose (UDP-Glc).[317] In humans, hUGDH is a hexamer that is allosterically regulated by the downstream metabolite UDP-α-D-xylose (UDP-Xyl) (Figure 1A).[319,328] Understanding the allosteric mechanism of hUGDH is an important step towards the design of therapeutics that can reduce the cellular levels of UDP-GlcA.

The activity of hUGDH is regulated by an atypical allosteric mechanism[318,320,322,323,372]. Briefly, the downstream metabolite UDP-Xyl competes with substrate, and upon binding triggers the Thr131-loop/α6-helix (allosteric switch) to change conformation and produce an inactive horseshoe-shaped complex (E^{Ω}) (Figure 3.1A,B).[318,321-323,372] This allosteric transition can be observed as positive cooperativity in inhibition studies.[318,320,323,372] In the E^{Ω} conformation, the Thr131-loop/α6-helix increases the affinity between adjacent subunits to produce a more stable hexamer, consistent with an allosteric transition.[320,321] In the absence of any ligand, the three dimers of hUGDH form a weakly associated hexamer called E* (Figure 3.1A,B).[318,320-325,372] The E* hexamer slowly isomerizes to the active E state upon binding substrate, which produces a lag in progress curves (hysteresis).[320]

The Thr131-loop of the allosteric switch is buried in the protein core, and must change conformation for the enzyme to transition between the E and E^{Ω} states (Figure 3.1C). The remarkable plasticity of the protein core originates from packing defects (cavities and deep surface

pockets) that provide flexibility and space for the allosteric transition (Figure 3.1C).[318,321,323] Comparing hUGDH to the non-allosteric homolog from *Streptococcus pyogenes* (spUGDH), it was shown that the packing defects evolved from six large-to-small residue substitutions in the protein core (Figure 3.1D).[321] It was hypothesized that these substitutions could be used as a sequence motif to identify other UGDHs that exhibit atypical allostery.[321] Here we have tested that hypothesis using *Caenorhabditis elegans* UGDH (cUGDH), which shares 61% sequence identity with hUGDH, but has a single substitution in one of the allosteric motif residues (A109P) (Figure 3.1D and 3.2). In the human enzyme, we have previously shown that the A109L substitution in this motif abolishes allostery[323], which suggests that cUGDH may not be allosteric. Here we show that cUGDH is not only allosteric, but that it also conserves hysteresis. In addition, we observe negative cooperativity with respect to NAD^+ binding, which has also been reported in the human enzyme.[318,323,324,372] The conservation of both hysteresis and negative cooperativity is strong evidence that these phenomena are defining characteristics of atypical allostery in UGDH. By combining these structural and kinetic studies with a phyologenetic analysis, we shed light on the evolution of atypical allostery in UGDH.

3.3 Materials and methods

Protein expression and purification

The enzymes cUGDH (UniProt ID: Q19905-1) and hUGDH (UniProt ID: O60701-1) were recombinantly expressed and purified as previously described.[318,320-322,324,325,372,373] Briefly, cUGDH was cloned into a modified pET-15b vectors (Norclone) with N-terminal hexahistidine affinity tags adjacent to a tobacco etch virus (TEV) cleavage site. Recombinant proteins were purified using a Talon® immobilized metal affinity column, and the affinity tags were removed with 5 μM TEV protease for ~16 hrs at 20 °C. Proteins were dialyzed into a storage buffer (25 mM

TRIS pH 8.0 and 50 mM NaCl) and then concentrated to 20 mg/mL using an Ultra-15 10K centrifugal filter. Proteins were quantified from dilution replicates ($N \geq 6$) using the molar extinction coefficients of 53806 M^{-1} cm^{-1} and 49850 M^{-1} cm^{-1} for cUGDH and hUGDH, respectively (calculated with the program PROTPARAM).[334] Proteins were aliquoted and flash frozen in liquid nitrogen, and then stored at -80 °C.

Protein Crystallization and Structure Determination

The cUGDH protein with bound UDP-Xyl was crystallized at 20 °C using the hanging drop vapor diffusion method, with a 2 μL drop that consisted of a 1:1 mixture of protein (10 mg/mL) and reservoir solution. Optimized crystals were obtained using a reservoir solution that consisted of citric acid buffer pH 5.0, 200 mM LiCl, 1 mM UDP-Xyl and 4% PEG 8000. These crystals were soaked in a cryoprotectant mixture composed of the same reservoir solution supplemented with 15% *v*/*v* of cryoprotectant solution (1:1:1 ethylene glycol:DMSO:glycerol ratio) and then rapidly plunged into liquid nitrogen.

Diffraction data were collected on the 22-ID beamline (SER-CAT) at Argonne National Laboratory (Argonne, IL) using a MAR 300 mm CCD detector. The 2.45 Å data set was processed with XDS in the space group $P22_12_1$, setting aside 5% of the data for cross-validation.[335,336] The structure was solved by molecular replacement using PHENIX[337] and a single dimer of the unliganded cUGDH crystal structure (PDB: 2O3J; unpublished) as a search model. During refinement, the structure was subjected to iterative cycles of manual model rebuilding using COOT and the automated refinement procedure that is implemented within the PHENIX software suite.[337-339] The R_{free} test set was corrupted during refinement, so a posterior R-free was calculated by selecting 7% of the reflections for a new cross-validation data set and subjecting the final model

to Cartesian dynamics at 4000º to decouple the test set.[336] Data collection and refinement statistics can be found in Table 3.1.

Structural Analyses

The program DynDom was used to identify the hinge bending axes for the domain rotation between the open and closed states and the rigid body rotation axis that relates the 32 symmetry hexamer to the horseshoe-shaped conformation.[374,375] PISA[376] and DIMPLOT[377] were used to identify interacting residues at the hexamer building interfaces in the cUGDH and hUGDH structures. The α6-helix rotation angle was calculated using PyMOL (https://pymolwiki.org/index.php/AngleBetweenHelices). Briefly, the monomers from the unliganded and UDP-Xyl bound structures were superimposed via least-squares-refinement using the Cα residues 1-132, and the rotation of the α6-helix was measured by using the Cα residues 141-151. Both PyMOL[378] and UCSF Chimera 1.13[379] were used to generate figures.

Sedimentation Velocity

Sedimentation velocity experiments were conducted as previously described.[318,320-322,325,372,373] Briefly, 9 μM cUGDH was dialyzed >12 hr at 4 °C into a buffer with or without 80 μM of UDP-Xyl containing 25 mM HEPES pH 7.5 and 150 mM KCl. Samples were loaded into 12 mm double-sector Epon centerpieces equipped with quartz window cells. Cells were then loaded into an An60 Ti rotor and equilibrated to 20 °C for 1.5 hrs. Sedimentation velocity data were collected using an Optima XLA analytical ultracentrifuge at 50,000 rpm for 8-10 hrs. Absorbance data were recorded at 280 nm in radial step sizes of 0.003 cm. SEDNTERP was used to estimate the partial specific volume of cUGDH (0.73895 mL/g) and the buffer density (1.00726 g/mL) and viscosity (0.01018 P).[343] Data were modelled as a continuous sedimentation coefficient (*c(s)*) distribution using SEDFIT.[344] The baseline, meniscus, frictional coefficient, and systematic

time-invariant, and radial invariant noise were fit.[345] Theoretical *s*-values were calculated using HYDROPRO[346] and the available crystal structures (PDB: 2O3J and 6OM8).

Steady-State and Transient-State Kinetics

All steady-state kinetic assays were performed as previously described.[318,320-322,324,325,372,373] Briefly, assays contained 100 nM enzyme and either saturating amounts of UDP-Glc or NAD^+ in the reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA). Solutions containing enzyme and substrate/cofactor were incubated separately at 25 °C for 5 mins, and then the reaction was initiated by rapidly mixing the two solutions. The progress of the reaction monitored NADH production ($\varepsilon_{340\,nm}$ = 6220 $M^{-1}cm^{-1}$) on an Agilent 8453 UV/vis spectrometer at 25 °C, with absorbance readings collected every 0.5 s. Because the progress curves for both cUGDH and hUGDH display hysteresis, the initial velocity (v_i) represents a transient that violates the steady-state approximation.[318,320,372,373] Thus, the initial steady-state velocities (v_{ss}) were derived from fitting progress curves to Frieden's equation for hysteresis before the depletion of 10% substrate:[318,320,342,372,373]:

$$P(t) = v_{ss}t - \tau(v_{ss} - v_i)(1 - e^{\frac{-t}{\tau}}) \qquad (1)$$

Where *P* is the product produced at time *t*, and τ is the relaxation time of the lag. The length of the lag is equal to $e\tau$. All data were fit using nonlinear regression analysis in PRISM (GraphPad Software). UDP-Glucose substrate saturation curves were fit to equation 2:

$$v_o = \frac{k_{cat}[E_t][S]}{K_M + [S]} \qquad (2)$$

Where v_o is the initial steady-state velocity (v_{ss} in equation 1), $[E_t]$ is the enzyme concentration (100 nM), and [S] is the variable substrate concentration. Both enzymes display negative cooperativity in NAD^+ saturation curves and were fit to the sigmoidal rate equation 3:

$$v_o = \frac{k_{cat}[E_t][S]^h}{(K_{0.5})^h + [S]^h} \qquad (3)$$

Where $K_{0.5}$ is the half saturation point and h is the Hill coefficient. The affinity (K_i) for UDP-Xyl was determined by globally fitting substrate saturation curves with and without inhibitor to a competitive inhibition model with cooperativity (equation 4) using PRISM:[318,320,372]

$$v_\circ = \frac{k_{cat}[E_t][S]^h}{(K_M^{app})^h + [S]^h} \text{ where } K_M^{app} = K_M\left(1 + \frac{[I]}{K_i}\right) \qquad (4)$$

The hysteresis (lag in activity) in progress curves was measured at 25 °C using an Applied Photophysics SX20 stopped-flow spectrophotometer as previously described.[372] Briefly, 500 nM cUGDH or hUGDH in the reaction buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 5 mM EDTA) was rapidly mixed with an equal volume of the same buffer supplemented with saturating amounts of both substrate and cofactor. Progress curves were monitored by NADH production (ε_{340} = 6220 M^{-1} cm^{-1}) and fit to equation 1.

Phylogenetic Analysis

Human UGDH orthologs were identified by running BLAST on the NCBI non-redundant database. A subset of 31 sequences from a total of over 70,000 UDGH related sequences were selected that represented the overall taxonomic diversity of UGDH sequences from bacteria to humans. The selected sequences were aligned using the MAPGAPS scheme.[380] The alignment and taxonomic information for the 31 sequences is provided in supplementary material (Supplementary Table 3.1). The aligned sequences were used for phylogenetic tree generation using IQTree v1.6.1 with the following options: -m MFP -bb 10000.[381] Bootstrap support values were calculated using UltraFast bootstrap approximation (UFBoot) implemented in IQ-Tree. Independent trees generated using FastTree showed topologies similar to IQTree.[382]

3.4 Results

cUGDH Forms the E^Q State

The crystal structure of cUGDH in complex with UDP-Xyl (cUGDH:UDP-Xyl) was solved and refined to a resolution of 2.45 Å (Figure 3.3A,B and Table 3.1). The asymmetric unit contains 12 chains that are structurally equivalent, with a small number of disordered residues at the N- and C-termini (≤7 and ≤11 residues, respectively). Each chain of cUGDH also contains a well-ordered molecule of UDP-Xyl in the active site, which conserves all of the same ligand interactions previously described[321] for the human enzyme (Figure 3.3C, D). In addition, there is second molecule of UDP-Xyl bound in the NAD^+ binding site that is most likely an artefact of the high UDP-Xyl concentrations (1 mM) in the crystallization buffer (not shown); at high concentrations, UDP-Xyl is known to bind promiscuously to the NAD^+ site in hUGDH.[372]

The cUGDH structure superimposes 453 corresponding Cα atoms onto hUGDH (PDB entry 3PTZ) with an rmsd of 0.7 Å and sequence identity of 61%. The domain structure is assigned based on homology to the human enzyme: NAD^+ binding domain (NB; res. 1-219), dimerization domain (DD; res. 220-329), and sugar-binding domain (SB; res. 330-481) (Figure 3.2). The application of crystallographic symmetry shows that all twelve chains of cUGDH assemble into horseshoe-shaped hexamers, similar to the UDP-Xyl bound E^{Ω} state[321] observed in the human enzyme (Figure 3.3B). The largest difference between the cUGDH and hUGDH E^{Ω} hexamers involves a 5.6° larger rigid body rotation of adjacent dimers, which produces a wider opening in the cUGDH 'horseshoe' complex.

UDP-Xyl binding induces formation of the E^{Ω} hexamer

Sedimentation velocity analysis of 9 μM unliganded cUGDH revels a *c(s)* distribution consisting of an 11.2 S (68.5%), 8.3 S (8.7%), 5.6 S (21.6%) and 3.3 S (1.2%) species (Figure 3.4A). This is consistent with the distribution of the hUGDH unliganded E* state, which exists as a mixture of hexamers, tetramers and dimers in rapid equilibrium under identical

conditions.[318,323,372] The 3.3 S species likely represents a small amount of misfolded monomer, which is also observed in sedimentation velocity studies with hUGDH.[318,323,372] The cUGDH hexamer (11.2 S) appears to be slightly smaller than the predicted value (11.5 S), which is likely a result of the fact that in a rapid-equilibrium system the modeled S-values are biased by the mean of the distribution.[345,383] Because the E* hexamer is less stable than the E^{Ω} complex, the allosteric transition can be easily observed in sedimentation velocity studies.[321] The addition of 80 μM UDP-Xyl shifts the distribution to a single 11.6 S species, which is nearly identical to the expected S value (11.7 S) of the E^{Ω} state, horseshoe-shaped hexameric complex (Figure 3.4B). The sharp peak observed near the origin indicates a slight buffer mismatch between the sample and reference cells.

A 1.88 Å resolution crystal structure of a unliganded cUGDH hexamer with 32 symmetry has been deposited as PDB entry 2O3J by the New York SGX Research Center for Structural Genomics (unpublished) (Figure 3.4C). The NB and SB domains of the unliganded cUGDH subunits have rotated to an open conformation (~11° and ~5°, respectively) relative to the E^{Ω} hexamer (Figure 3.4D). This open domain conformation resembles the E* state of hUGDH, which represents an intermediate in the allosteric transition from E to E^{Ω} (Figure 3.1B).[318,320,323] The Thr136-Loop/α6-helix (allosteric switch) in cUGDH was identified based on homology with the human enzyme (Figure 3.2). A comparison of the unliganded cUGDH structure with the human enzyme confirms that the allosteric switch is in the E* conformation (Figure 3.4E). This analysis also shows that cUGDH and hUGDH undergo a similar conformational change in the transition from the E* and E^{Ω} state (Figure 3.4E).[321,322] The only significant difference in the Thr136-Loop portion of the allosteric switch involves the conserved residue Val139 (cUGDH numbering) (Figure 3.5A). In cUGDH, the side chain of Val139 is weakly ordered, and is modeled in the *gauche+*, *gauche-*, and *trans* rotamers in the different peptides of the crystal structure, while in

hUGDH the side chain is in the *trans* conformation (-170°). The rotational disorder of the valine appears to be in response to the bulky pyrrolidine ring of Pro109, which displaces the Val139 Cα atom by 1.4 Å relative to the human enzyme. This displacement also induces the conserved residue Leu228 to change rotamers (Figure 3.5A). The largest conformational difference between the enzymes involves the α6-helix; in the allosteric transition from the E* hexamer to the E^{Ω} complex the cUGDH helix rotates by 14 ± 3°, but only 6 ± 2° in the human enzyme (Figure 3.5B). The difference in rotation originates from several changes in the packing interactions between the α6-helix and the groove formed between the α5-helix and the β6-strand. First, the Cβ atom Ala141 and the Cγ atom of Val114 are ~4.7 Å apart, which creates a small void that weakens the packing of the α6-helix in cUGDH (Figure 3.5C). In hUGDH, Val114 is replaced with an isoleucine which fills the void with a Cδ atom and packs tightly against the alanine. Another significant change involves an alanine-to-serine substitution at position 118 in cUGDH, which is buried in a hydrophobic pocket where it weakens a hydrogen bond in the α5-helix by competing for the main chain carbonyl oxygen of Val114 (Figure 3.5D). The most significant change in the α6-helix involves Ala152 in cUGDH, which packs into a buried hydrophobic cluster formed by Ala122, Leu149 and Phe161 (Figure 3.5E). In the human enzyme, the hydrophobic cluster is maintained with conservative substitutions, but Ala152 is replaced by an asparagine (Asn147 in hUGDH, Figure 3.5E). As a result, the C-terminus of the α6-helix in hUGDH has locally unfolded to expose the Asn to solvent. There is also a two-residue insertion in the loop following the α6-helix in cUGDH, but it does not appear to contribute directly to the change in helix rotation (Figure 3.2 and 3.5E).

There are several structural changes in the cUGDH hexamer-building interface that accommodate the altered conformation of the α6-helix (Figure 3.5F and Figure 3.S1). The most

significant differences involve the conserved residues Phe330 and Pro440, both of which pack directly against the α6-helix. In cUGDH, Phe330 adopts a χ_1 torsion angle of -69° and packs against the α6-helix residue Cys147 (Figure 3.5F). In contrast, Cys147 is replaced with an arginine in the human enzyme, which forces the corresponding Phe into a strained, weakly ordered rotamer of -102°. The conserved residues Glu143 and Ser144 change conformations to produce a complementary packing surface for Phe330 in each enzyme. The conserved residue Pro440 is also displaced by ~1.9 Å relative to the human enzyme (Figure 3.5F). This shift is a consequence of the altered α6-helix conformation and the replacement of alanine by Glu151 in cUGDH.

The Atypical Allostery Motif in cUGDH

With the exception of the alanine-to-proline substitution at position 109, hUGDH and cUGDH conserve the atypical allostery motif and most of the associated packing defects in the E^{Ω} hexamer (cavity C-1^{Ω} and deep surface pockets D-1^{Ω} and D-2^{Ω}) (Figure 3.2 and 3.6A-C). In hUGDH, the main-chain amide of the alanine forms a hydrogen bond with the carboxamide of an asparagine (Asn283, human numbering) (Figure 3.6D). The proline substitution in cUGDH removes the amide, which would break the hydrogen bond (Figure 3.6E). As a result, the selective pressure on the asparagine is relaxed, and it has been replaced with Ser290 which forms a hydrogen bond with an ordered water molecule in the deep surface pocket D-2^{Ω}. The pyrrolidine ring of Pro109 also alters the structure of deep surface pocket D-1^{Ω} by introducing a steric clash that induces the rotameric changes in Val139 and Leu228 that were described in the previous section (Figure 3.5A). The only other change in D-1^{Ω} involves Val114, which replaces an isoleucine in the human enzyme.

cUGDH Conserves Hysteresis and Cooperativity

The atypical allosteric hUGDH undergoes a slow, substrate-induced isomerization from the inactive E* to the active E state, which can be observed as a lag in progress curves.[318] The cUGDH progress curves also display a similar hysteresis, with a lag that is ~30% shorter than that observed in hUGDH (Figure 3.7A,B). The kinetic parameters for cUGDH were determined using the steady-state velocities derived from Frieden's equation for enzyme hysteresis.[342] The NAD^+ saturation curve exhibits negative cooperativity (Hill Coefficient = 0.82), with a $K_{0.5}$ and k_{cat} of 333 μM and 1.7 s^{-1}, respectively (Figure 3.7C and Table 3.2). This is comparable to the NAD^+ negative cooperativity in saturation curves that has previously been reported[320,324,325] for the human enzyme (Figure 3.7D and Table 3.2). In contrast, the UDP-Glc saturation curve is hyperbolic, with a K_M and k_{cat} of 55 μM and 1.2 s^{-1}, respectively (Figure 3.7E and Table 3.2). The lower k_{cat} is a result of the negative cooperativity associated with NAD^+ binding, which makes it difficult to saturate cUGDH with co-enzyme in order to achieve pseudo-first order conditions.[318,323]

Steady-state analysis of UDP-Xyl inhibition yields a K_i of 6.9 μM, which corresponds to a ~16-fold weaker affinity for the inhibitor than that seen in hUGDH (Figure 3.7E,F and Table 3.3). The cUGDH UDP-Glc saturation curves also exhibit positive cooperativity in the presence of 15 μM UDP-Xyl (Hill coefficient of 1.4) (Figure 3.7E and Table 3.3). Substrate cooperativity in the presence of saturating UDP-Xyl has been observed in hUGDH, and interpreted as evidence of the allosteric transition from a low UDP-Glc affinity E^{Ω} complex to a higher affinity E hexamer.[318,320,323,372]

Evolutionary Analysis of Atypical Allostery in UGDH

The evolution of atypical allostery in UGDH was examined using over 70,000 sequences. A phylogenetic tree of 31 representative UGDH sequences from diverse eukaryotic and prokaryotic phyla is shown in Figure 8. Based on overall sequence similarity and taxonomic

representation, these 31 sequences can be broadly classified into 8 clades. The nature of amino acids present in the atypical allostery motif (including the newly defined S/N290 position) was compared across phyla (Figure 3.8). While both the atypical allosteric motif residues and the packing defect residues are generally conserved across the eukaryotic and prokaryotic phyla (clades 1-7), they are strikingly different in the non-allosteric *S. pyogenes* clade (clade 8), which was used as an outgroup (see methods) (Figures 3.8 and 3.9). If a clade contains a single, conservative variation, then it is identified as likely allosteric, but if there are significant changes, then it is assumed to be non-allosteric (Figure 3.9). Sequences in clades 1-2 strictly conserve the allosteric motif residues, while notable variations are observed in clades 3-7. In clade 3, the Capsaspora phylum absolutely conserves the motif, while the Platyhelminthes substitute a serine for the alanine at position 141. The impact of the serine is likely to be negligible since the Oγ would point into the deep surface pocket D-2^{Ω}, which is large enough to accomadate a water molecule for hydrogen bonding. In clades 5-8, we observe the persistent conservation of a Thr substitution in position 141. This is most pronounced in the Stramenopiles, Bacillariophyta and Phaeophyceae phyla of clade 5, where the Thr141 substitution represents the only significant variation in either the atypical motif or packing defect residues. In contrast, the remaining phyla of clade 5 also contain a tryptophan substituion in the packing defect residue position 114, which will likely introduce significant clashes in the inhibited state of the enzyme. The Ascomycota (clade 7) contain a leucine substitution at position 109, which is interesting because the equivalent substitution in the human enzyme has been shown to eliminate allostery by introducing steric clashes in the deep surface pocket D-2^{Ω}.[323] In general, the overall conservation of the motif residues in fungal and cyanobacterial clades (clades 6-7) suggests that the atypical allosteric mechanism is likely ancient and evolved progressively in higher eukaryotes (clades 1-2) (Figure

3.9). Because of the specificity for UDP-Xyl, it is unlikely that the atypical allosteric inhibition mechanism would be present in the UGDHs of organisms that do not produce this nucleotide-sugar. Unfortunately, data concerning the distribution of UDP-Xyl throughout various phyla is poorly characterized and incomplete.

3.5 Discussion

What makes allostery in UGDH atypical is that the feedback inhibitor competes with substrate for the active site to trigger a transition to the E^{Ω} state (Figure 3.1B).[320,321] Key to this remarkable transition is the atypical allostery motif, which identifies a series of large-to-small amino acid substitutions that produce the packing defects in the protein core that are necessary to accommodate both the E and E^{Ω} conformations of the buried allosteric switch (Figures 3.4E and 3.6A-C).[321,323] This motif is conserved in cUGDH with the exception of an alanine to proline substitution in position 109 (Figure 3.2). In an early study with the human enzyme, it was shown that a leucine substitution at the equivalent position abolished allostery.[323] Here, we show that despite the A109P substitution, the crystal structure of cUGDH:UDP-Xyl still forms the allosterically inhibited E^{Ω} hexamer (Figure 3.3B). The structure also suggests a reason for why Ser290 is replaced with an asparagine in hUGDH (Figures 3.6D,E). There, the asparagine forms a hydrogen bond with the main-chain amide of the alanine in position 109 of the atypical motif (Ala104 in hUGDH) (Figures 3.6D,E). The Pro109 substitution in cUGDH removes the amide, which relaxes the selective pressure on the asparagine and allows the Ser290 substitution.

A comparison of the E* and E^{Ω} cUGDH structures shows that the allosteric switch undergoes a conformational change similar to that observed in the human enzyme, but with a larger rotation of α6-helix (Figure 4E and 5B). It has been shown that changes in the conformation of the α6 helix can affect the stability of the hUGDH hexamer and the affinity for UDP-Xyl.[318,321,323]

Here, we have shown that the E* state of cUGDH is a weakly associated hexamer, while the allosterically inhibited enzyme forms a stable E^{Ω} complex, similar to human enzyme (Figure 3.4A,B).[321,372] However, the UDP-Xyl affinity of cUGDH is ~16-fold weaker than that of the human enzyme, despite the fact that both enzymes conserve the same interactions in the binding site (Table 3.3 and Figure 3.3D). Since UDP-Xyl is coupled to the hexamer-building interface through the allosteric switch, it is likely that the changes in the rotation of the α6-helix are responsible for the reduced affinity (Figure 3.4B).

Hysteresis and allostery have been shown to be coupled in hUGDH.[318,320,321,323] Our observation of hysteresis in cUGDH suggests that this is a defining characteristic of the atypical allosteric mechanism (Figure 3.7A,B). The allosteric transition can observed as positive cooperativity in steady-state inhibition studies (Figure 3.7E,F, Table 3.3).[320] Briefly, the E^{Ω} conformation has low affinity for the substrate UDP-Glc (Figure 3.1D).[320,321] As UDP-Glc outcompetes the inhibitor, the enzyme allosterically switches to the high substrate affinity E state, resulting in a sigmoidal substrate saturation curve.[320] Both hUGDH and cUGDH also display negative cooperativity with respect to NAD^+ binding (Figure 3.7C,D and Table 3.2). The negative cooperativity in hUGDH is poorly understood at a structural level, but it is known to originate from a substrate-induced asymmetry in the enzyme.[324] Negative cooperativity is believed to be linked to the allosteric mechanism, because amino acid substitutions that disrupt allostery also abolish negative cooperativity.[318,323] The conservation of negative cooperativity in cUGDH strongly suggests that this phenomenon is also a defining characteristic of the atypical allosteric mechanism.

To obtain new insights into the evolution of allostery in UGDH, we have examined the patterns of conservation and variation in the atypical allostery motif residues across diverse

sequences (Figures 3.8 and 3.9). Phyla that conserve Ala109 also conserve the Asn290, supporting our hypothesis that these residues have co-evolved. The UGDHs of higher eukaryotes (clades 1-3) generally conserve the motif and are likely to be allosteric, while clade 8 is non-allosteric due to striking divergence in the motif residues (Figures 3.8 and 3.9). However, the remaining clades (clades 4-7) will require additional analysis to identify allosteric UGDHs. Specifically, we are intrigued by the selective conservation of threonine at position 141 in clades 4-7 (Figure 3.9). If that substitution prevents the formation of the E^{Ω} complex, then it may be that we have identified a key step in the evolution atypical allostery in UGDH. Future work will focus on mapping these evolutionary steps through the experimental characterization of UGDHs from the Stramenopiles, Bacillariophyta and Phaeophyceae phyla in clade 5. Because these sequences contain the Thr141 substitution and conserve most of the packing defect residues, they appear to be important transition points between non-allosteric bacterial UGDH and allosteric human UGDH.

3.6 Acknowledgements

We thank Dr. Renuka Kadirvelraj for advice in the refinement of the cUGDH crystal structure

3.7 Figures

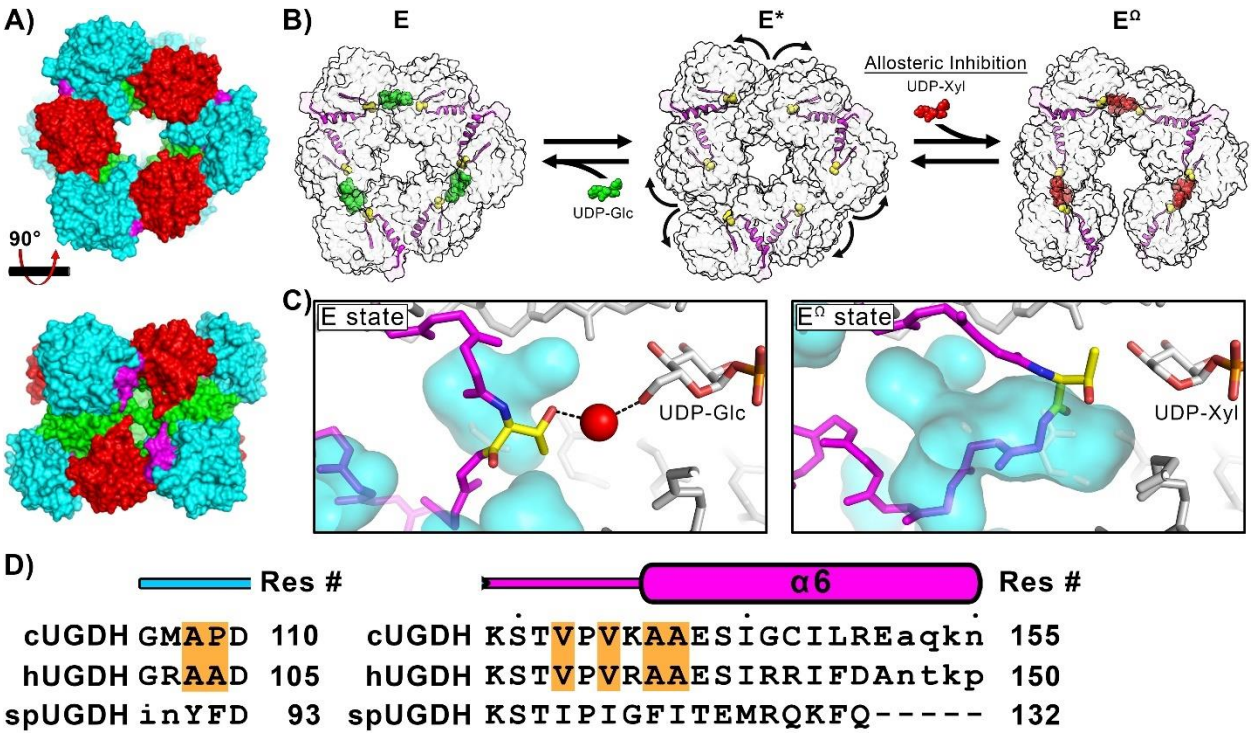


Figure 3.1: The allosteric transition in hUGDH. (A) Hexameric hUGDH forms from the association of three dimers. The allosteric switch (magenta), NAD^+ binding (NB) domain (cyan), dimerization domain (green), and nucleotide-sugar binding (SB) domain (red) are identified. **(B)** In the inactive E* state, the NB and SB domains adopt an 'open' conformation (curved arrows). The binding of substrate (green spheres) induces the formation of the active, E state, while UDP-Xyl (red spheres) produces the inactive, horseshoe-shaped E^{Ω} complex. **(C)** The E and E^{Ω} conformations of the allosteric switch (magenta). The packing defects in the protein core are colored cyan. In the E state, a water molecule (red sphere) forms hydrogen bonds with the C5'CH_2OH of UDP-Glc and Thr131 of the allosteric switch. UDP-Xyl lacks the C5'CH_2OH, which triggers the allosteric switch to produce the E^{Ω} state. **(D)** The atypical allostery motif residues (orange boxes) hypothesized to facilitate allostery in UGDH.[321]

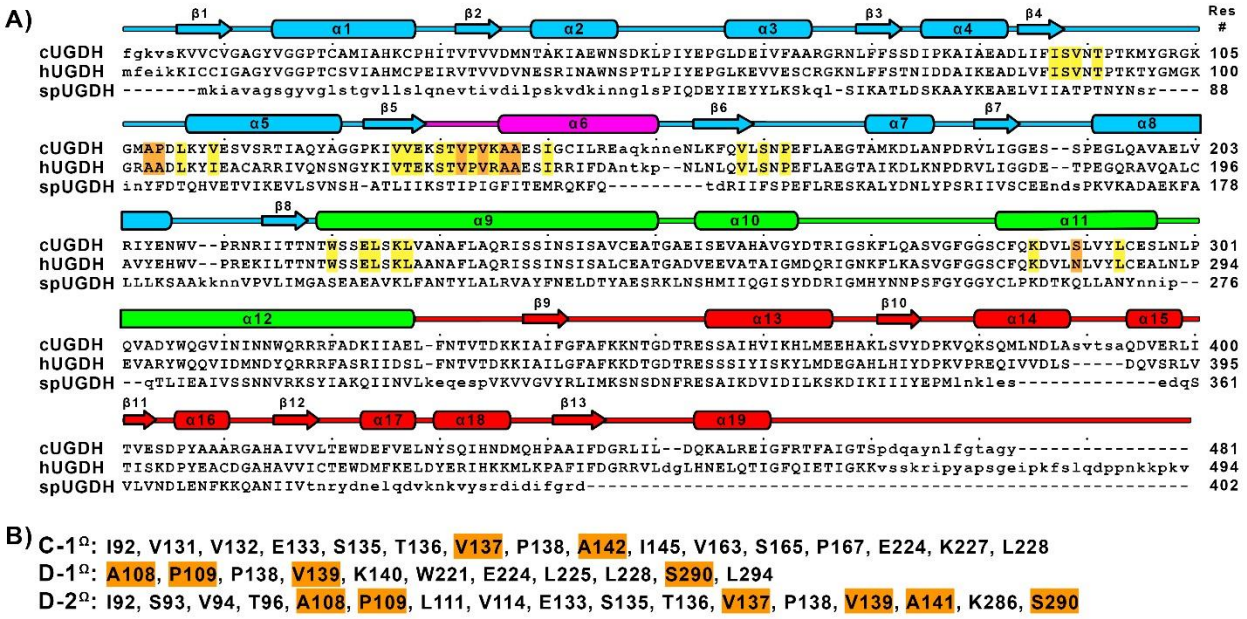


Figure 3.2: Structure-based sequence alignment. (A) The structure-based sequence alignment of cUGDH (residues 5-481), hUGDH (residues 1-494), and *S. pyogenes* UGDH **(**spUGDH; residues 1-402) constructed by superimposing the corresponding crystal structures (PDB entries 2O3J, 2Q3E and 1DLI, respectively) using Sequoia[384]. Uppercase letters identify structurally equivalent residues. The structural elements are colored according domain: NB domain (cyan), dimerization domain (green), and SB domain (red). The allosteric switch (magenta), atypical allostery motif (orange boxes), and packing defect residues (yellow boxes) are depicted. **(B)** The packing defect residues in cUGDH that form cavity C-1^{Ω} and the deep surface pockets D-1^{Ω} and D-2^{Ω}, with the atypical allostery motif residues highlighted in orange.

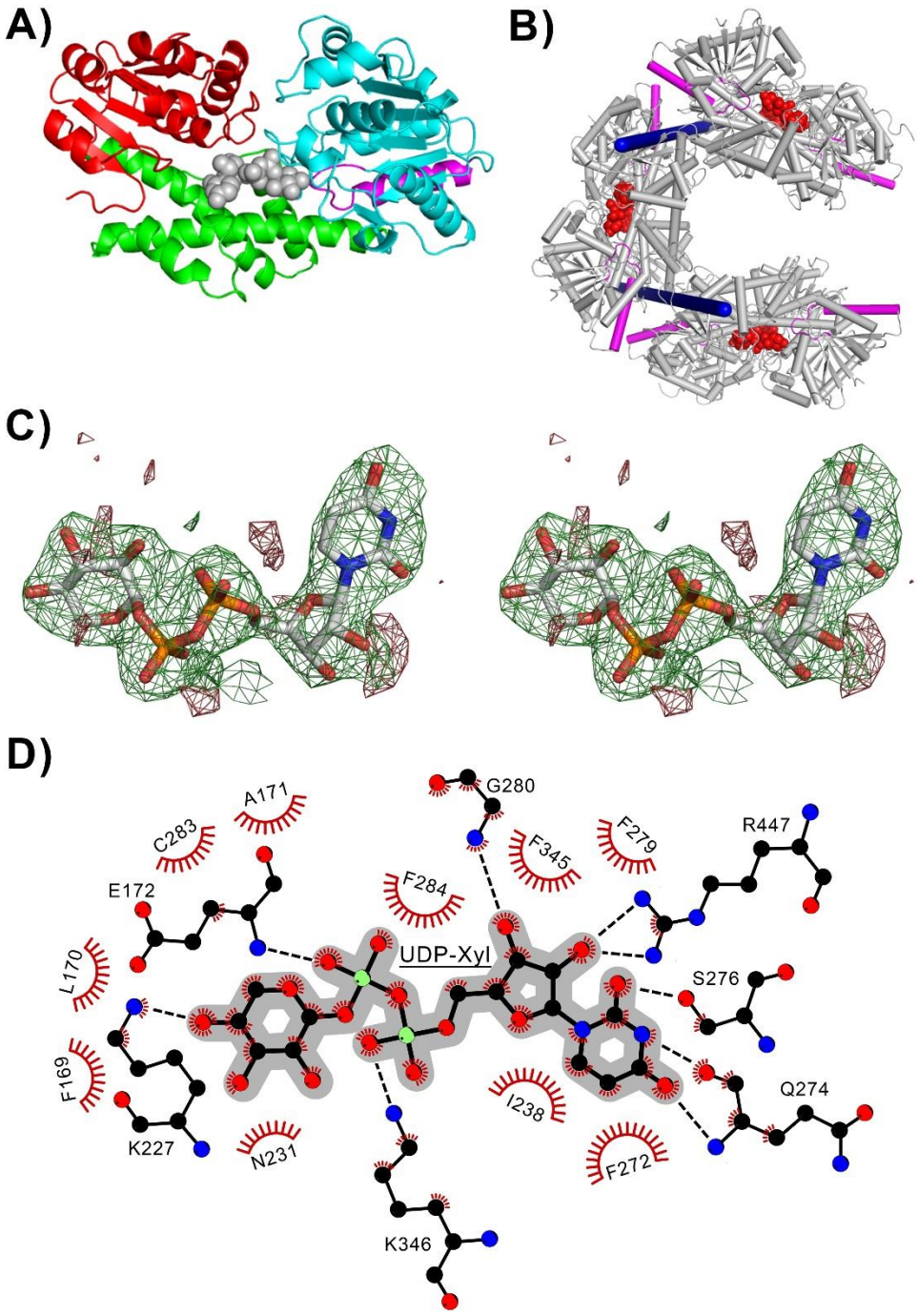


Figure 3.3: The Inhibited E^{Ω} State of cUGDH. (A) The NB domain (cyan), dimerization domain (green), SB domain (red) and allosteric switch (magenta) are shown in single chain of cUGDH in complex with UDP-Xyl (grey spheres) **(B)** The cUGDH horseshoe-shaped hexameric complex (E^{Ω}), depicted with rigid body rotation axes (blue rods) in the hexamer interfaces. The active sites with UDP-Xyl (red spheres) and allosteric switches (magenta) are also identified. **(C)** A stereodiagram of the F_o - F_c density map for UDP-Xyl contoured at 3σ and calculated by omitting UDP-Xyl subjecting the model to simulated annealing. **(D)** LigPlot[377] displaying hydrogen bonds (dashed lines) and van der Waals (red feathers) interactions with UDP-Xyl (grey highlighting).

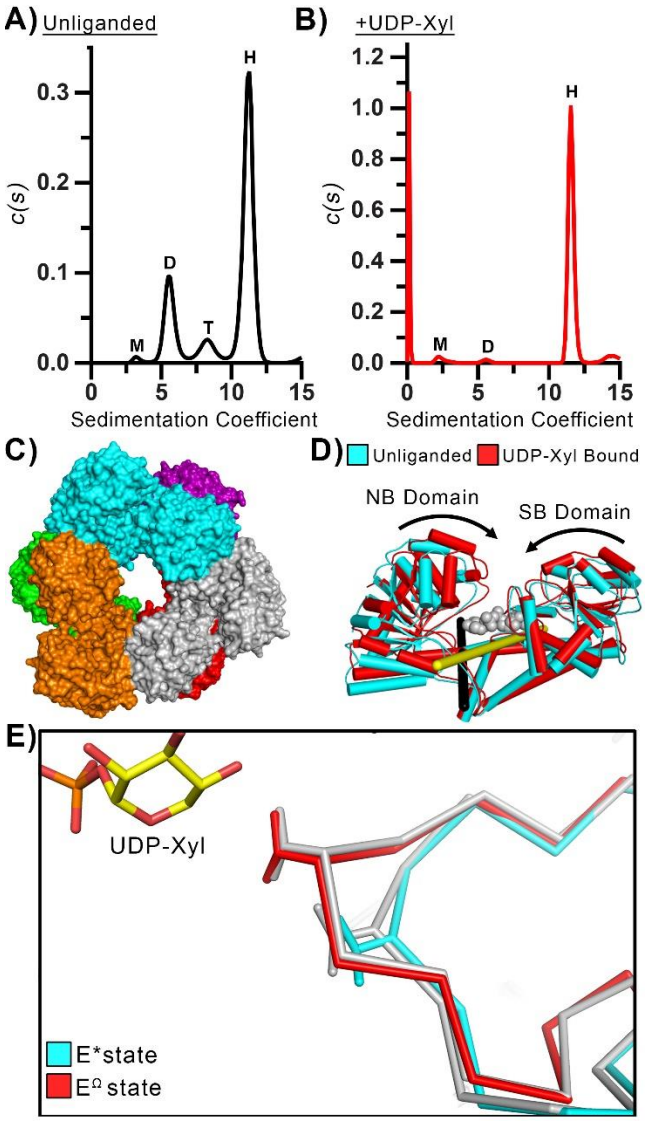


Figure 3.4: UDP-Xyl binding induces formation of the E^{Ω} state. (A) Sedimentation velocity analysis of unliganded cUGDH reveals a distribution corresponding to a (M)onomer, (D)imer, (T)etramer, and (H)examer species. **(B)** UDP-Xyl shifts the distribution to a single 11.6 S species consistent with the E^{Ω} hexamer. **(C)** Unliganded cUGDH structure forms a 32 symmetry hexamer (PDB: 2O3J; unpublished). **(D)** A superposition of an unliganded (cyan) and UDP-Xyl bound (red) monomer of cUGDH illustrates the rotation of the NB and SB domains. DynDom[374] detects hinge-bending axes in the NB (black rod) and SB (yellow rod) domains. The SB hinge-bending axis is only detected in 6 of the 12 chains. **(E)** The cUGDH allosteric switch (Cα trace with Thr136 in sticks) in the E* (cyan) and E^{Ω} (red) states are similar to the equivalent structural states of hUGDH (grey).

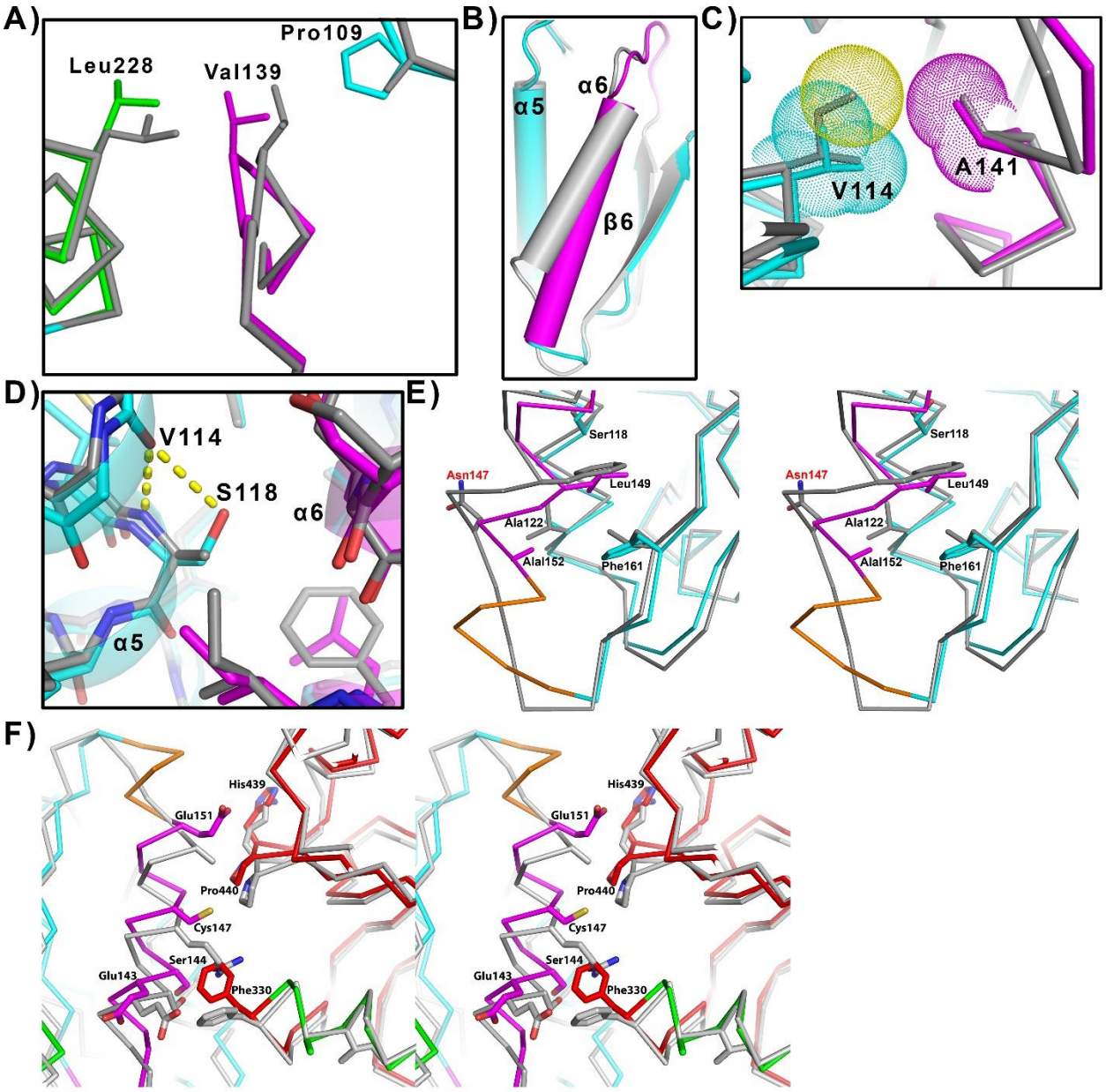


Figure 3.5: Structural divergence in the cUGDH allosteric switch. (A) The Pro109 substitution in cUGDH displaces the Val139 (magenta) which causes Leu228 to change rotamer relative to hUGDH (grey). **(B)** The α6-helix in the E^{Ω} conformation of cUGDH (magenta, cyan) is rotated more than that observed in hUGDH. **(C)** Val114 (cyan) replaces an isoleucine in hUGDH (grey). VDWs distances (dots) and the α6-helix (magenta) are identified. **(D)** Ser118 of cUGDH (cyan) replaces an alanine in hUGDH (grey). **(E)** Stereoview depicting significant changes in the packing of the α6-helix of cUGDH (magenta) against the α5-helix and β6-strand (cyan) compared to the hUGDH (grey). cUGDH contains an insertion that lengthens the α6-helix (orange). **(F)** Stereoview of changes in the packing interactions that form the hexamer building interface of cUGDH (colored by domain; NB-cyan, allosteric switch-magenta, dimerization domain-green, and SB-red). hUGDH is colored grey.

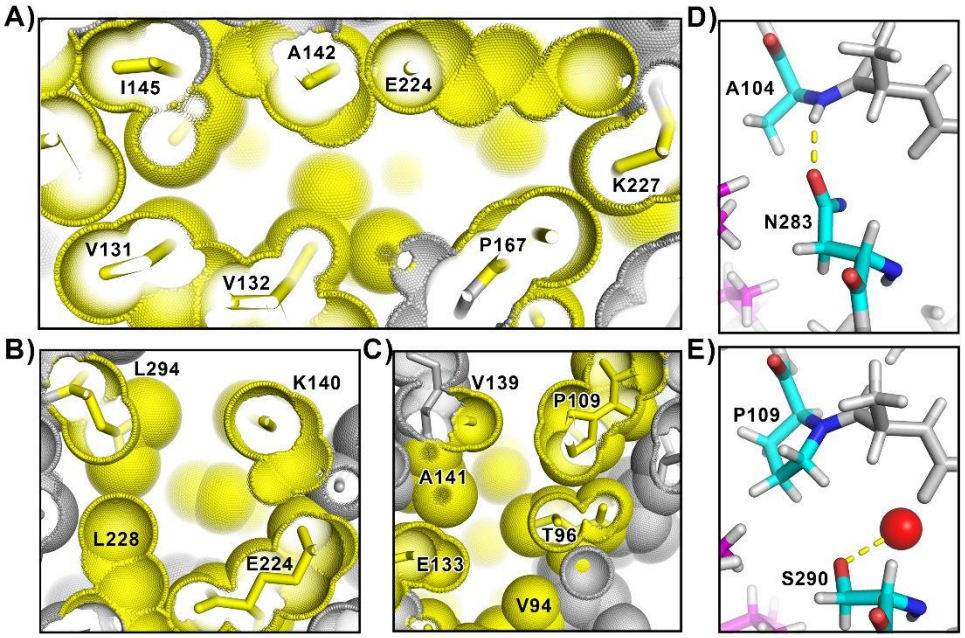


Figure 3.6: The packing defect residues (PD residues) are conserved in cUGDH. (A) A cutaway of the cUGDH E^{Ω} structure protein core depicting cavity C-1^{Ω}. **(B, C)** Same as in (A), depicting deep surface pockets D-1^{Ω} and D-2^{Ω}, respectively. **(D)** In hUGDH, Asn283 in D-2^{Ω} forms a hydrogen bond with the backbone amide of Ala104 (hUGDH numbering). **(E)** In cUGDH, the alanine and asparagine are replaced with a Pro109 and Ser290, respectively. Ser290 forms a hydrogen bond with a water molecule (red sphere).

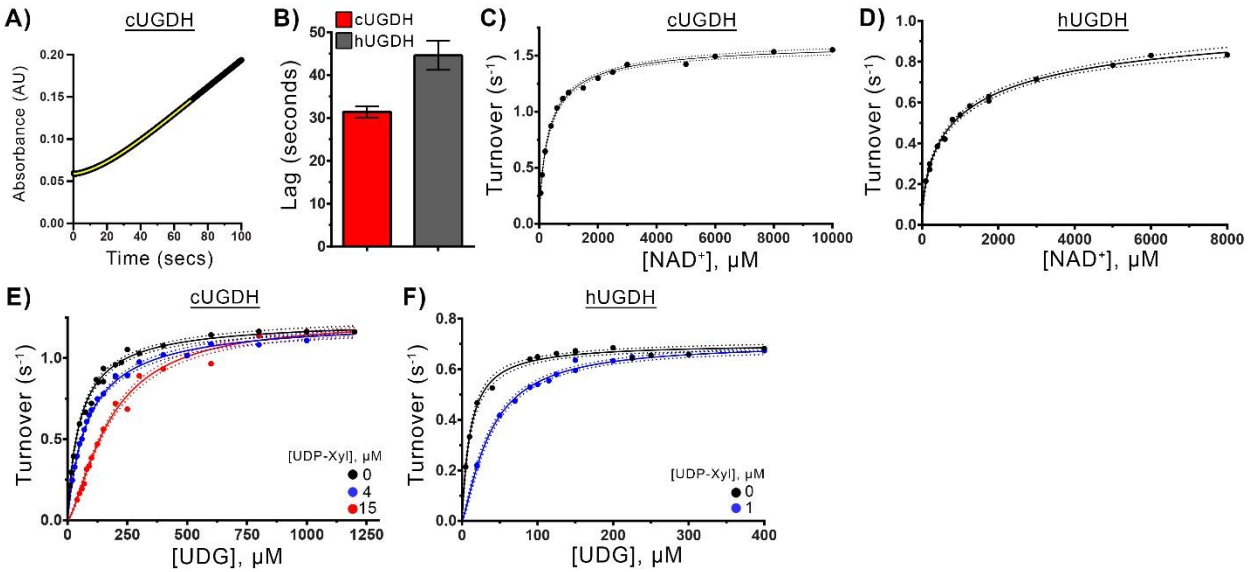


Figure 3.7: Kinetic analyses of cUGDH and hUGDH. (A) Representative cUGDH progress curve (black line) from stopped-flow absorbance spectroscopy, which is fit to equation 1 (yellow line) (eq 1). **(B)** The length of the progress curve lag under the same conditions for cUGDH and hUGDH. **(C)** NAD^+ saturation curve with cUGDH fit to eq 3 (thin line). Rates (black dots) have been normalized by enzyme concentration to give turnover, and dashed lines represent 95% confidence intervals. **(D)** NAD^+ saturation curve with hUGDH. **(E)** cUGDH substrate saturation curves with 0 μM (black), 4 μM (blue), and 15 μM UDP-Xyl (red), which were globally fit to eq. 4. **(F)** hUGDH substrate saturation curves with 0 μM (black) and 1 μM (blue) UDP-Xyl, which were globally fit to eq. 4. Kinetic parameters derived from all fits can be found in Tables 3.2 and 3.3.

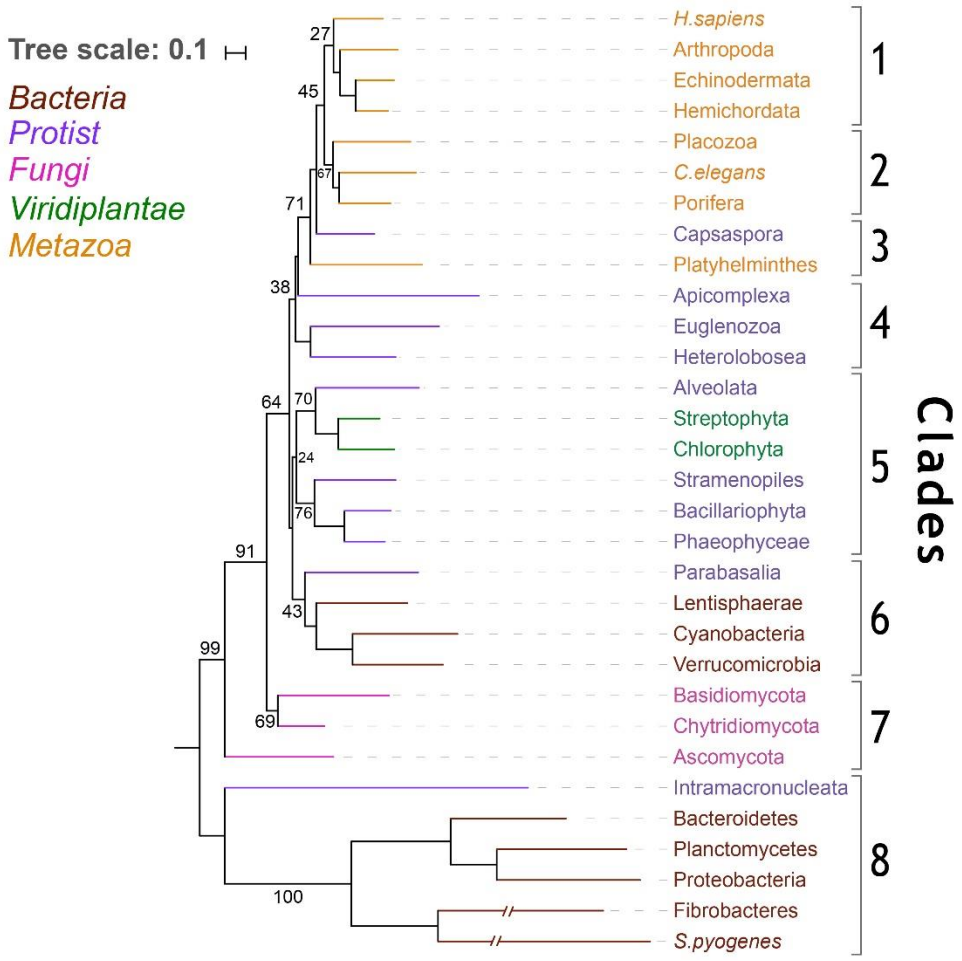


Figure 3.8: Phylogenetic analysis of UGDH. Using taxonomically diverse UGDH sequences, the evolution of UGDH was followed. This analysis distinguishes eight distinct clades of UGDH. The long branch lengths of the non-allosteric outgroup (clade 8) indicates that these phyla contain highly divergent UGDHs. Bootstrap values are indicated.

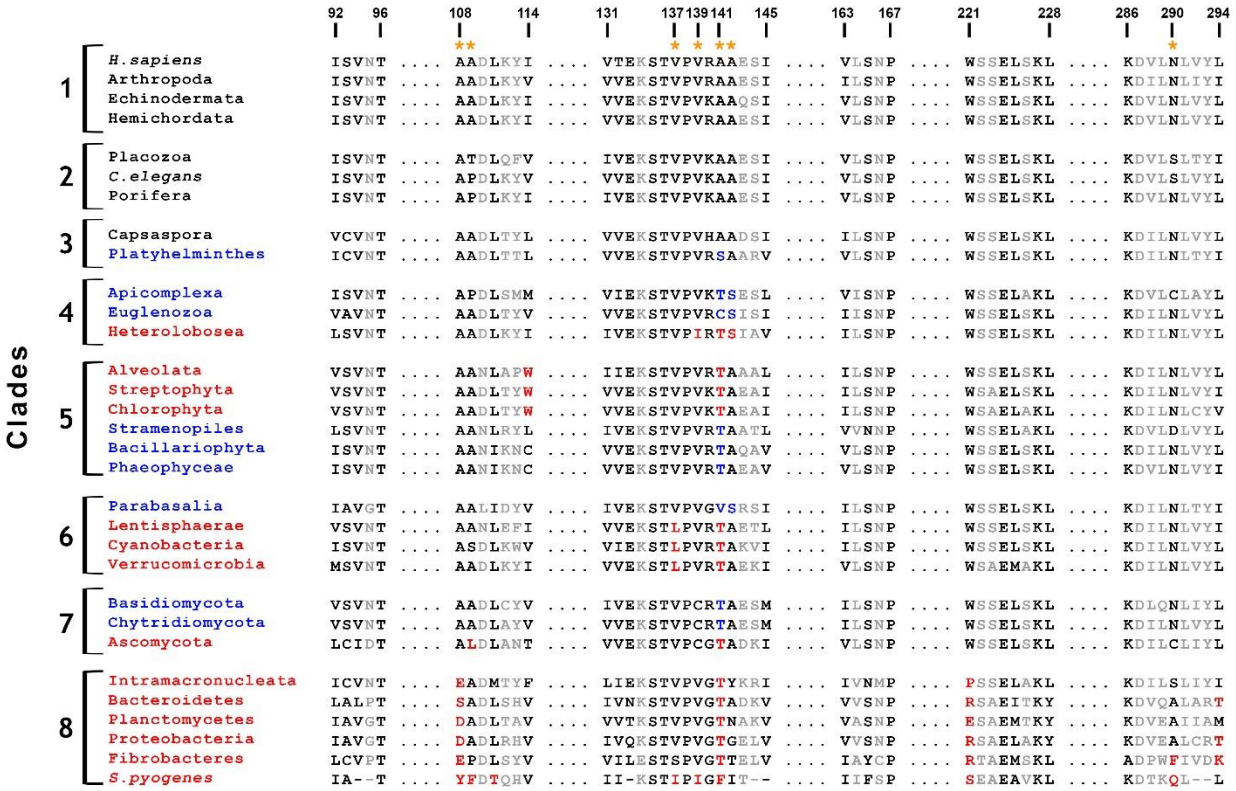


Figure 3.9: Representive sequences from each phyla in the evolutionary analysis of UGDH. Residues that make up the allosteric sequence motif (orange stars) and the packing defects (bold black lettering), or simply adjacent in sequence (light grey). The prediction of allostery in the UGDHs that comprise clades 3-7 is purely speculative (see text for details). Within each clade, the phyla that are expected to be allosteric are shown in black. If it is unknown whether or not a specific phylum conserves the allosteric mechanism, it is colored blue along with the questionable substitutions. If a specific phylum is expected to be non-allosteric, it is colored red along with the questionable substitutions that are unlikely to be accommodated in the E^{Ω} structure.

3.8 Tables

Table 3.1: Data Collection and Refinement Statistics for UDP-Xyl bound cUGDH

Data Collection	
Protein Data Bank Entry	6OM8
Space group	P 2 2_1 2_1
Unit cell dimensions (Å)	157.7, 168.2, 279.6
Completeness (%)	98.1 (98.4)[a]
No. reflections	266577 (19605)
Redundancy	6.3 (6.1)
I / σ(I)	9.1 (1.3)[a]
$CC_{1/2}$[b]	99.6 (45.5)[a]
R_{meas} (%)[c]	15.3 (164.6)
R_{sym} (%)	14.1 (150.6)
Refinement	
Resolution (Å)	2.45
R_{work} / R_{free}	0.19 / 0.22[d]
No. atoms: Protein / Ligand / Water	42921 / 816 / 1194
B-factors (Å^2): Protein / Ligand / Water	53.3 / 49.1 / 49.0
Stereochemical Ideality	
Bond lengths (Å^2)	0.008
Bond angles (°)	0.905
φ,ψ Most favored (%)[d]	96.6
φ,ψ Additionally allowed (%)	3.4

[a]Values in parentheses are for highest-resolution shell (2.51-2.45)
[b]$CC_{1/2}$ is the percentage of correlation between intensities from random half-data sets[341]
[c]R_{meas} is the redundancy-independent merging R factor[340]
[d]Posterior R_{free} (see methods for details)

Table 3.2: Kinetic parameters for cUGDH and hUGDH

Enzyme	Ligand	K_M (μM)	k_{cat}[a] (s^{-1})	$K_{0.5}$ (μM)	k_{cat}[a] (s^{-1})	Hill (*h*)
cUGDH	UDP-Glc	55 ± 3	1.23 ± 0.03			1.0
	NAD^+			333 ± 18	1.7 ± 0.1	0.8 ± 0.1
hUGDH	UDP-Glc	11.0 ± 0.8	0.71 ± 0.02			1.0
	NAD^+			1094 ± 244	1.1 ± 0.1	0.6 ± 0.1

[a]One complete catalytic turnover produces two molecules of NADH.

Table 3.3: Global Analysis of UDP-xylose Inhibition

Enzyme	UDP-Xyl (μM)	K_M (μM)[a]	k_{cat} (s^{-1})[a]	K_i (μM)[a]	Hill (*h*)
cUGDH	0	54 ± 3	1.23 ± 0.03	6.9 ± 0.6	1.0
	4	-	-	-	1.0 ± 0.1
	15	-	-	-	1.4 ± 0.1
hUGDH	0	11.0 ± 0.8	0.70 ± 0.01	0.42 ± 0.04	1.0
	1	-	-	-	1.3 ± 0.1

[a]Shared parameter in global fitting.

3.9 Supplementary Figures and Tables

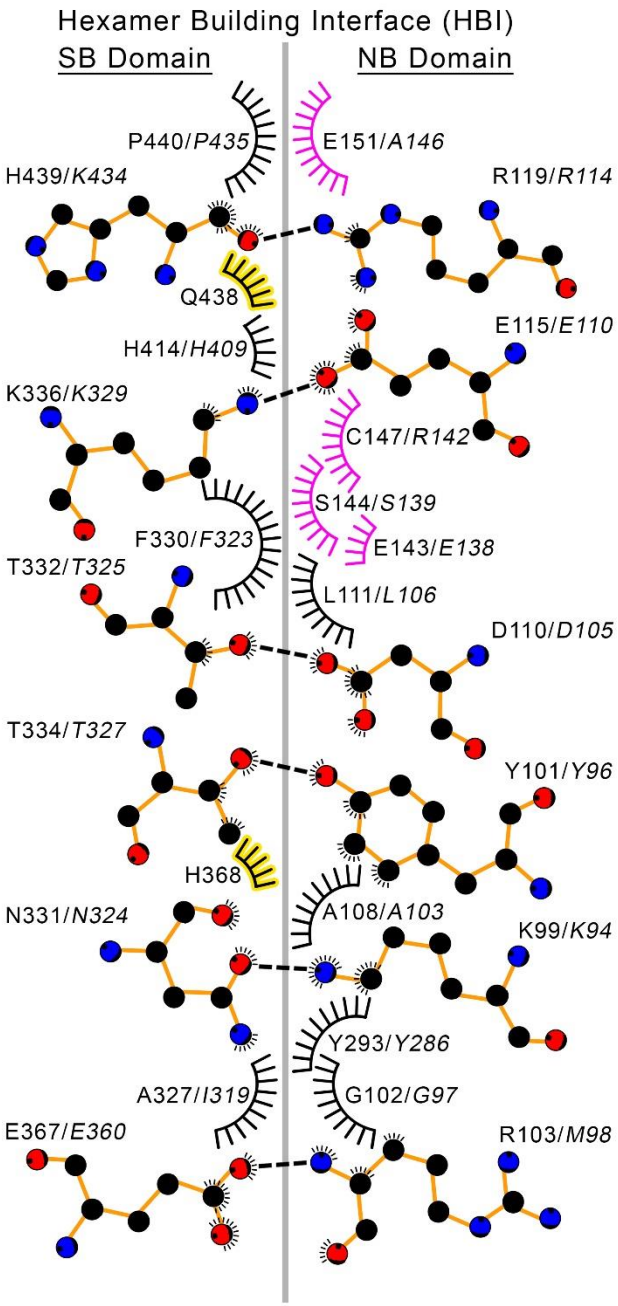


Figure S3.1: The hexamer building interface residues of cUGDH and hUGDH. The hexamer building interface (grey line) formed from the packing of SB and NB domains from adjoining dimers. The residues that pack at the interface for cUGDH are shown first, with the equivalent residues in hUGDH shown second and italicized. Hydrogen bonds (dashed lines) and van der Waals interactions (feathered lines) are depicted. Allosteric switch residues that interact in the interface are colored magenta. Yellow highlighting identifies residues that are present in cUGDH but missing in hUGDH.

Table S3.1: UGDH phyla and associated Genbank/Uniprot ID's that were used in the phylogenetic analysis.

Phylum	Genbank/Uniprot ID	Species	Kingdom	Clade Number
Chordata	O60701.1	Homo sapiens	Metazoa	1
Arthropoda	AAC97125.1	Drosophila melanogaster	Metazoa	1
Echinodermata	XP_011683830.1	Strongylocentrotus purpuratus	Metazoa	1
Hemichordata	XP_002741382.1	Saccoglossus kowalevskii	Metazoa	1
Nematoda	NP_505730.1	Caenorhabditis elegans	Metazoa	2
Placozoa	XP_002113339.1	Trichoplax adhaerens	Metazoa	2
Porifera	XP_003383497.1	Amphimedon queenslandica	Metazoa	2
Platyhelminthes	CDS17287.1	Echinococcus granulosus	Metazoa	3
Capsaspora	XP_004365811.1	Capsaspora owczarzaki ATCC 30864	Protozoa	3
Euglenozoa	XP_813451.1	Trypanosoma cruzi strain CL	Protozoa	4
Heterolobosea	XP_002676546.1	Naegleria gruberi	Protozoa	4
Apicomplexa	XP_627000.1	Cryptosporidium parvum Iowa II	Protozoa	4
Streptophyta	NP_197053.1	Arabidopsis thaliana	Viridiplantae	5
Chlorophyta	XP_001703656.1	Chlamydomonas reinhardtii	Viridiplantae	5
Alveolata	XP_002766755.1	Perkinsus marinus ATCC 50983	Protozoa	5
Bacillariophyta	GAX09934.1	Fistulifera solaris	Protozoa	5
Phaeophyceae	CBJ28343.1	Ectocarpus siliculosus	Protozoa	5
Stramenopiles	XP_012897639.1	Blastocystis hominis	Protozoa	5
Parabasalia	XP_001310931.1	Trichomonas vaginalis G3	Protozoa	6
Cyanobacteria	WP_011818811.1	Prochlorococcus marinus	Bacteria	6
Lentisphaerae	WP_007280104.1	Lentisphaera araneosa	Bacteria	6
Verrucomicrobia	PHX69165.1	Opitutae bacterium	Bacteria	6
Ascomycota	RAQ69836.1	Aspergillus flavus	Fungi	7
Basidiomycota	XP_025347162.1	Pseudomicrostroma glucosiphilum	Fungi	7
Chytridiomycota	EGF83877.1	Batrachochytrium dendrobatidis JAM81	Fungi	7
Intramacronucleata	CDW89424.1	Stylonychia lemnae	Protozoa	8
Bacteroidetes	WP_014682690.1	Solitalea canadensis	Bacteria	8
Fibrobacteres	WP_022637184.1	Chitinivibrio alkaliphilus	Bacteria	8
Firmicutes	P0C0F4.1	Streptococcus pyogenes	Bacteria	8
Planctomycetes	WP_082839084.1	Gemmata sp. SH-PL17	Bacteria	8
Proteobacteria	WP_013900935.1	Ramlibacter tataouinensis	Bacteria	8

CHAPTER 4

CONCLUSIONS

4.1 Broad Implications

Approximately 44% of the human proteome contains intrinsically disordered (ID) peptide segments that are >30 residues in length[32], many of which are in proteins that have disease relevance.[198,200,385-387] However, most ID segments have not yet been functionally annotated.[22,33,36] We have recently shown that the entropic force (i.e.; net pulling force) of an ID segment can shape the conformational landscape of a protein to tune function.[81] This mechanism of an "entropic rectifier" depends only on the length, not the sequence, of the ID segment.[81] Given the absence of sequence and structural constraints on an entropic rectifier, it is an easily acquired adaptation. Thus, it is likely that evolution utilizes entropic rectifiers (ID segments) to modify the energy landscape and function of proteins (Figure 4.1). This sequence-independent mechanism of "entropic rectification" may explain the persistence of low complexity ID segments in the proteome.[32,81]

Presented below are pieces of evidence from the literature that support the idea of evolution altering the length of ID segments to rectify the energy landscape of proteins (Figure 4.1):

1) There is transproteomic evidence that thermophilic organisms contain proteins with shorter loops than those found in their mesophilic homologs.[388] Thermophilic proteins have adapted to function at high temperatures (e.g.; ~40-122 °C), and as a result, they are more thermally stable than their mesophilic homologs.[58,389,390] In thermophilic environments, the increased reservoir of

heat (k_BT) would subject protein structures to larger fluctuations, which could impair function or induce unfolding. In the latter case, this would increase the propensity for proteins to aggregate. Because of the cytotoxic and wasteful nature associated with protein aggregation[254,391-394], thermophilic protein structures evolved to become tightly packed.[395-401] In other words, the packing interactions between the atoms that make up the core of thermophilic proteins has likely been optimized to counteract the increase in thermal fluctuations (k_BT). However, there is some debate on whether tighter packing is the source of increased thermostability.[402-404] An alternative mechanism suggests that increased thermal stability is correlated with an increase in the number of internal hydrogen bonds and electrostatic interactions.[405] Both of these thermal adaptation strategies could originate from changes to the length of surface loops (disordered segments). Decreasing the length of a loop would bring secondary structural elements into closer contact (i.e.; tighter packing), and it could also introduce new interactions (e.g.; hydrogen bonds).[388] Because loop regions facilitate protein motion[22,33,75,144,390], shortening loops would reduce the conformational entropy of a protein structure.

The loop deletion mechanism utilized by evolution to increase thermostability[388] is supported by several protein engineering studies.[287,289,290,292] These studies show that reducing loop length increases the thermal stability of some proteins.[287,289,290,292,293] Reducing loop length may be an evolutionary strategy for increasing the thermostability of proteins. This mechanism would decrease the conformational entropy of a protein, which should shift the free energy landscape to favor the 'native', low-energy folded-state. However, this mechanism would require the protein to still maintain enough flexibility for function. Nonetheless, these observations and studies suggest that evolution alters the length of ID segments to rectify the energy landscape of proteins.

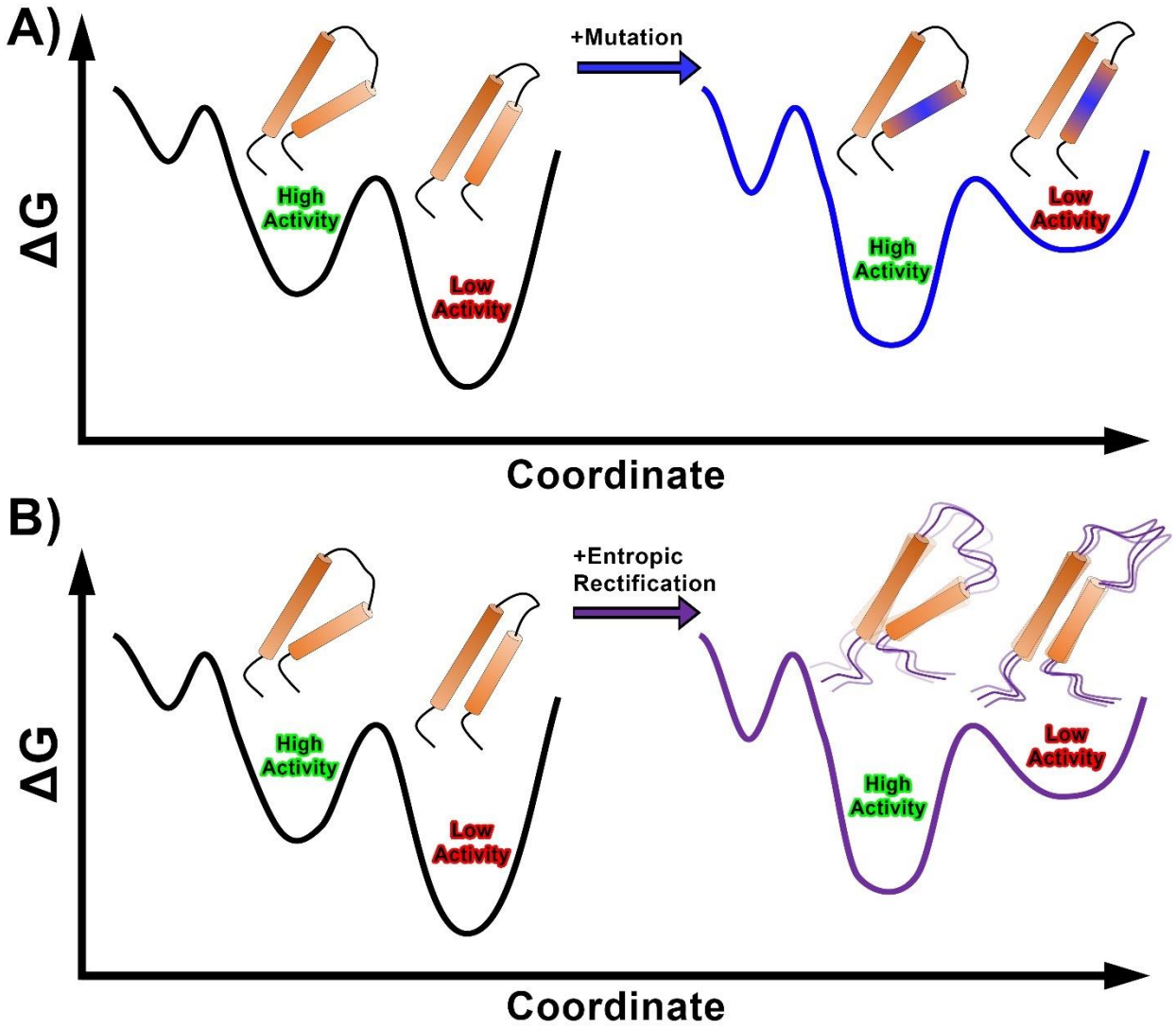


Figure 4.1: The evolutionary mechanisms for re-modeling the energy landscape and function of proteins. (A) The hypothetical 2-dimensional free energy landscape for a protein (orange cylinders as helices and black lines as loops) is depicted. The protein adopts two distinct conformations with different relative free energies that have distinct activities (high and low). *Left:* The distribution of states favors the low activity state of the protein. *Right*: A mutation to the structured region of the protein (blue) redistributes the conformational ensemble of states to favor the high activity state of the protein. Any mutation made to the structure of a protein will, in theory, perturb the free energy landscape.[64,68,71,74-76,115,117,118,154,156,159,162,171,178] In the traditional view of protein evolution, mutations that alter function cannot significantly compromise the folded 'native' state of the protein structure. In other words, the selection for mutations that alter function is restricted by the necessity to maintain the integrity of the folded structure.[68,74-76,124,256] **(B)** *Left:* The identical 2-dimensional free energy landscape from panel A is depicted. *Right:* The lengthening of disordered loops (purple) entropically rectifies the free energy landscape by altering the dynamics of the protein (orange cylinders). This redistributes the conformational ensemble of states to favor the high activity state of the protein. Disordered segments have low sequence complexity[21,22,36,39,43,45] and are capable of perturbing the functional dynamics of

proteins.[33,75,81,162,210,326,333] Therefore, the evolution of the free energy landscape for some proteins may occur through changes to the length of disordered segments. Importantly, this "entropic rectification" mechanism has no specific sequence constraints, and as a result, it could be an easily acquired adaptation for tuning protein function.[81] Credits to Zac Wood for the original idea behind this re-created figure.

2) Enzymes that are adapted to function in cold environments (psychrophilic enzymes) typically have longer disordered loop regions than their mesophilic homologs.[406-414] In order to maintain function (e.g.; rate of catalysis) at lower temperatures, psychrophilic enzymes might have adapted longer surface loops to increase their conformational heterogeneity. In fact, recent evidence suggests that increasing the dynamics of surface loops can facilitate the cold-adaptation of enzymes.[162,408] Increasing loop length can destabilize a protein by increasing its conformational entropy[288,408] (i.e.; a protein with longer surface loops should exhibit larger structural fluctuations). Therefore, it is possible that evolution rectifies the energy landscape of psychrophilic enzymes by extending surface loops. This mechanism would lead to an increase in the conformational entropy of psychrophilic enzymes and might be the source for maintaining function in cold environments. In other words, the likelihood of sampling functionally important conformations is lower in a cold environment because of the reduction in thermal energy (k_BT). To combat the reduction in thermal energy, cold-adapted enzymes might have evolved to become more dynamic through the extension of surface loops.[162,210,408]

4.2 Future Directions (Aim 1)

Understanding the functional capacity of entropic rectifiers would be extremely valuable in protein engineering efforts (e.g.; designing enzymes to be more efficient or thermally stable).[33,287] A systematic study with a small protein (less than 20 kDa) investigating the functional effects upon the elongation or truncation of native, disordered loops would be a great start in identifying the 'rules' associated with entropic rectifiers. Examining the changes in dynamics of

those mutants that show functional alterations would reveal mechanistic insight into how the "entropic quake" of an ID segment propagates through a protein structure. Because of the size of the target protein, both NMR and MD could be utilized to investigate the mechanism of entropic rectifiers at an atomistic level of detail. The proposed study is outlined in more detail below (Aim 1):

Aim 1: Determine the functional capacity of entropic rectifiers. Working hypothesis: The entropic force generated by ID segments is dependent on both location and length. Future work should focus on identifying where and how the entropic forces of ID segments can be utilized to modulate protein function. By using a small enzyme that possesses native disordered regions, we could systematically assess the importance of ID segment length and location with regards to protein function. In other words, we could elongate and truncate native ID-segments and assay for functional changes (e.g.; catalytic turnover or binding affinities for substrate/inhibitors). Those mutants that display functional alterations would then be subjected to chemical/temperature denaturation experiments to assess for changes in protein stability. Using NMR relaxation experiments and MD simulations, we could then evaluate how the entropic force perturbs the dynamics of the protein structure across a wide range of timescales. Project workflow and specific experiments are outlined below:

Identify systems for studying the effect of loop elongation/truncation on the function and dynamics of a protein: **1)** Identify a small protein (less than 20 kDa) that is easy to express and purify recombinantly; **2)** The target protein ideally has multiple crystal structures or NMR solution structures available with different substrate(s); **3)** The target protein contains some natively disordered loop regions (internal and/or termini). Identify and name disordered regions: *i)* Run disorder predictors (IUPred or PONDR are my favorites) on the amino acid sequence to help

identify regions of disorder; *ii)* Utilize crystal-structures (missing density or high *B*-factors), NMR solutions structures, or information from NMR relaxation experiments to identify loops whose dynamics are perturbed upon ligand binding. In this way, the use of *a priori* knowledge can help avoid wasting any time on targeting loops that do not appear important for protein function; **4)** The target protein is easy to assay for functional changes. Ideally, there are at least two different ways to measure changes in function from loop mutagenesis. Suggested experiments for elucidating functional changes quickly and with high accuracy: *i)* Fluorescence equilibrium binding, monitoring the change in intrinsic tryptophan fluorescence that occurs during the titration of ligand; *ii)* Steady-state kinetics (absorbance based, monitoring product formation or substrate depletion); *iii)* Isothermal titration calorimetry with ligand(s) (note: requires high concentrations of protein, so if protein production yields are low, I would advise against the use of this technique). **5)** The target protein is a reversible folder (two-state folder). Use CD/Fluorescence/DSC to assay for stability changes in the mutants that showed functional alterations upon loop elongation or truncation. **6)** Those targets that show clear relationships between protein function and stability are examined at the atomistic level of detail to investigate how the ID segment changes the protein dynamics. Using NMR with the isotopically labelled protein, one could pursue solving NMR solution structures of the wild-type protein and the mutant(s) of interest. NMR relaxation experiments that access different timescale motions (^{1}H-^{15}N NOE experiments for sub-nanosecond motions, CPMG Relaxation and residual dipolar coupling measurements to access the slower, nano-to-millisecond slower timescales) would be useful.[55,67,134,155,156,230] These experiments would need to be done in the presence and absence of ligand, with both the wild-type and the mutant(s) of interest. This work could unveil how the ID segment modulates the dynamics and function of the protein at an atomistic level of detail. These experiments and analyses would be cumbersome,

but they are exciting and have great potential to yield mechanistic insight into how an entropic rectifier transmits its "entropic quake" across the protein structure to alter function. Molecular dynamics simulations (currently can access up to the microsecond timescale) would be useful in elucidating the atomistic mechanism of an entropic rectifier[164,415,416], especially if the simulation utilizes restraints from NMR solution structures. Monte-carlo simulations of the disordered region at different lengths could also be useful. With this method, the number of states that the free disordered region can access (not connected to the protein) can be compared to the number of possible states it can access within the confined geometry of the structured protein. This would allow for an estimation of the entropic cost or gain induced by elongating or truncating a loop region, respectively. Finally, *a priori* knowledge of established, published methods will be useful in identifying a viable protein system to study the effect of entropic rectifiers at an atomistic level of detail. There is a saying at MIT *"One hour of reading will save you one hundred hours at the bench"*.

4.3 Future Directions (Aim 2)

The functions of some proteins might be entropically rectified by post-translational modifications (PTMs) to their disordered region.[124,200,211,417,418] These modifications often result in a disorder-to-order transition[10,13,22,33,44,197,201,209], which would reduce the overall length of the disordered segment and subsequently the entropic force it generates. However, a recent study has shown that a PTM can increase the end-to-end length of a disordered segment, which may increase the entropic force it exerts.[417]

Aim 2. Determine the effect of post-translational modifications on the entropic force generated by ID segments. Working hypothesis: The length of an ID segment is altered when it is post-translationally modified, which changes the entropic force it can exert. Post-translational

modifications are a common mechanism used to regulate many enzymes.[10,22,33,44,124,197,199,200,202,206-208,213] Recently it has been shown that a PTM can increase the end-to-end length of a disordered peptide, possibly allowing it to access more degrees of freedom (i.e.; increased entropy).[417] In accord with our results[81] and simulation studies[297,298,300,301] on the entropic cost of confinement, this PTM-induced change in the length of an ID segment would increase the amount of entropic force it produces. Therefore, acting through the entropic force mechanism, protein function could be tuned via the PTM-induced changes to the conformational freedom of a disordered region. In both the modified and unmodified states, the protein of interest should be examined using the above workflow from Aim 1, starting at step 4.

REFERENCES

1 Berg, J. M., Tymoczko, J. L., Gatto, G. J. & Stryer, L. *Biochemistry*. Eighth edition. edn, (W.H. Freeman & Company, a Macmillan Education Imprint, 2015).

2 Fersht, A. *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*. (World Scientific, 2017).

3 Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

4 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).

5 Pauling, L. & Corey, R. B. Atomic Coordinates and Structure Factors for 2 Helical Configurations of Polypeptide Chains. *P Natl Acad Sci USA* **37**, 235-240, doi:DOI 10.1073/pnas.37.5.235 (1951).

6 Pauling, L., Corey, R. B. & Branson, H. R. The Structure of Proteins - 2 Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *P Natl Acad Sci USA* **37**, 205-211, doi:DOI 10.1073/pnas.37.4.205 (1951).

7 Perutz, M. F. *et al.* Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **185**, 416-422 (1960).

8 Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **41**, 6573-6582, doi:10.1021/bi012159+ (2002).

9 Dunker, A. K. *et al.* Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59 (2001).

10 Dunker, A. K. & Obradovic, Z. The protein trinity--linking function and disorder. *Nat Biotechnol* **19**, 805-806, doi:10.1038/nbt0901-805 (2001).

11 Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr Opin Struc Biol* **18**, 756-764, doi:10.1016/j.sbi.2008.10.002 (2008).

12 Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Bio* **6**, 197-208, doi:10.1038/nrm1589 (2005).

13 Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing Protein Intrinsic Disorder. *Chemical Reviews* **114**, 6561-6588, doi:10.1021/cr400514h (2014).

14 Manalan, A. S. & Klee, C. B. Activation of calcineurin by limited proteolysis. *Proc Natl Acad Sci U S A* **80**, 4291-4295, doi:10.1073/pnas.80.14.4291 (1983).

15 Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553-584, doi:10.1146/annurev-biochem-072711-164947 (2014).

16 Pullen, R. A., Jenkins, J. A., Tickle, I. J., Wood, S. P. & Blundell, T. L. The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen. *Mol Cell Biochem* **8**, 5-20 (1975).

17 Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021-1025, doi:10.1038/nature05858 (2007).

18 Tompa, P. Unstructural biology coming of age. *Curr Opin Struc Biol* **21**, 419-425, doi:10.1016/j.sbi.2011.03.012 (2011).

19 Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends in Biochemical Sciences* **37**, 509-516, doi:10.1016/j.tibs.2012.08.004 (2012).

20 Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell B* **43**, 1090-1103, doi:10.1016/j.biocel.2011.04.001 (2011).

21 Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415-427 (2000).

22 van der Lee, R. *et al.* Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **114**, 6589-6631, doi:10.1021/cr400525m (2014).

23 Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293**, 321-331, doi:DOI 10.1006/jmbi.1999.3110 (1999).

24 Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Bio* **16**, 18-29, doi:10.1038/nrm3920 (2015).

25 Spolar, R. S. & Record, M. T., Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784 (1994).

26 Lewis, M. *et al.* Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-1254, doi:DOI 10.1126/science.271.5253.1247 (1996).

27 Cary, P. D., Moss, T. & Bradbury, E. M. High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur J Biochem* **89**, 475-482 (1978).

28 Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. Protein Disk of Tobacco Mosaic-Virus at 2.8-a Resolution Showing Interactions within and between Subunits. *Nature* **276**, 362-368, doi:DOI 10.1038/276362a0 (1978).

29 Alber, T., Gilbert, W. A., Ponzi, D. R. & Petsko, G. A. The Role of Mobility in the Substrate Binding and Catalytic Machinery of Enzymes. *Ciba F Symp* **93**, 4-18 (1983).

30 Bode, W. Transition of Bovine Trypsinogen to a Trypsin-Like State Upon Strong Ligand-Binding .2. Binding of the Pancreatic Trypsin-Inhibitor and of Isoleucine-Valine and of Sequentially Related Peptides to Trypsinogen and to P-Guanidinobenzoate-Trypsinogen. *Journal of Molecular Biology* **127**, 357-374, doi:Doi 10.1016/0022-2836(79)90227-4 (1979).

31 Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9 Suppl 2**, S1, doi:10.1186/1471-2164-9-S2-S1 (2008).

32 Oates, M. E. *et al.* D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* **41**, D508-516, doi:10.1093/nar/gks1226 (2013).

33 Papaleo, E. *et al.* The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem Rev* **116**, 6391-6423, doi:10.1021/acs.chemrev.5b00623 (2016).

34 Radivojac, P. *et al.* Intrinsic disorder and functional proteomics. *Biophys J* **92**, 1439-1456, doi:10.1529/biophysj.106.094045 (2007).

35 Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N. & Dunker, A. K. Intrinsic disorder in the Protein Data Bank. *J Biomol Struct Dyn* **24**, 325-342, doi:10.1080/07391102.2007.10507123 (2007).

36 He, B. *et al.* Predicting intrinsic disorder in proteins: an overview. *Cell Res* **19**, 929-949, doi:10.1038/cr.2009.87 (2009).

37 Oldfield, C. J. *et al.* Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989-2000, doi:10.1021/bi047993o (2005).

38 Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434, doi:10.1093/bioinformatics/bti541 (2005).

39 Obradovic, Z. *et al.* Predicting intrinsic disorder from amino acid sequence. *Proteins* **53 Suppl 6**, 566-572, doi:10.1002/prot.10532 (2003).

40 Linding, R. *et al.* Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453-1459 (2003).

41 Nielsen, J. T. & Mulder, F. A. A. Quality and bias of protein disorder predictors. *Sci Rep* **9**, 5137, doi:10.1038/s41598-019-41644-w (2019).

42 Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**, 3701-3708, doi:10.1093/nar/gkg519 (2003).

43 Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38-48 (2001).

44 Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim Biophys Acta* **1804**, 1231-1264, doi:10.1016/j.bbapap.2010.01.017 (2010).

45 Tompa, P. Intrinsically unstructured proteins. *Trends Biochem Sci* **27**, 527-533 (2002).

46 Fontana, A. *et al.* Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **25**, 1847-1851 (1986).

47 Disfani, F. M. *et al.* MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75-83, doi:10.1093/bioinformatics/bts209 (2012).

48 Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E. & Zambonin, M. Probing the partly folded states of proteins by limited proteolysis. *Fold Des* **2**, R17-26, doi:10.1016/S1359-0278(97)00010-2 (1997).

49 Tran, H. T., Mao, A. & Pappu, R. V. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc* **130**, 7380-7392, doi:10.1021/ja710446s (2008).

50 Li, M., Cao, H., Lai, L. & Liu, Z. Disordered linkers in multidomain allosteric proteins: Entropic effect to favor the open state or enhanced local concentration to favor the closed state? *Protein Sci* **27**, 1600-1610, doi:10.1002/pro.3475 (2018).

51 Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**, 5129-5148, doi:10.1111/j.1742-4658.2005.04948.x (2005).

52 Fontana, A. *et al.* Probing protein structure by limited proteolysis. *Acta Biochim Pol* **51**, 299-321, doi:035001299 (2004).

53 Radivojac, P. *et al.* Protein flexibility and intrinsic disorder. *Protein Sci* **13**, 71-80, doi:10.1110/ps.03128904 (2004).

54 Arora, K. & Brooks Iii, C. L., 3rd. Functionally important conformations of the Met20 loop in dihydrofolate reductase are populated by rapid thermal fluctuations. *J Am Chem Soc* **131**, 5642-5647, doi:10.1021/ja9000135 (2009).

55 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972, doi:10.1038/nature06522 (2007).

56 Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **254**, 1598-1603, doi:DOI 10.1126/science.1749933 (1991).

57 Careri, G., Fasella, P. & Gratton, E. Enzyme dynamics: the statistical physics approach. *Annu Rev Biophys Bioeng* **8**, 69-97, doi:10.1146/annurev.bb.08.060179.000441 (1979).

58 Tang, K. E. & Dill, K. A. Native protein fluctuations: the conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. *J Biomol Struct Dyn* **16**, 397-411, doi:10.1080/07391102.1998.10508256 (1998).

59 Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. Slaving: solvent fluctuations dominate protein dynamics and functions. *Proc Natl Acad Sci U S A* **99**, 16047-16051, doi:10.1073/pnas.212637899 (2002).

60 Brooks, C. L., 3rd & Karplus, M. Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. *J Mol Biol* **208**, 159-181 (1989).

61 Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280**, 558-563 (1979).

62 Gavish, B. & Werber, M. M. Viscosity-dependent structural fluctuations in enzyme catalysis. *Biochemistry* **18**, 1269-1275 (1979).

63 Berendsen, H. J. C. & Hayward, S. Collective protein dynamics in relation to function. *Curr Opin Struc Biol* **10**, 165-169, doi:Doi 10.1016/S0959-440x(00)00061-0 (2000).

64 Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638-1642, doi:10.1126/science.1130258 (2006).

65 McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).

66 Karplus, M. & McCammon, J. A. Dynamics of proteins: elements and function. *Annu Rev Biochem* **52**, 263-300, doi:10.1146/annurev.bi.52.070183.001403 (1983).

67 Kay, L. E. Protein dynamics from NMR. *Nat Struct Biol* **5 Suppl**, 513-517, doi:10.1038/755 (1998).

68 Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203-207, doi:10.1126/science.1169375 (2009).

69 Glowacki, D. R., Harvey, J. N. & Mulholland, A. J. Taking Ockham's razor to enzyme dynamics and catalysis. *Nat Chem* **4**, 169-176, doi:10.1038/nchem.1244 (2012).

70 Kamerlin, S. C. & Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78**, 1339-1375, doi:10.1002/prot.22654 (2010).

71 Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234-238, doi:10.1126/science.1198542 (2011).

72 Kumar, S., Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **9**, 10-19, doi:10.1110/ps.9.1.10 (2000).

73 Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669-673, doi:10.1038/nature08615 (2009).

74 Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789-796, doi:10.1038/nchembio.232 (2009).

75 Campbell, E. *et al.* The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol* **12**, 944-950, doi:10.1038/nchembio.2175 (2016).

76 James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* **28**, 361-368, doi:10.1016/S0968-0004(03)00135-X (2003).

77 Khersonsky, O. & Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem* **79**, 471-505, doi:10.1146/annurev-biochem-030409-143718 (2010).

78 Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **35**, 539-546, doi:10.1016/j.tibs.2010.04.009 (2010).

79 Tzeng, S. R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **488**, 236-240, doi:10.1038/nature11271 (2012).

80 Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U. & Nidetzky, B. Structure and mechanism of human UDP-glucose 6-dehydrogenase. *J Biol Chem* **286**, 23877-23887, doi:10.1074/jbc.M111.234682 (2011).

81 Keul, N. D. *et al.* The entropic force generated by intrinsically disordered segments tunes protein function. *Nature* **563**, 584-588, doi:10.1038/s41586-018-0699-5 (2018).

82 Monod, J. & Jacob, F. General Conclusions - Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb Sym* **26**, 389-&, doi:Doi 10.1101/Sqb.1961.026.01.048 (1961).

83 Changeux, J. P. 50th anniversary of the word "allosteric". *Protein Sci* **20**, 1119-1124, doi:10.1002/pro.658 (2011).

84 Monod, J., Wyman, J. & Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol* **12**, 88-118 (1965).

85 Monod, J., Changeux, J. P. & Jacob, F. Allosteric proteins and cellular control systems. *J Mol Biol* **6**, 306-329 (1963).

86 Monod, J. *Chance and necessity; an essay on the natural philosophy of modern biology*. (Vintage Books, 1972).

87 Bohr, C., Hasselbalch, K. & Krogh, A. About a new biological relation of high importance that the blood carbonic acid tension exercises on its oxygen binding. *Skand Arch Physiol* **16**, 402-412, doi:DOI 10.1111/j.1748-1716.1904.tb01382.x (1904).

88 Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Sci* **17**, 1295-1307, doi:10.1110/ps.03259908 (2008).

89 Changeux, J. P. Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys* **41**, 103-133, doi:10.1146/annurev-biophys-050511-102222 (2012).

90 Changeux, J. P. & Edelstein, S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep* **3**, 19, doi:10.3410/B3-19 (2011).

91 Levitzki, A. & Koshland, D. E., Jr. Negative cooperativity in regulatory enzymes. *Proc Natl Acad Sci U S A* **62**, 1121-1128, doi:10.1073/pnas.62.4.1121 (1969).

92 Rubin, M. M. & Changeux, J. P. On the nature of allosteric transitions: implications of non-exclusive ligand binding. *J Mol Biol* **21**, 265-274 (1966).

93 Laskowski, R. A., Gerick, F. & Thornton, J. M. The structural basis of allosteric regulation in proteins. *Febs Lett* **583**, 1692-1698, doi:10.1016/j.febslet.2009.03.019 (2009).

94 Koshland, D. E., Jr. The structural basis of negative cooperativity: receptors and enzymes. *Curr Opin Struct Biol* **6**, 757-761 (1996).

95 Pauling, L. The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation. *Proc Natl Acad Sci U S A* **21**, 186-191, doi:10.1073/pnas.21.4.186 (1935).

96 Perutz, M. F., Bolton, W., Diamond, R., Muirhead, H. & Watson, H. C. Structure of Haemoglobin. An X-Ray Examination of Reduced Horse Haemoglobin. *Nature* **203**, 687-690 (1964).

97 Henri, V. *General Laws for the Action of Diastases*, University of Göttingen, (1903).

98 Michaelis, L., Menten, M. L., Johnson, K. A. & Goody, R. S. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* **50**, 8264-8269, doi:10.1021/bi201284u (2011).

99 Hill, A. V. The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. I. *Biochem J* **7**, 471-480, doi:10.1042/bj0070471 (1913).

100 Weiss, J. N. The Hill equation revisited: uses and misuses. *Faseb Journal* **11**, 835-841 (1997).

101 Koshland, D. E., Jr., Nemethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365-385 (1966).

102 Koshland, D. E., Jr. Enzyme flexibility and enzyme action. *J Cell Comp Physiol* **54**, 245-258 (1959).

103 Sullivan, S. M. & Holyoak, T. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *P Natl Acad Sci USA* **105**, 13829-13834, doi:10.1073/pnas.0805364105 (2008).

104 Wirsching, P., Ashley, J. A., Benkovic, S. J., Janda, K. D. & Lerner, R. A. An unexpectedly efficient catalytic antibody operating by ping-pong and induced fit mechanisms. *Science* **252**, 680-685 (1991).

105 Changeux, J. P. & Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **308**, 1424-1428, doi:10.1126/science.1108595 (2005).

106 Changeux, J. P., Thiery, J., Tung, Y. & Kittel, C. On the cooperativity of biological membranes. *Proc Natl Acad Sci U S A* **57**, 335-341, doi:10.1073/pnas.57.2.335 (1967).

107 Wlodarski, T. & Zagrovic, B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci U S A* **106**, 19346-19351, doi:10.1073/pnas.0906966106 (2009).

108 Hammes, G. G., Chang, Y. C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci U S A* **106**, 13737-13741, doi:10.1073/pnas.0907195106 (2009).

109 Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331-339, doi:10.1038/nature13001 (2014).

110 Oakley, A. J., Lo Bello, M., Ricci, G., Federici, G. & Parker, M. W. Evidence for an induced-fit mechanism operating in pi class glutathione transferases. *Biochemistry* **37**, 9912-9917, doi:10.1021/bi980323w (1998).

111 Yousef, M. S. *et al.* Induced fit in guanidino kinases--comparison of substrate-free and transition state analog structures of arginine kinase. *Protein Sci* **12**, 103-111, doi:10.1110/ps.0226303 (2003).

112 Rodnina, M. V. & Wintermeyer, W. Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem Sci* **26**, 124-130 (2001).

113 Perutz, M. F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726-739 (1970).

114 Burgen, A. S. Conformational changes and drug action. *Fed Proc* **40**, 2723-2728 (1981).

115 Nussinov, R. & Tsai, C. J. Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struc Biol* **30**, 17-24, doi:10.1016/j.sbi.2014.11.005 (2015).

116 Tsai, C. J., Ma, B. & Nussinov, R. Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci U S A* **96**, 9970-9972, doi:10.1073/pnas.96.18.9970 (1999).

117 Ma, B., Kumar, S., Tsai, C. J. & Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng* **12**, 713-720 (1999).

118 Tsai, C. J., Kumar, S., Ma, B. & Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci* **8**, 1181-1190, doi:10.1110/ps.8.6.1181 (1999).

119 Weber, G. Ligand binding and internal equilibria in proteins. *Biochemistry* **11**, 864-878 (1972).

120 Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**, 7524-7528, doi:10.1073/pnas.84.21.7524 (1987).

121 Tsai, C. J. & Nussinov, R. A Unified View of "How Allostery Works". *Plos Computational Biology* **10**, doi:10.1371/journal.pcbi.1003394 (2014).

122 Malmendal, A., Evenas, J., Forsen, S. & Akke, M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. *J Mol Biol* **293**, 883-899, doi:10.1006/jmbi.1999.3188 (1999).

123 Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429-2433, doi:10.1126/science.291.5512.2429 (2001).

124 Smock, R. G. & Gierasch, L. M. Sending signals dynamically. *Science* **324**, 198-203, doi:10.1126/science.1169377 (2009).

125 Berger, C. *et al.* Antigen recognition by conformational selection. *Febs Lett* **450**, 149-153 (1999).

126 Clark, A. C. Caspase Allostery and Conformational Selection. *Chem Rev* **116**, 6666-6706, doi:10.1021/acs.chemrev.5b00540 (2016).

127 Pozzi, N., Vogt, A. D., Gohara, D. W. & Di Cera, E. Conformational selection in trypsin-like proteases. *Curr Opin Struc Biol* **22**, 421-431, doi:10.1016/j.sbi.2012.05.006 (2012).

128 Gsponer, J. *et al.* A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure* **16**, 736-746, doi:10.1016/j.str.2008.02.017 (2008).

129 Vogt, A. D., Pozzi, N., Chen, Z. W. & Di Cera, E. Essential role of conformational selection in ligand binding. *Biophys Chem* **186**, 13-21, doi:10.1016/j.bpc.2013.09.003 (2014).

130 Bahar, I., Chennubhotla, C. & Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* **17**, 633-640, doi:10.1016/j.sbi.2007.09.011 (2007).

131 Gunasekaran, K., Ma, B. Y. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins-Structure Function and Bioinformatics* **57**, 433-443, doi:10.1002/prot.20232 (2004).

132 Junker, J. P., Ziegler, F. & Rief, M. Ligand-Dependent Equilibrium Fluctuations of Single Calmodulin Molecules. *Science* **323**, 633-637, doi:10.1126/science.1166191 (2009).

133 Nevo, R. *et al.* A molecular switch between alternative conformational states in the complex of Ran and importin beta 1. *Nature Structural Biology* **10**, 553-557, doi:10.1038/nsb940 (2003).

134 Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078-1082, doi:10.1038/nature06232 (2007).

135 Fenwick, R. B. *et al.* Solution structure and dynamics of the small GTPase RalB in its active conformation: significance for effector protein binding. *Biochemistry* **48**, 2192-2206, doi:10.1021/bi802129d (2009).

136 Saitoh, T. *et al.* Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *Embo J* **26**, 4777-4787, doi:10.1038/sj.emboj.7601888 (2007).

137 Daily, M. D. & Gray, J. J. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLoS Comput Biol* **5**, e1000293, doi:10.1371/journal.pcbi.1000293 (2009).

138 Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838-844, doi:10.1038/nature06410 (2007).

139 Bosshard, H. R. Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci* **16**, 171-173 (2001).

140 Debye, P. Interference of x rays and heat movement. *Ann Phys-Berlin* **43**, 49-95 (1913).

141 Waller, I. On the question of the influence of thermal motion on the interference of X-rays. *Z Phys* **17**, 398-408, doi:Doi 10.1007/Bf01328696 (1923).

142 Vihinen, M., Torkkila, E. & Riikonen, P. Accuracy of Protein Flexibility Predictions. *Proteins-Structure Function and Bioinformatics* **19**, 141-149, doi:DOI 10.1002/prot.340190207 (1994).

143 Carugo, O. & Argos, P. Correlation between side chain mobility and conformation in protein structures. *Protein Engineering* **10**, 777-787, doi:DOI 10.1093/protein/10.7.777 (1997).

144 Halle, B. Flexibility and packing in proteins. *Proc Natl Acad Sci U S A* **99**, 1274-1279, doi:10.1073/pnas.032522499 (2002).

145 Tang, K. E. S. & Dill, K. A. How experiments see fluctuations of native proteins: Perspective from an exact model. *Int J Quantum Chem* **75**, 147-164, doi:Doi 10.1002/(Sici)1097-461x(1999)75:3<147::Aid-Qua5>3.0.Co;2-P (1999).

146 Zhang, X. J., Wozniak, J. A. & Matthews, B. W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol* **250**, 527-552, doi:10.1006/jmbi.1995.0396 (1995).

147 Parthasarathy, S. & Murthy, M. R. N. Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Engineering* **13**, 9-13, doi:DOI 10.1093/protein/13.1.9 (2000).

148 Matthews, B. W. Structural and genetic analysis of protein stability. *Annu Rev Biochem* **62**, 139-160, doi:10.1146/annurev.bi.62.070193.001035 (1993).

149 Bahar, I., Atilgan, A. R., Demirel, M. C. & Erman, B. Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. *Phys Rev Lett* **80**, 2733-2736, doi:DOI 10.1103/PhysRevLett.80.2733 (1998).

150 Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins* **46**, 105-109 (2002).

151 Hilser, V. J. An Ensemble View of Allostery. *Science* **327**, 653-654, doi:10.1126/science.1186121 (2010).

152 Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155 (1990).

153 Cooper, A. & Dryden, D. T. F. Allostery without Conformational Change - a Plausible Model. *Eur Biophys J Biophy* **11**, 103-109, doi:Doi 10.1007/Bf00276625 (1984).

154 Tsai, C. J., del Sol, A. & Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J Mol Biol* **378**, 1-11, doi:10.1016/j.jmb.2008.02.034 (2008).

155 Wand, A. J. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol* **23**, 75-81, doi:10.1016/j.sbi.2012.11.005 (2013).

156 Sekhar, A. & Kay, L. E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci U S A* **110**, 12867-12874, doi:10.1073/pnas.1305688110 (2013).

157 Tlapak-Simmons, V. L. & Reinhart, G. D. Obfuscation of allosteric structure-function relationships by enthalpy-entropy compensation. *Biophys J* **75**, 1010-1015, doi:10.1016/S0006-3495(98)77589-7 (1998).

158 Tsai, C. J., Del Sol, A. & Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol Biosyst* **5**, 207-216, doi:10.1039/b819720b (2009).

159 Kern, D. & Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* **13**, 748-757 (2003).

160 Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A* **106**, 18249-18254, doi:10.1073/pnas.0904492106 (2009).

161 Guo, J. & Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chem Rev* **116**, 6503-6515, doi:10.1021/acs.chemrev.5b00590 (2016).

162 Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J. & Hilser, V. J. Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324-328, doi:10.1038/s41586-018-0183-2 (2018).

163 Holliday, M. J., Camilloni, C., Armstrong, G. S., Vendruscolo, M. & Eisenmesser, E. Z. Networks of Dynamic Allostery Regulate Enzyme Function. *Structure* **25**, 276-286, doi:10.1016/j.str.2016.12.003 (2017).

164 Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **102**, 6679-6685, doi:10.1073/pnas.0408930102 (2005).

165 Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **23**, 105-116, doi:10.1002/jmr.961 (2010).

166 Liu, J. & Nussinov, R. Energetic redistribution in allostery to execute protein function. *Proc Natl Acad Sci U S A* **114**, 7480-7482, doi:10.1073/pnas.1709071114 (2017).

167 Shen, T., Zong, C., Hamelberg, D., McCammon, J. A. & Wolynes, P. G. The folding energy landscape and phosphorylation: modeling the conformational switch of the NFAT regulatory domain. *FASEB J* **19**, 1389-1395, doi:10.1096/fj.04-3590hyp (2005).

168 Zimmermann, J. *et al.* Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proc Natl Acad Sci U S A* **103**, 13722-13727, doi:10.1073/pnas.0603282103 (2006).

169 Fenton, A. W. Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem Sci* **33**, 420-425, doi:10.1016/j.tibs.2008.05.009 (2008).

170 Freiburger, L. A. *et al.* Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. *Nat Struct Mol Biol* **18**, 288-294, doi:10.1038/nsmb.1978 (2011).

171 Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* **10**, 866-876, doi:10.1038/nrm2805 (2009).

172 Ma, B. & Nussinov, R. Protein dynamics: Conformational footprints. *Nat Chem Biol* **12**, 890-891, doi:10.1038/nchembio.2212 (2016).

173 Colombo, M. F., Rau, D. C. & Parsegian, V. A. Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science* **256**, 655-659 (1992).

174 Uversky, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* **22**, 693-724, doi:10.1002/pro.2261 (2013).

175 Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *P Natl Acad Sci USA* **104**, 8311-8315, doi:10.1073/pnas.0700323104 (2007).

176 Tantos, A., Han, K. H. & Tompa, P. Intrinsic disorder in cell signaling and gene transcription. *Mol Cell Endocrinol* **348**, 457-465, doi:10.1016/j.mce.2011.07.015 (2012).

177 Cortese, M. S., Uversky, V. N. & Dunker, A. K. Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol* **98**, 85-106, doi:10.1016/j.pbiomolbio.2008.05.007 (2008).

178 Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433-443, doi:10.1002/prot.20232 (2004).

179 Singh, G. P., Ganapathi, M. & Dash, D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* **66**, 761-765, doi:10.1002/prot.21281 (2007).

180 Schlessinger, A. *et al.* Protein disorder--a breakthrough invention of evolution? *Curr Opin Struct Biol* **21**, 412-418, doi:10.1016/j.sbi.2011.03.014 (2011).

181 Fuxreiter, M., Tompa, P. & Simon, I. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950-956, doi:10.1093/bioinformatics/btm035 (2007).

182 Latysheva, N. S., Flock, T., Weatheritt, R. J., Chavali, S. & Babu, M. M. How do disordered regions achieve comparable functions to structured domains? *Protein Science* **24**, 909-922, doi:10.1002/pro.2674 (2015).

183 Uversky, V. N. The multifaceted roles of intrinsic disorder in protein complexes. *Febs Lett* **589**, 2498-2506, doi:10.1016/j.febslet.2015.06.004 (2015).

184 Berg, O. G. & von Hippel, P. H. Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Biophys Chem* **14**, 131-160, doi:10.1146/annurev.bb.14.060185.001023 (1985).

185 Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**, 8868-8873, doi:10.1073/pnas.160259697 (2000).

186 Levy, Y., Onuchic, J. N. & Wolynes, P. G. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* **129**, 738-739, doi:10.1021/ja065531n (2007).

187 Chen, J. Intrinsically disordered p53 extreme C-terminus binds to S100B(betabeta) through "fly-casting". *J Am Chem Soc* **131**, 2088-2089 (2009).

188 Huang, Y. & Liu, Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J Mol Biol* **393**, 1143-1159, doi:10.1016/j.jmb.2009.09.010 (2009).

189 Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr Opin Struct Biol* **19**, 31-38, doi:10.1016/j.sbi.2008.12.003 (2009).

190 Zhou, H. X., Pang, X. & Lu, C. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Phys Chem Chem Phys* **14**, 10466-10476, doi:10.1039/c2cp41196b (2012).

191 Pontius, B. W. Close Encounters - Why Unstructured, Polymeric Domains Can Increase Rates of Specific Macromolecular Association. *Trends in Biochemical Sciences* **18**, 181-186, doi:Doi 10.1016/0968-0004(93)90111-Y (1993).

192 Umezawa, K., Ohnuki, J., Higo, J. & Takano, M. Intrinsic disorder accelerates dissociation rather than association. *Proteins-Structure Function and Bioinformatics* **84**, 1124-1133, doi:10.1002/prot.25057 (2016).

193 Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I. & Tompa, P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* **5**, 2985-2995, doi:10.1021/pr060101o (2006).

194 Haynes, C. *et al.* Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *Plos Computational Biology* **2**, 890-901, doi:ARTN e100 10.1371/journal.pcbi.0020100 (2006).

195 Kim, P. M., Sboner, A., Xia, Y. & Gerstein, M. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol* **4**, 179, doi:10.1038/msb.2008.16 (2008).

196 Magidovich, E., Orr, I., Fass, D., Abdu, U. & Yifrach, O. Intrinsic disorder in the C-terminal domain of the Shaker voltage-activated K+ channel modulates its interaction with scaffold proteins. *Proc Natl Acad Sci U S A* **104**, 13022-13027, doi:10.1073/pnas.0704059104 (2007).

197 Liu, Z. R. & Huang, Y. Q. Advantages of proteins being disordered. *Protein Science* **23**, 539-550, doi:10.1002/pro.2443 (2014).

198 Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z. & Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**, 573-584 (2002).

199 Gao, J. J. & Xu, D. Correlation between Posttranslational Modification and Intrinsic Disorder in Protein. *Biocomput-Pac Sym*, 94-103 (2012).

200 Xie, H. B. *et al.* Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res* **6**, 1917-1932, doi:10.1021/pr060394e (2007).

201 Darling, A. L. & Uversky, V. N. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet* **9**, 158, doi:10.3389/fgene.2018.00158 (2018).

202 Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnology* **21**, 255-261, doi:10.1038/nbt0303-255 (2003).

203 Pejaver, V. *et al.* The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Science* **23**, 1077-1093, doi:10.1002/pro.2494 (2014).

204 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).

205 Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* **1**, doi:10.1038/srep00090 (2011).

206 Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med* **40**, 271-280, doi:10.3892/ijmm.2017.3036 (2017).

207 Cohen, P. The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-130, doi:10.1038/ncb0502-e127 (2002).

208 Rubin, C. S. & Rosen, O. M. Protein phosphorylation. *Annu Rev Biochem* **44**, 831-887, doi:10.1146/annurev.bi.44.070175.004151 (1975).

209 Bah, A. & Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J Biol Chem* **291**, 6696-6705, doi:10.1074/jbc.R115.695056 (2016).

210 Flock, T., Weatheritt, R. J., Latysheva, N. S. & Babu, M. M. Controlling entropy to tune the functions of intrinsically disordered regions. *Curr Opin Struct Biol* **26**, 62-72, doi:10.1016/j.sbi.2014.05.007 (2014).

211 Iakoucheva, L. M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037-1049, doi:10.1093/nar/gkh253 (2004).

212 Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* **18**, 343-384, doi:10.1002/jmr.747 (2005).

213 Cohen, P. The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem Sci* **25**, 596-601 (2000).

214 Van Roey, K., Dinkel, H., Weatheritt, R. J., Gibson, T. J. & Davey, N. E. The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci Signal* **6**, rs7, doi:10.1126/scisignal.2003345 (2013).

215 Naik, M. T. *et al.* Staphylococcus aureus Sortase A transpeptidase - Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *Journal of Biological Chemistry* **281**, 1817-1826, doi:DOI 10.1074/jbc.M506123200 (2006).

216 Bienkiewicz, E. A., Adkins, J. N. & Lumb, K. J. Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). *Biochemistry* **41**, 752-759 (2002).

217 Tait, S. *et al.* Local control of a disorder-order transition in 4E-BP1 underpins regulation of translation via eIF4E. *Proc Natl Acad Sci U S A* **107**, 17627-17632, doi:10.1073/pnas.1008242107 (2010).

218 Thomas, D. D., Ramachandran, S., Roopnarine, O., Hayden, D. W. & Ostap, E. M. The mechanism of force generation in myosin: a disorder-to-order transition, coupled to internal structural changes. *Biophys J* **68**, 135S-141S (1995).

219 Borg, M. *et al.* Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Natl Acad Sci U S A* **104**, 9650-9655, doi:10.1073/pnas.0702580104 (2007).

220 Fuxreiter, M., Simon, I. & Bondos, S. Dynamic protein-DNA recognition: beyond what can be seen. *Trends in Biochemical Sciences* **36**, 415-423, doi:10.1016/j.tibs.2011.04.006 (2011).

221 Fuxreiter, M. & Tompa, P. Fuzzy complexes: a more stochastic view of protein function. *Adv Exp Med Biol* **725**, 1-14, doi:10.1007/978-1-4614-0659-4_1 (2012).

222 Fuxreiter, M. & Tompa, P. Fuzzy Complexes: Polymorphism And Structural Disorder In Protein-protein Interactions. *Biophys J* **96**, 319a-319a (2009).

223 Sharma, R., Raduly, Z., Miskei, M. & Fuxreiter, M. Fuzzy complexes: Specific binding without complete folding. *Febs Lett* **589**, 2533-2542, doi:10.1016/j.febslet.2015.07.022 (2015).

224 Olsson, T. S., Ladbury, J. E., Pitt, W. R. & Williams, M. A. Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci* **20**, 1607-1618, doi:10.1002/pro.692 (2011).

225 Machius, M., Declerck, N., Huber, R. & Wiegand, G. Activation of Bacillus licheniformis alpha-amylase through a disorder-->order transition of the substrate-binding site mediated by a calcium-sodium-calcium metal triad. *Structure* **6**, 281-292 (1998).

226 Oldfield, C. J. *et al.* Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454-12470, doi:10.1021/bi050736e (2005).

227 Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struc Biol* **12**, 54-60, doi:Unsp 0959-440x/02/$
Doi 10.1016/S0959-440x(02)00289-0 (2002).

228 Devarakonda, S. *et al.* Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1 alpha/ERR gamma complex. *P Natl Acad Sci USA* **108**, 18678-18683, doi:10.1073/pnas.1113813108 (2011).

229 Love, J. J., Li, X., Chung, J., Dyson, H. J. & Wright, P. E. The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry* **43**, 8725-8734, doi:10.1021/bi049591m (2004).

230 Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology* **5**, 808-814, doi:10.1038/nchembio.238 (2009).

231 Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296, doi:10.1002/prot.340110407 (1991).

232 Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates

binding diversity. *Proc Natl Acad Sci U S A* **93**, 11504-11509, doi:10.1073/pnas.93.21.11504 (1996).

233 Santner, A. A. *et al.* Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* **51**, 7250-7262, doi:10.1021/bi300653m (2012).

234 Frenkel, D. Entropy-driven phase transitions. *Physica A* **263**, 26-38, doi:Doi 10.1016/S0378-4371(98)00501-9 (1999).

235 Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr Opin Struct Biol* **44**, 18-30, doi:10.1016/j.sbi.2016.10.015 (2017).

236 Frenkel, D. Order through Disorder - Entropy Strikes Back. *Phys World* **6**, 24-25, doi:Doi 10.1088/2058-7058/6/2/24 (1993).

237 Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420-435, doi:10.1016/j.tcb.2018.02.004 (2018).

238 Vucetic, S. *et al.* DisProt: a database of protein disorder. *Bioinformatics* **21**, 137-140, doi:10.1093/bioinformatics/bth476 (2005).

239 Lopez-Montero, I. *et al.* Intrinsic disorder of the bacterial cell division protein ZipA: coil-to-brush conformational transition. *FASEB J* **27**, 3363-3375, doi:10.1096/fj.12-224337 (2013).

240 DeForte, S. & Uversky, V. N. Order, Disorder, and Everything in Between. *Molecules* **21**, doi:ARTN 1090
10.3390/molecules21081090 (2016).

241 Pancsa, R. & Tompa, P. Structural disorder in eukaryotes. *Plos One* **7**, e34687, doi:10.1371/journal.pone.0034687 (2012).

242 Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10**, 411-421 (1999).

243 Braun, P. *et al.* Proteome-scale purification of human proteins from bacteria. *Proc Natl Acad Sci U S A* **99**, 2654-2659, doi:10.1073/pnas.042684199 (2002).

244 LaVallie, E. R. & McCoy, J. M. Gene fusion expression systems in Escherichia coli. *Curr Opin Biotechnol* **6**, 501-506 (1995).

245 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172, doi:10.3389/fmicb.2014.00172 (2014).

246 Diguan, C., Li, P., Riggs, P. D. & Inouye, H. Vectors That Facilitate the Expression and Purification of Foreign Peptides in Escherichia-Coli by Fusion to Maltose-Binding Protein. *Gene* **67**, 21-30, doi:Doi 10.1016/0378-1119(88)90004-2 (1988).

247 Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* **17**, 353-358, doi:10.1016/j.copbio.2006.06.003 (2006).

248 Ellis, R. J. & Minton, A. P. Cell biology: join the crowd. *Nature* **425**, 27-28, doi:10.1038/425027a (2003).

249 Seyrek, E., Dubin, P. L., Tribet, C. & Gamble, E. A. Ionic strength dependence of protein-polyelectrolyte interactions. *Biomacromolecules* **4**, 273-282, doi:10.1021/bm025664a (2003).

250 Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11**, 114-119 (2001).

251 Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26**, 597-604 (2001).

252 Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* **20**, 1325-1336 (2003).

253 Morris, A. M., Watzky, M. A. & Finke, R. G. Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. *Biochim Biophys Acta* **1794**, 375-397, doi:10.1016/j.bbapap.2008.10.016 (2009).

254 Fink, A. L. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3**, R9-23, doi:10.1016/S1359-0278(98)00002-9 (1998).

255 Schmittschmitt, J. P. & Scholtz, J. M. The role of protein stability, solubility, and net charge in amyloid fibril formation. *Protein Sci* **12**, 2374-2378, doi:10.1110/ps.03152903 (2003).

256 Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900-904, doi:10.1038/nature02264 (2003).

257 Sasahara, K., McPhie, P. & Minton, A. P. Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J Mol Biol* **326**, 1227-1237 (2003).

258 Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* **35**, 133-152 (1999).

259 Despa, F., Orgill, D. P. & Lee, R. C. Molecular crowding effects on protein stability. *Ann N Y Acad Sci* **1066**, 54-66, doi:10.1196/annals.1363.005 (2005).

260 Cheung, M. S., Klimov, D. & Thirumalai, D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci U S A* **102**, 4753-4758, doi:10.1073/pnas.0409630102 (2005).

261 Chan, H. S. & Dill, K. A. Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* **20**, 447-490, doi:10.1146/annurev.bb.20.060191.002311 (1991).

262 Miklos, A. C., Sarkar, M., Wang, Y. & Pielak, G. J. Protein crowding tunes protein stability. *J Am Chem Soc* **133**, 7116-7120, doi:10.1021/ja200067p (2011).

263 Schellman, J. A. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophys J* **85**, 108-125, doi:10.1016/S0006-3495(03)74459-2 (2003).

264 Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **24**, 1501-1509 (1985).

265 Minde, D. P., Halff, E. F. & Tans, S. Designing disorder: Tales of the unexpected tails. *Intrinsically Disord Proteins* **1**, e26790, doi:10.4161/idp.26790 (2013).

266 Grana-Montes, R., Marinelli, P., Reverter, D. & Ventura, S. N-Terminal Protein Tails Act as Aggregation Protective Entropic Bristles: The SUMO Case. *Biomacromolecules* **15**, 1194-1203, doi:10.1021/bm401776z (2014).

267 Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* **5**, 15-22, doi:10.1038/nchembio.131 (2009).

268 Zhou, H. X. & Dill, K. A. Stabilization of proteins in confined spaces. *Biochemistry* **40**, 11289-11293 (2001).

269 Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* **105**, 8256-8261, doi:10.1073/pnas.0801340105 (2008).

270 Shental-Bechor, D. & Levy, Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol* **19**, 524-533, doi:10.1016/j.sbi.2009.07.002 (2009).

271 Rodriguez-Martinez, J. A. *et al.* Stabilization of alpha-chymotrypsin upon PEGylation correlates with reduced structural dynamics. *Biotechnol Bioeng* **101**, 1142-1149, doi:10.1002/bit.22014 (2008).

272 Muller-Spath, S. *et al.* Charge interactions can dominate the dimensions of intrinsically disordered proteins. *P Natl Acad Sci USA* **107**, 14609-14614, doi:10.1073/pnas.1001743107 (2010).

273 Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **107**, 8183-8188, doi:10.1073/pnas.0911107107 (2010).

274 Hofmann, H. *et al.* Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *P Natl Acad Sci USA* **109**, 16155-16160, doi:10.1073/pnas.1207719109 (2012).

275 Vitalis, A., Wang, X. L. & Pappu, R. V. Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophys J* **93**, 1923-1937, doi:10.1529/biophysj.107.110080 (2007).

276 Hoh, J. H. Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins* **32**, 223-228 (1998).

277 Gsponer, J. & Babu, M. M. Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Rep* **2**, 1425-1437, doi:10.1016/j.celrep.2012.09.036 (2012).

278 Homouz, D., Perham, M., Samiotakis, A., Cheung, M. S. & Wittung-Stafshede, P. Crowded, cell-like environment induces shape changes in aspherical protein. *P Natl Acad Sci USA* **105**, 11754-11759, doi:10.1073/pnas.0803672105 (2008).

279 Pace, C. N. *et al.* Contribution of Hydrophobic Interactions to Protein Stability. *Journal of Molecular Biology* **408**, 514-528, doi:10.1016/j.jmb.2011.02.053 (2011).

280 Tanford, C. Contribution of Hydrophobic Interactions to Stability of Globular Conformation of Proteins. *Journal of the American Chemical Society* **84**, 4240-&, doi:DOI 10.1021/ja00881a009 (1962).

281 Karplus, P. A. Hydrophobicity regained. *Protein Science* **6**, 1302-1307 (1997).

282 Eisenberg, D. & Mclachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **319**, 199-203, doi:DOI 10.1038/319199a0 (1986).

283 Nick Pace, C., Scholtz, J. M. & Grimsley, G. R. Forces stabilizing proteins. *Febs Lett* **588**, 2177-2184, doi:10.1016/j.febslet.2014.05.006 (2014).

284 Wriggers, W., Chakravarty, S. & Jennings, P. A. Control of protein functional dynamics by peptide linkers. *Biopolymers* **80**, 736-746, doi:10.1002/bip.20291 (2005).

285 Gerstein, M., Lesk, A. M. & Chothia, C. Structural mechanisms for domain movements in proteins. *Biochemistry* **33**, 6739-6749 (1994).

286 Boehr, D. D., D'Amico, R. N. & O'Rourke, K. F. Engineered control of enzyme structural dynamics and function. *Protein Science* **27**, 825-838, doi:10.1002/pro.3379 (2018).

287 Nestl, B. M. & Hauer, B. Engineering of Flexible Loops in Enzymes. *Acs Catal* **4**, 3201-3211, doi:10.1021/cs500325p (2014).

288 Ladurner, A. G. & Fersht, A. R. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *Journal of Molecular Biology* **273**, 330-337, doi:DOI 10.1006/jmbi.1997.1304 (1997).

289 Nagi, A. D. & Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* **2**, 67-75, doi:Doi 10.1016/S1359-0278(97)00007-2 (1997).

290 Wang, L. P., Rivera, E. V., Benavides-Garcia, M. G. & Nall, B. T. Loop entropy and cytochrome c stability. *Journal of Molecular Biology* **353**, 719-729, doi:10.1016/j.jmb.2005.08.038 (2005).

291 Robinson, C. R. & Sauer, R. T. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *P Natl Acad Sci USA* **95**, 5929-5934, doi:DOI 10.1073/pnas.95.11.5929 (1998).

292 Gavrilov, Y., Dagan, S. & Levy, Y. Shortening a loop can increase protein native state entropy. *Proteins-Structure Function and Bioinformatics* **83**, 2137-2146, doi:10.1002/prot.24926 (2015).

293 Nagi, A. D., Anderson, K. S. & Regan, L. Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *Journal of Molecular Biology* **286**, 257-265, doi:DOI 10.1006/jmbi.1998.2474 (1999).

294 Jackson, S. E. & Fersht, A. R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-10435 (1991).

295 Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* **254**, 260-288, doi:10.1006/jmbi.1995.0616 (1995).

296 Neira, J. L. *et al.* Towards the complete structural characterization of a protein folding pathway: the structures of the denatured, transition and native states for the association/folding of two complementary fragments of cleaved chymotrypsin inhibitor 2. Direct evidence for a nucleation-condensation mechanism. *Fold Des* **1**, 189-208 (1996).

297 Carmichael, S. P. & Shell, M. S. Entropic (de)stabilization of surface-bound peptides conjugated with polymers. *J Chem Phys* **143**, doi:10.1063/1.4929592 (2015).

298 Smyda, M. R. & Harvey, S. C. The entropic cost of polymer confinement. *J Phys Chem B* **116**, 10928-10934, doi:10.1021/jp302807r (2012).

299 Sakaue, T. & Raphael, E. Polymer chains in confined spaces and flow-injection problems: Some remarks. *Macromolecules* **39**, 2621-2628, doi:10.1021/ma0514424 (2006).

300 Bickel, T., Jeppesen, C. & Marques, C. M. Local entropic effects of polymers grafted to soft interfaces. *Eur Phys J E* **4**, 33-43, doi:10.1007/s101890170140 (2001).

301 Bickel, T., Marques, C. & Jeppesen, C. Pressure patches for membranes: the induced pinch of a grafted polymer. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **62**, 1124-1127 (2000).

302 Waters, J. T. & Kim, H. D. Calculation of a fluctuating entropic force by phase space sampling. *Phys Rev E* **92**, doi:10.1103/PhysRevE.92.013308 (2015).

303 Busch, D. J. *et al.* Intrinsically disordered proteins drive membrane curvature. *Nat Commun* **6**, 7875, doi:10.1038/ncomms8875 (2015).

304 Liu, T. *et al.* Direct measurement of the mechanical work during translocation by the ribosome. *Elife* **3**, e03406, doi:10.7554/eLife.03406 (2014).

305 Perales-Calvo, J., Giganti, D., Stirnemann, G. & Garcia-Manyes, S. The force-dependent mechanism of DnaK-mediated mechanical folding. *Sci Adv* **4**, eaaq0243, doi:10.1126/sciadv.aaq0243 (2018).

306 Leininger, S., Narayan, K. B., Deutsch, C. & O'Brien, E. P. Mechanochemistry in Translation. *Biochemistry*, doi:10.1021/acs.biochem.9b00260 (2019).

307 Fritch, B. *et al.* Origins of the Mechanochemical Coupling of Peptide Bond Formation to Protein Synthesis. *J Am Chem Soc* **140**, 5077-5087, doi:10.1021/jacs.7b11044 (2018).

308 Wilson, D. N. & Beckmann, R. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Curr Opin Struc Biol* **21**, 274-282, doi:10.1016/j.sbi.2011.01.007 (2011).

309 Goldman, D. H. *et al.* Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* **348**, 457-460, doi:10.1126/science.1261909 (2015).

310 Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469, doi:10.1016/j.cell.2011.04.010 (2011).

311 Augustyniak, R. & Kay, L. E. Cotranslocational processing of the protein substrate calmodulin by an AAA+ unfoldase occurs via unfolding and refolding intermediates. *Proc Natl Acad Sci U S A* **115**, E4786-E4795, doi:10.1073/pnas.1721811115 (2018).

312 Shariff, K., Ghosal, S. & Matouschek, A. The force exerted by the membrane potential during protein import into the mitochondrial matrix. *Biophys J* **86**, 3647-3652, doi:10.1529/biophysj.104.040865 (2004).

313 De Los Rios, P., Ben-Zvi, A., Slutsky, O., Azem, A. & Goloubinoff, P. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc Natl Acad Sci U S A* **103**, 6166-6171, doi:10.1073/pnas.0510496103 (2006).

314 Goloubinoff, P. & De Los Rios, P. The mechanism of Hsp70 chaperones: (entropic) pulling the models together. *Trends in Biochemical Sciences* **32**, 372-380, doi:10.1016/j.tibs.2007.06.008 (2007).

315 Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat Commun* **5**, 4841, doi:10.1038/ncomms5841 (2014).

316 Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat Struct Mol Biol* **11**, 830-837, doi:10.1038/nsmb814 (2004).

317 Maxwell, E. S., Kalckar, H. M. & Strominger, J. L. Some Properties of Uridine Diphosphoglucose Dehydrogenase. *Archives of Biochemistry and Biophysics* **65**, 2-10, doi:Doi 10.1016/0003-9861(56)90171-0 (1956).

318 Beattie, N. R., Keul, N. D., Sidlo, A. M. & Wood, Z. A. Allostery and Hysteresis Are Coupled in Human UDP-Glucose Dehydrogenase. *Biochemistry* **56**, 202-211, doi:10.1021/acs.biochem.6b01044 (2017).

319 Gainey, P. A. & Phelps, C. F. Interactions of Uridine-Diphosphate Glucose Dehydrogenase with Inhibitor Uridine-Diphosphate Xylose. *Biochem J* **145**, 129-134, doi:DOI 10.1042/bj1450129 (1975).

320 Kadirvelraj, R. *et al.* Hysteresis in Human UDP-Glucose Dehydrogenase Is Due to a Restrained Hexameric Structure That Favors Feedback Inhibition. *Biochemistry* **53**, 8043-8051, doi:10.1021/bi500594x (2014).

321 Kadirvelraj, R., Sennett, N. C., Polizzi, S. J., Weitzel, S. & Wood, Z. A. Role of packing defects in the evolution of allostery and induced fit in human UDP-glucose dehydrogenase. *Biochemistry* **50**, 5780-5789, doi:10.1021/bi2005637 (2011).

322 Sennett, N. C., Kadirvelraj, R. & Wood, Z. A. Conformational flexibility in the allosteric regulation of human UDP-alpha-D-glucose 6-dehydrogenase. *Biochemistry* **50**, 9651-9663, doi:10.1021/bi201381e (2011).

323 Beattie, N. R., Pioso, B. J., Sidlo, A. M., Keul, N. D. & Wood, Z. A. Hysteresis and Allostery in Human UDP-Glucose Dehydrogenase Require a Flexible Protein Core. *Biochemistry*, doi:10.1021/acs.biochem.8b00497 (2018).

324 Kadirvelraj, R., Sennett, N. C., Custer, G. S., Phillips, R. S. & Wood, Z. A. Hysteresis and negative cooperativity in human UDP-glucose dehydrogenase. *Biochemistry* **52**, 1456-1465, doi:10.1021/bi301593c (2013).

325 Sennett, N. C., Kadirvelraj, R. & Wood, Z. A. Cofactor binding triggers a molecular switch to allosterically activate human UDP-alpha-D-glucose 6-dehydrogenase. *Biochemistry* **51**, 9364-9374, doi:10.1021/bi301067w (2012).

326 Li, J. *et al.* Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. *Elife* **6**, doi:10.7554/eLife.30688 (2017).

327 Ferreon, A. C., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390-394, doi:10.1038/nature12294 (2013).

328 Neufeld, E. F. & Hall, C. W. Inhibition of Udp-D-Glucose Dehydrogenase by Udp-D-Xylose - a Possible Regulatory Mechanism. *Biochem Bioph Res Co* **19**, 456-&, doi:Doi 10.1016/0006-291x(65)90146-4 (1965).

329 Theillet, F. X. *et al.* The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24360, doi:10.4161/idp.24360 (2013).

330 Uversky, V. N. The intrinsic disorder alphabet. III. Dual personality of serine. *Intrinsically Disord Proteins* **3**, e1027032, doi:10.1080/21690707.2015.1027032 (2015).

331 Balasubramaniam, D. & Komives, E. A. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim Biophys Acta* **1834**, 1202-1209, doi:10.1016/j.bbapap.2012.10.009 (2013).

332 Chen, J. Z. Y. Theory of wormlike polymer chains in confinement. *Prog Polym Sci* **54-55**, 3-46, doi:10.1016/j.progpolymsci.2015.09.002 (2016).

333 Thielges, M. C., Chung, J. K., Axup, J. Y. & Fayer, M. D. Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry* **50**, 5799-5805, doi:10.1021/bi2003923 (2011).

334 Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552 (1999).

335 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

336 Brunger, A. T. Free R value: cross-validation in crystallography. *Methods Enzymol* **277**, 366-396 (1997).

337 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

338 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

339 Urzhumtsev, A., Afonine, P. V. & Adams, P. D. TLS from fundamentals to practice. *Crystallogr Rev* **19**, 230-270, doi:10.1080/0889311X.2013.835806 (2013).

340 Diederichs, K. & Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**, 269-275 (1997).

341 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).

342 Frieden, C. Kinetic aspects of regulation of metabolic processes. The hysteretic enzyme concept. *J Biol Chem* **245**, 5788-5799 (1970).

343 Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. *Analytical Ultracentrifugation*. (The Royal Society of Chemistry, 1992).

344 Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-1619, doi:10.1016/S0006-3495(00)76713-0 (2000).

345 Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem* **320**, 104-124 (2003).

346 Ortega, A., Amoros, D. & Garcia de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**, 892-898, doi:10.1016/j.bpj.2011.06.046 (2011).

347 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

348 Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275-282 (1992).

349 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).

350 Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95-99 (1963).

351 Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of Molecular Biology* **285**, 1735-1747, doi:DOI 10.1006/jmbi.1998.2401 (1999).

352 Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* **79**, 28 29 21-14, doi:10.1002/0471140864.ps2809s79 (2015).

353 Fang, J. *et al.* Conformational dynamics of the Escherichia coli DNA polymerase manager proteins UmuD and UmuD'. *J Mol Biol* **398**, 40-53, doi:10.1016/j.jmb.2010.02.040 (2010).

354 Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* **25**, 158-170, doi:10.1002/mas.20064 (2006).

355 Johnson, B. *et al.* Dimerization Controls Marburg Virus VP24-dependent Modulation of Host Antioxidative Stress Responses. *J Mol Biol* **428**, 3483-3494, doi:10.1016/j.jmb.2016.07.020 (2016).

356 Chen, E. *et al.* Broadly neutralizing epitopes in the Plasmodium vivax vaccine candidate Duffy Binding Protein. *Proc Natl Acad Sci U S A* **113**, 6277-6282, doi:10.1073/pnas.1600488113 (2016).

357 Yan, Y., Grant, G. A. & Gross, M. L. Hydrogen-Deuterium Exchange Mass Spectrometry Reveals Unique Conformational and Chemical Transformations Occurring upon [4Fe-4S] Cluster Binding in the Type 2 L-Serine Dehydratase from Legionella pneumophila. *Biochemistry* **54**, 5322-5328, doi:10.1021/acs.biochem.5b00761 (2015).

358 Palmier, M. O. & Van Doren, S. R. Rapid determination of enzyme kinetics from fluorescence: overcoming the inner filter effect. *Anal Biochem* **371**, 43-51, doi:10.1016/j.ab.2007.07.008 (2007).

359 Johnson, K. A. Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol* **467**, 601-626, doi:10.1016/S0076-6879(09)67023-3 (2009).

360 Johnson, K. A., Simpson, Z. B. & Blom, T. Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. *Analytical Biochemistry* **387**, 20-29, doi:10.1016/j.ab.2008.12.024 (2009).

361 Johnson, K. A., Simpson, Z. B. & Blom, T. FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal Biochem* **387**, 30-41, doi:10.1016/j.ab.2008.12.025 (2009).

362 Guillemette, C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J* **3**, 136-158, doi:10.1038/sj.tpj.6500171 (2003).

363 Lin, J. H. & Lu, A. Y. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol Rev* **49**, 403-449 (1997).

364 Tukey, R. H. & Strassburg, C. P. Human UDP-glucuronosyltransferases: Metabolism, expression, and disease. *Annu Rev Pharmacol* **40**, 581-616, doi:DOI 10.1146/annurev.pharmtox.40.1.581 (2000).

365 Cummings, J. *et al.* Enhanced clearance of topoisomerase I inhibitors from human colon cancer cells by glucuronidation. *Biochem Pharmacol* **63**, 607-613 (2002).

366 Cummings, J. *et al.* Glucuronidation as a mechanism of intrinsic drug resistance in human colon cancer: Reversal of resistance by food additives. *Cancer Res* **63**, 8443-8450 (2003).

367 de Almagro, M. C. *et al.* UDP-glucuronosyltransferase 1A6 overexpression in breast cancer cells resistant to methotrexate. *Biochemical Pharmacology* **81**, 60-70, doi:10.1016/j.bcp.2010.09.008 (2011).

368 Oguri, T. *et al.* UGT1A10 is responsible for SN-38 glucuronidation and its expression in human lung cancers. *Anticancer Res* **24**, 2893-2896 (2004).

369 Kakizaki, I. *et al.* A novel mechanism for the inhibition of hyaluronan biosynthesis by 4-methylumbelliferone. *J Biol Chem* **279**, 33281-33289, doi:10.1074/jbc.M405918200 (2004).

370 Kultti, A. *et al.* 4-Methylumbelliferone inhibits hyaluronan synthesis by depletion of cellular UDP-glucuronic acid and downregulation of hyaluronan synthase 2 and 3. *Exp Cell Res* **315**, 1914-1923, doi:10.1016/j.yexcr.2009.03.002 (2009).

371 Wei, Q., Galbenus, R., Raza, A., Cerny, R. L. & Simpson, M. A. Androgen-Stimulated UDP-Glucose Dehydrogenase Expression Limits Prostate Androgen Availability without Impacting Hyaluronan Levels. *Cancer Res* **69**, 2332-2339, doi:10.1158/0008-5472.Can-08-3083 (2009).

372 Keul, N. D. *et al.* The entropic force generated by intrinsically disordered segments tunes protein function. *Nature*, doi:10.1038/s41586-018-0699-5 (2018).

373 Beattie, N. R., Pioso, B., Sidlo, A. M., Keul, N. D. & Wood, Z. A. Hysteresis and Allostery in Human UDP-Glucose Dehydrogenase Require a Flexible Protein Core. *Biochemistry*, doi:10.1021/acs.biochem.8b00497 (2018).

374 Hayward, S. & Berendsen, H. J. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144-154 (1998).

375 Girdlestone, C. & Hayward, S. The DynDom3D Webserver for the Analysis of Domain Movements in Multimeric Proteins. *J Comput Biol* **23**, 21-26, doi:10.1089/cmb.2015.0143 (2016).

376 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

377 Laskowski, R. A. & Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-2786, doi:10.1021/ci200227u (2011).

378 Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).

379 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

380 Neuwald, A. F. Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. *Bioinformatics* **25**, 1869-1875, doi:10.1093/bioinformatics/btp342 (2009).

381 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

382 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One* **5**, doi:ARTN e9490 10.1371/journal.pone.0009490 (2010).

383 Zhao, H., Balbo, A., Brown, P. H. & Schuck, P. The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity. *Methods* **54**, 16-30, doi:10.1016/j.ymeth.2011.01.010 (2011).

384 Bruns, C. M., Hubatsch, I., Ridderstrom, M., Mannervik, B. & Tainer, J. A. Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J Mol Biol* **288**, 427-439, doi:10.1006/jmbi.1999.2697 (1999).

385 Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K. & Uversky, V. N. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* **45**, 10448-10460, doi:10.1021/bi060981d (2006).

386 Uversky, V. N. Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front Biosci (Landmark Ed)* **14**, 5188-5238 (2009).

387 Uversky, V. N. *et al.* Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* **10 Suppl 1**, S7, doi:10.1186/1471-2164-10-S1-S7 (2009).

388 Thompson, M. J. & Eisenberg, D. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol* **290**, 595-604, doi:10.1006/jmbi.1999.2889 (1999).

389 Vihinen, M. Relationship of protein flexibility to thermostability. *Protein Eng* **1**, 477-480 (1987).

390 Fields, P. A. Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp Biochem Physiol A Mol Integr Physiol* **129**, 417-431 (2001).

391 Drummond, D. A. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* **10**, 715-724, doi:10.1038/nrg2662 (2009).

392 Bucciantini, M. *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507-511, doi:DOI 10.1038/416507a (2002).

393 Dobson, C. M. Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**, 3-16, doi:10.1016/j.semcdb.2003.12.008 (2004).

394 Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med (Berl)* **81**, 678-699, doi:10.1007/s00109-003-0464-5 (2003).

395 Goldstein, R. A. Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: insights from the quasi-chemical approximation. *Protein Sci* **16**, 1887-1895, doi:10.1110/ps.072947007 (2007).

396 Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **355**, 371-373, doi:10.1038/355371a0 (1992).

397 Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. The crystal structure of citrate synthase from the hyperthermophilic archaeon pyrococcus furiosus at 1.9 A resolution. *Biochemistry* **36**, 9983-9994, doi:10.1021/bi9705321 (1997).

398 Berezovsky, I. N., Zeldovich, K. B. & Shakhnovich, E. I. Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput Biol* **3**, e52, doi:10.1371/journal.pcbi.0030052 (2007).

399 Jaenicke, R. & Zavodszky, P. Proteins under extreme physical conditions. *Febs Lett* **268**, 344-349 (1990).

400 Dalhus, B. *et al.* Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. *J Mol Biol* **318**, 707-721, doi:10.1016/S0022-2836(02)00050-5 (2002).

401 Meruelo, A. D., Han, S. K., Kim, S. & Bowie, J. U. Structural differences between thermophilic and mesophilic membrane proteins. *Protein Sci* **21**, 1746-1753, doi:10.1002/pro.2157 (2012).

402 Szilagyi, A. & Zavodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493-504, doi:Doi 10.1016/S0969-2126(00)00133-7 (2000).

403 Vogt, G., Woell, S. & Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* **269**, 631-643, doi:10.1006/jmbi.1997.1042 (1997).

404 Vogt, G. & Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Fold Des* **2**, S40-46 (1997).

405 Karshikoff, A. & Ladenstein, R. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Engineering* **11**, 867-872, doi:DOI 10.1093/protein/11.10.867 (1998).

406 Siddiqui, K. S. & Cavicchioli, R. Cold-adapted enzymes. *Annu Rev Biochem* **75**, 403-433, doi:10.1146/annurev.biochem.75.103004.142723 (2006).

407 Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N. & Arnold, F. H. Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* **297**, 1015-1026, doi:10.1006/jmbi.2000.3612 (2000).

408 Isaksen, G. V., Aqvist, J. & Brandsdal, B. O. Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *P Natl Acad Sci USA* **113**, 7822-7827, doi:10.1073/pnas.1605237113 (2016).

409 Feller, G. *et al.* Enzymes from psychrophilic organisms. *Fems Microbiol Rev* **18**, 189-202, doi:DOI 10.1111/j.1574-6976.1996.tb00236.x (1996).

410 Arnold, F. H., Wintrode, P. L., Miyazaki, K. & Gershenson, A. How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci* **26**, 100-106 (2001).

411 Feller, G., Arpigny, J. L., Narinx, E. & Gerday, C. Molecular adaptations of enzymes from psychrophilic organisms. *Comp Biochem Phys A* **118**, 495-499, doi:Doi 10.1016/S0300-9629(97)00011-X (1997).

412 Struvay, C. & Feller, G. Optimization to Low Temperature Activity in Psychrophilic Enzymes. *Int J Mol Sci* **13**, 11643-11665, doi:10.3390/ijms130911643 (2012).

413 Georlette, D. *et al.* Some like it cold: biocatalysis at low temperatures. *Fems Microbiol Rev* **28**, 25-42, doi:10.1016/j.femsre.2003.07.003 (2004).

414 Russell, N. J. Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extremophiles* **4**, 83-90 (2000).

415 Piana, S., Klepeis, J. L. & Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr Opin Struc Biol* **24**, 98-105, doi:10.1016/j.sbi.2013.12.006 (2014).

416 Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* **19**, 120-127, doi:10.1016/j.sbi.2009.03.004 (2009).

417 Chin, A. F., Toptygin, D., Elam, W. A., Schrank, T. P. & Hilser, V. J. Phosphorylation Increases Persistence Length and End-to-End Distance of a Segment of Tau Protein. *Biophys J* **110**, 362-371, doi:10.1016/j.bpj.2015.12.013 (2016).

418 Grimmler, M. *et al.* Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell* **128**, 269-280, doi:10.1016/j.cell.2006.11.047 (2007).

419 Kurganov, B. I. Analysis of negative cooperativity for glutamate dehydrogenase. *Biophysical Chemistry* **87**, 185-199 (2000).

APPENDIX A

HYSTERESIS IN HUMAN UDP-GLUCOSE DEHYDROGENASE IS DUE TO A RESTRAINED HEXAMERIC STRUCTURE THAT FAVORS FEEDBACK INHIBITION

Keul, N.D. conducted and analyzed the enzyme hysteresis assays at different pH's. In addition, Keul, N.D. contributed in editorial revisions of the manuscript.

A1. Abstract

Human UDP-α-D-glucose-6-dehydrogenase (hUGDH) displays hysteresis because of a slow isomerization from an inactive state (E*) to an active state (E). Here we show that the structure of E* constrains hUGDH in a conformation that favors feedback inhibition at physiological pH. The feedback inhibitor UDP-α-D-xylose (UDP-Xyl) competes with the substrate UDP-α-D-glucose for the active site. Upon binding, UDP-Xyl triggers an allosteric switch that changes the structure and affinity of the intersubunit interface to form a stable but inactive horseshoe-shaped hexamer. Using sedimentation velocity studies and a new crystal structure, we show that E* represents a stable conformational intermediate between the active and feedback-inhibited conformations. Because the allosteric switch occludes the cofactor and substrate binding sites in the inactive hexamer, the intermediate conformation observed in the crystal structure is consistent with the E* transient observed in relaxation studies. Steady-state analysis shows that the E* conformation enhances the affinity of hUGDH for the allosteric inhibitor UDP-Xyl by 8.6-fold (K_i = 810 nM). We present a model in which the constrained quaternary structure permits a small effector molecule to leverage a disproportionately large allosteric response.

A2. Introduction

Human UDP-α-D-glucose-6-dehydrogenase (hUGDH) catalyzes the NAD+-dependent oxidation of UDP-α-D-glucose (UDP-Glc) to produce UDP-α-D-glucuronic acid.[1-4] In solution, apo-hUGDH forms a hexamer built from three weakly interacting dimers.[5-7] Because of the weak affinity between subunits, the apo hexamer easily dissociates at low protein concentrations. We have previously shown that dimer association is under allosteric control.[5, 6, 8] The feedback inhibitor UDP-α-D-xylose (UDP-Xyl) binds to the active site and triggers an allosteric switch that increases the affinity between the subunit interfaces to form a stable, horseshoe-shaped hexamer (Figure A1.A,B).[8] In the horseshoe conformation, the allosteric switch sterically occludes the binding sites for the cofactor and the C5′-hydroxymethyl on the substrate to inactivate the enzyme (Figure A1.B). Because the allosteric switch responds to the absence of a C5′-hydroxymethyl in UDP-Xyl, the substrate UDP-Glc has no effect on the affinity between subunits or hexamer formation (Figure A1.B).[5, 6, 8] In contrast, binding of the cofactor (NAD^+) allosterically increases the affinity between subunits and induces the formation of an active, 32 symmetry hexamer (Figure A1.A).[6] hUGDH is unusual in that the allosteric and active sites are present as a single site with distinct induced-fit responses that control subunit affinity and quaternary structure.[8]

In addition to allostery, hUGDH also displays hysteresis, which can be observed as a lag in progress curves (Figure A1.C).[7,9] Hysteresis can occur when an enzyme undergoes a slow transition from an inactive conformation (E*) to a more active state (E) in the presence of substrate or an allosteric effector (Figure A1.D).[10] We have recently shown that the lag in hUGDH is not due to the association of dimers or the dissociation of the hexamer or another aggregate.[7] Instead, hUGDH hysteresis is due to a cofactor and substrate-induced isomerization of the E* conformation to form the active enzyme (Figure A1.D).[7] Previous work[7] has shown that the hysteretic transition

in hUGDH is most consistent with the ligand-induced slow transition (LIST) model in which the binding of the cofactor to the binary E*:UDP-Glc complex induces a slow isomerization to the active (or more active) E:UDP-Glc:NAD+ complex (Figure A1.D).[10] Here we show that the E* structure represents a conformational intermediate between the active 32 symmetry and inactive horseshoe hexamers. Because the only difference between UDP-Xyl and UDP-Glc is the absence of a C5′- hydroxymethyl group in the inhibitor, we argue that the function of the intermediate, E*, is to leverage the conformation of hUGDH into a state that can easily switch to the inhibited complex.

A3. Materials and Methods

Protein Expression and Crystallization

Wild-type hUGDH was recombinantly expressed and purified, and the His tag was removed as described previously.[8] The purified protein was dialyzed into storage buffer [25 mM Tris (pH 8) and 200 mM NaCl] and concentrated to 10 mg/mL. Crystals were grown at 20 °C using the hanging drop vapor diffusion method with 2 μL drops mixed in a 1:1 ratio (protein:reservoir). The reservoir contained 50 mM sodium formate, 15% PEG 3350, and 100 mM HEPES (pH 7.0).

Data Collection, Structure Solution, and Refinement

Crystals were cryoprotected with the reservoir solution containing 21% glycerol, ethylene glycol, and dimethyl sulfoxide in a 1:1:1 ratio and plunged into liquid nitrogen prior to data collection. The 2.7 Å resolution data were collected at the 22-BM beamline (SER-CAT) at the Argonne National Laboratory (Argonne, IL) using a MAR 225 mm CCD detector, a wavelength of 1 Å, and an oscillation step of 1.0° for 360°. Data were processed with XDS,[11] and 8% of the data were set aside for cross-validation. The crystal parameters and data collection statistics are summarized in Table A1.

The structure of E* was determined by molecular replacement using the PHENIX software suite.[12] Molecular replacement was conducted using an unpublished crystal structure of hUGDH lacking the extreme C-terminus (residues 467−494) as the search model. The model was subjected to rigid body and positional refinement using the PHENIX software suite.[12] Refinement consisted of iterative cycles of manual rebuilding using COOT followed by automated refinement with PHENIX.[12,13] The B factors were refined using TLS.[14] Because the molecular replacement search model and the final structure were in isomorphous space groups, we inadvertently corrupted the cross-validation test set. To calculate an *a posteriori* R_{free}, we selected a new test set representing 10% of the data.[15] The final model was subjected to simulated annealing using Cartesian molecular dynamics at 2500 K to decouple the test set.[15] The final model statistics are listed in Table A1.

Sedimentation Velocity

Solutions of 0.91 μM hUGDH, 150 mM KCl, and 25 mM buffer [MES (pH 6.5), HEPES (pH 7.5), or Tris (pH 8.5)] were loaded into 12 mm cells and placed in an Optima XLA analytical ultracentrifuge set to 20 °C and 50000 rpm. The hUGDHM11 dimer (9.1 μM) was analyzed similarly in 150 mM KCl and 25 mM HEPES (pH 7.5). Sedimentation within the cells was monitored in 0.003 cm step sizes at 230 nm (0.9 μM) or 280 nm (9.0 μM) for 8 h. Raw sedimentation data were modeled using SEDFIT as *c(s)* distributions that fit baseline and meniscus as well as systematic and time invariant noise.[16] SEDENTERP was used to estimate a partial specific volume of 0.7384 mL g^{-1} and densities of 1.00532, 1.00726, and 1.00603 g mL^{-1} for pH 6.5, 7.5, and 8.5 solutions, respectively.[17] Viscosities for the same pH range were estimated to be 0.01003, 0.010176, and 0.010068 P, respectively.

Steady-State Kinetics and UDP-Xyl Inhibition

Nucleotides NAD^+, NADH, and UDP-Glc were obtained from Sigma. UDP-Xyl was obtained from Carbosource (University of Georgia). All experiments were conducted in a reaction buffer containing 50 mM HEPES or 50 mM Tris, 50 mM NaCl, and 5 mM EDTA. Steady-state kinetic analysis of wild-type hUGDH was conducted using saturating amounts of 2.5 mM NAD^+ with 100 nM hUGDH at pH 7.5 and 230 nM enzyme at pH 8.5. Prior to the reaction, the enzyme and substrates were preincubated separately for 5 min at 25 °C and the reaction was initiated by a rapid manual mixing of the two solutions. An Agilent 8453 UV/vis spectrometer equipped with a Peltier temperature controller was used to follow enzyme reactions by continually monitoring the formation of NADH at 340 nm (molar absorptivity coefficient of 6220 M^{-1} cm^{-1}). All data were analyzed and modeled using nonlinear regression as implemented in PRISM (GraphPad Software Inc., San Diego, CA). Steady-state velocities were calculated by fitting the progress curves to the hysteresis equation derived by Frieden[10]:

$$P(t) = v_{ss}t - \tau(v_{ss} - v_i)(1 - e^{-t/\tau}) \quad \text{(equation 1)}$$

where P is the concentration of product at time t and τ is the relaxation time. τ is equal to $1/k_{obs}$, where k_{obs} is the apparent rate constant for the transition between the initial (v_i) and steady-state velocities (v_{ss}) (Figure A1.C). It is important to note that v_i in eq 1 is the initial velocity prior to steady state, and only v_{ss} is appropriate for calculating initial steady-state velocities for use in substrate saturation curves (with the caveat that < 10% of substrate has been consumed prior to reaching v_{ss}). For steady-state analysis, the substrate saturation curves were fit to a sigmoidal model based on residual analysis[18]:

$$v_o = \frac{k_{cat}[E_t][S]^h}{K_M{}^h + [S]^h} \quad \text{(equation 2)}$$

To determine the K_i for the competitive inhibitor UDP-Xyl, steady-state analysis was conducted using different concentrations of UDP-Xyl (see Figure A3 for concentrations). The data

from each concentration of UDP-Xyl were simultaneously fit to the equation for competitive inhibition using global analysis in PRISM:

$$v_o = \frac{k_{cat}[E_t][S]^h}{(K_M^{app})^h + [S]^h} \text{ where } K_M^{app} = K_M\left(1 + \frac{[I]}{K_i}\right) \quad \text{(equation 3)}$$

Specific activity studies were conducted using saturating amounts of 2.5 mM NAD^+ and 0.3 mM UDP-Glc with 230 nM enzyme at pH 8.4. To determine the pH optimum, reaction mixtures were prepared as separate substrate (2.5 mM NAD^+ and 0.6 mM UDP-Glc) and enzyme (230 nM) solutions in 50 mM buffer (pH 6.0–6.4 for MES, pH 7.0–7.4 for HEPES, pH 7.8–9.0 for Tris, and pH 9.4–10.4 for glycine), 5 mM EDTA, and 50 mM NaCl. Substrate and enzyme solutions were preincubated separately for 5 min at a reaction temperature of 25 °C, and the reaction was initiated by a rapid mixing of the two solutions.

Modeling the Dependence of Specific Activity on Protein Concentration

In our previous work, we showed that the tetramer is a transient in the formation and dissociation of the hexamer.[5,6] In modeling the dependence of hUGDH specific activity on protein concentration, we make the simplifying assumption that the tetramer makes a negligible contribution to specific activity. Thus, our derivation assumes an equilibrium between three low-activity dimers (E_D) and a higher-activity hexamer (E_H):

$$3E_D \overset{K_a}{\leftrightarrow} E_H \quad \text{(equation 4)}$$

The hexamer concentration can be expressed as:

$$[E_H] = K_a[E_D]^3 \quad \text{(equation 5)}$$

The total concentration of enzyme, $[E_T]$, is:

$$[E_T] = 2[E_D] + 6[E_H] \quad \text{(equation 6)}$$

which can be expressed in terms of the dimer concentration:

$$[E_T] = 2[E_D] + 6K_a[E_D]^3 \quad \text{(equation 7)}$$

or

$$0 = 6K_a[E_D]^3 + 2[E_D] - [E_T] \quad \text{(equation 8)}$$

Using the general solution for a cubic polynomial, we solve for $[E_D]$ in terms of total enzyme concentration:

$$[E_D] = \sqrt[3]{\frac{[E_T]}{12K_a} + \sqrt{\left(\frac{[E_T]}{12K_a}\right)^2 + \left(\frac{1}{9K_a}\right)^3}} + \sqrt[3]{\frac{[E_T]}{12K_a} - \sqrt{\left(\frac{[E_T]}{12K_a}\right)^2 + \left(\frac{1}{9K_a}\right)^3}} \quad \text{(equation 9)}$$

The specific activity (SA) of hUGDH is the sum of the specific activities of the hexamer (SA_H) and dimer (SA_D) multiplied by their molar fraction:

$$SA = \frac{6[E_H]}{[E_T]}SA_H + \frac{2[E_D]}{[E_T]}SA_D \quad \text{(equation 10)}$$

Expressing the SA in terms of the dimer concentration (eq 4):

$$SA = \frac{6K_a[E_D]^3}{[E_T]}SA_H + \frac{2[E_D]}{[E_T]}SA_D \quad \text{(equation 11)}$$

We can now substitute eq 9 for $[E_D]$ in the expression given above and solve for total specific activity in terms of total enzyme concentration. Specific activity data at different enzyme concentrations were fit using the program PRISM.

A4. Results

The Hexameric Structure of hUGDH Is Destabilized at Alkaline pH

We have previously shown that the apo-hUGDH hexamer dissociates at low protein concentrations.[5] Here, we used analytical ultracentrifugation to study the effects of pH on the oligomeric state. Sedimentation velocity studies of 0.9 μM hUGDH at pH 7.5 show that the *c(s)* distribution is dominated by an 11.4 S species (77%) corresponding to the hUGDH hexamer. We also observe a 8.6 and 5.4 S species, indicating the presence of a small amount of tetramer (8%)

and dimer (15%), respectively (Figure A2.A). The hexamer is also favored at pH 6.5 (85% of the distribution), with smaller amounts of tetramer (4%) and dimer (11%) (Figure A2.A). In contrast, the hexamer dissociates at alkaline pH. At pH 8.5, 85% of the *c(s)* distribution is represented by a 5.7 S species consistent with the hUGDH dimer, and the remaining 15% sediments at 3.7 S, indicating a small amount of monomer (Figure A2.A). The s-values obtained from rapid equilibrium systems are biased by the mean of the *c(s)* distribution; thus, the apparent deviation in s-values for the dimer at pH 8.5 and 7.5 (5.7 and 5.4 S, respectively) does not necessarily imply a significant change in the conformation of the protein.[19] These results show that physiological pH stabilizes the hUGDH hexamer.

The Hexameric Structure Alters the Activity of hUGDH

We determined the pH optimum of hUGDH to be ~8.6, which is consistent with earlier reports (Figure A2.B).[2, 20] The alkaline pH optimum appears to be at odds with our observation of a dimer at pH 8.5, given that the hexamer is required for maximal catalytic activity (Figure A2.A).[6, 21] Because cofactor and substrate binding will allosterically induce hexamer formation, it may be that the hexamer is formed under steady-state conditions.[6] To test this possibility, we examined the change in specific activity with respect to enzyme concentration. The specific activity of hUGDH displays a hyperbolic dependence on enzyme concentration at pH 8.4 (Figure A2.C). The saturable behavior of hUGDH specific activity is consistent with three low-activity dimers (D) undergoing a concentration-dependent association to form a higher-activity hexamer[6]:

$$3[\mathrm{D}] \overset{K}{\leftrightarrow} [\mathrm{D}]^3$$

This model assumes that the tetramer is a negligible species in the association and dissociation of the hexamer and is based on our previous work showing that the tetramer behaves as a transient in the dimer–hexamer equilibrium.[5, 6] The change in specific activity due to hexamer

formation can be fit to an isotherm relating specific activity to protein self-association using eq 11 (Figure A2.C). The fit predicts specific activities of 1.2 ± 0.2 and 2.6 ± 0.1 μmol min^{-1} mg^{-1} for the dimer and hexamer, respectively, and a dissociation constant, *K*d, of 1.8 nM^2 for the active hexamer at pH 8.4 (Figure A2.C). On the basis of this analysis, the saturating concentration of substrate and cofactor will stabilize the 0.9 μM hUGDH as a hexamer at pH 8.4. In fact, we have previously used sedimentation velocity analysis to show that saturating concentrations of cofactor and substrate will induce hexamer formation at pH 8.0.[6] Thus, the observed pH optimum of 8.6 for hUGDH is associated with the formation of the active hexamer.

Next, we examined the pH optimum of dimeric hUGDH. We have previously described $hUGDH_{K94E}$, a stable dimer of hUGDH that is the result of a charge-switch substitution designed to disrupt the hexamer.[6] $hUGDH_{K94E}$ has a pH optimum of ∼7.4, but the activity (0.023 μmol min^{-1} mg^{-1}) is 37-fold lower than that of the wild-type enzyme at the same pH (0.85 μmol min^{-1} mg^{-1}) (Figure A2.B). It has been suggested that the reduced activity in the $hUGDH_{K94E}$ dimer may be due to the destabilization of loop 88–110, which is involved in cofactor binding.[22] To address this concern, we designed a new, stable dimer construct ($hUGDH_{M11}$) using a triple substitution (F323T/N324T/T325D) in loop 323–325, which is buried in the hexamer interface. Sedimentation velocity analysis of 9.1 μM $hUGDH_{M11}$ at pH 7.5 reveals a 5.7 S species, consistent with the hUGDH dimer (Figure A2.D). This is in contrast to the wild-type enzyme, which forms a stable hexamer at a much lower protein concentration of 0.91 μM (Figure A2.A). The activity of $hUGDH_{M11}$ is only 4.7-fold lower than that of the wild-type enzyme hexamer at pH 7.4, which is similar to the 5-fold decrease in activity for the hUGDH dimer reported by others.[22] In addition, both $hUGDH_{M11}$ and $hUGDH_{K94E}$ have the same pH optima of 7.4 (Figure A2.E). At pH 8.4, the specific activity of $hUGDH_{M11}$ is ∼22 times lower than what we expected for the wild-type dimer

on the basis of our modeling of the specific activity at pH 8.4 (Figure A2.B,C,E). It may be that the tetramer is not a negligible species at low protein concentrations in the presence of saturating cofactor and substrate, and that the activity of the tetramer is significant. Thus, the activity that we assigned to the dimer in our model may also include a substantial amount of active tetramer. Still, the goal of the modeling experiment was to determine the protein concentration that stabilizes the hexamer in the presence of saturating substrate and cofactor. In that context, the model predicts 0.9 μM hUGDH will stabilize the hexamer under saturating conditions at pH 8.4, which is consistent with our earlier results from sedimentation velocity analysis.[6]

Hysteresis in hUGDH Is Favored at Physiological pH

Because the E* conformation is an intrinsic property of the hexamer,[7] and the hexamer stability is sensitive to pH, we examined the pH dependency of hUGDH hysteresis. We determined the length of the hysteretic lag, defined as the time taken for the enzyme to reach steady-state velocity. The duration of lag, initial velocity (v_i), and steady-state velocity (v_{ss}) were calculated by fitting progress curves to eq 1 (see Materials and Methods). The lag increases at lower pH, suggesting that the stabilities of E* and the hexamer are correlated (Table A2). We estimated the fraction of active enzyme present when the reaction was initiated (E/E*) by examining the ratio of initial and steady-state velocities obtained from the progress curves (v_i/v_{ss}). If we assume that the E* conformation has no activity (the extreme case), then the v_i/v_{ss} ratio shows that the E* conformation is strongly favored at pH 6.5, with only 0.67% of the enzyme in the E state (Table A2). In contrast, at pH 7.5, 10.3% of the enzyme is present in the active state. When the pH increases to 8.5, 64% of hUGDH is observed to exist in the active, E conformation (Table A2). Even if the E* conformation has a low level of activity, it is clear from this analysis that the E* conformation is favored at lower pH.

hUGDH Is More Sensitive to Inhibition by UDP-Xylose at Physiological pH

At pH 8.5, the substrate saturation curves of hUGDH with respect to UDP-Glc show a small degree of negative cooperativity (Hill coefficient of 0.87 ± 0.08) (Figure A3.A and Table A3). In the presence of 20 μM UDP-Xyl, the substrate saturation curves become sigmoidal (Hill coefficient of 1.3 ± 0.08). At the higher concentration of 40 μM UDP-Xyl, the sigmoidicity increases to a Hill coefficient of 1.9 ± 0.07. This type of cooperative behavior in the presence of UDP-Xyl was observed in earlier studies of UGDH.[3, 23] UDP-Xyl stabilizes the horseshoe-shaped hexamer, while the cofactor in the presence of substrate stabilizes the active, 32 symmetry complex (Figure A1.A).[6, 8] In the horseshoe-shaped hexamer, the Thr131 loop occupies the space that would normally accommodate the C5′-hydroxymethyl of UDP-Glc (Figure A1.B).[8] The partially occluded active site would reduce the affinity for UDP-Glc. Thus, the substrate cooperativity we observe in the presence of UDP-Xyl is consistent with the allosteric transition of the horseshoe-shaped complex with low affinity for UDP-Glc into the 32 symmetry hexamer with higher affinity for the substrate (Figure A1.A).

hUGDH is sensitive to product inhibition with a K_i of 27 μM for NADH at pH 7.5.[7] At high protein concentrations, hUGDH can produce inhibitory amounts of NADH before the hysteretic lag ends, thereby complicating steady-state analysis. This can be avoided by keeping protein concentrations low and measuring the amount of NADH produced before reaching steady-state velocity, as we have previously described.[7] Here, our data show that UDP-Xyl inhibition increases the lag (Figure A3.C). Conditions that increase the length of the lag at a constant protein concentration could produce inhibitory concentrations of NADH before reaching the steady state. In fact, at 40 μM UDP-Xyl (~6 x Ki), >18 μM NADH is produced before the lag ends (data not shown). This explains the apparent decrease in kcat for the substrate saturation curve at 40 μM

UDP-Xyl (Figure A3.A and Table A3). Therefore, we excluded the 40 μM UDP-Xyl data from our determination of the K_i for UDP-Xyl. Because UDP-Glc and the allosteric inhibitor compete for the active site, we fit the 0 and 20 μM UDP-Xyl data to a cooperative model for competitive inhibition using global analysis (eq 3 in Materials and Methods and Table A3). This analysis shows that the K_i for UDP-Xyl at pH 8.5 is 6.98 ± 1.6 μM. Next, we examined UDP-Xyl inhibition at pH 7.5. To prevent product inhibition from occurring during the lag at pH 7.5, we reduced the protein concentration to 100 nM. The substrate saturation curve at pH 7.5 is hyperbolic with respect to UDP-Glc (Figure A3.A and Table A3). However, in the presence of 0.6 μM UDP-Xyl, the saturation curves begin to display substrate cooperativity, with a Hill coefficient of 1.24 ± 0.06. The K_i for UDP-Xyl at pH 7.5 is 0.81 ± 0.12 μM, representing a 8.6-fold higher affinity for UDP-Xyl at physiological pH (Figure A3.B and Table A3).

Crystal Structure of the E Complex*

Our data show that physiological pH promotes hUGDH hysteresis, favors UDP-Xyl inhibition, and stabilizes the hexamer. To crystallize the E* complex of hUGDH, we targeted the narrow pH range of 6.0–7.5 using nonionic precipitants for crystallization. Crystals were obtained at pH 7.0 and diffracted to a resolution of 2.7 Å. On the basis of our hysteresis data, >90% of hUGDH should be in the E* conformation at pH 7.0 (Table A2). The crystal structure contains three chains in the asymmetric unit corresponding to the top half of a complete hexamer (Figure A4.A). The complete hexamer is formed by the operation of a crystallographic 2-fold axis. Residues 383–388 in chains A and B, residues 383–387 in chain C, and the C-terminal tail (residues 467–494) in all three chains in the asymmetric unit are disordered.

The NAD^+ and nucleotide sugar binding domains of hUGDH are flexible and can rotate from a closed to an open conformation in a concerted motion without disrupting the hexamer

(Figure A4.B).[5] The open form allows the Thr131 loop to change conformation and repack in response to UDP-Xyl binding.[5] The domains of the E* crystal structure are in the "open" conformation. DynDom analysis shows that the NAD^+ binding domains have rotated open by 10.3° relative to the closed conformation of hUGDH (PDB entry 2Q3E).[24] The hinge bending axis is between residues 219 and 222 and has been described in detail elsewhere.[4, 5] The structure of E* is most similar to that of the $hUGDH_{\Delta 132}$ construct (PDB entry 3TF5); 2730 corresponding Cα atoms of the E* and $hUGDH_{\Delta 132}$ hexamers superimpose with a 0.9 Å root-mean-square deviation. With the exception of the mutation in $hUGDH_{\Delta 132}$, a minor difference in domain conformation is the only variation between the structures; the domains of the E* structure are 2° less open than those in the $hUGDH_{\Delta 132}$ crystal structure. The strong similarities between the E* and $hUGDH_{\Delta 132}$ hexamers are intriguing, given that the $hUGDH_{\Delta 132}$ construct was designed to trap the allosteric Thr131-loop−α6 switch in an intermediate conformation between the active and inactive states.[5] In fact, the most significant observation in the E* structure involves the Thr131 loop−α6 allosteric switch, which does adopt an intermediate conformation (Figure A4.C). In our earlier work, we showed that the hUGDH domains can close only if the Thr131 loop is in the active or inactive state.[5] The intermediate E* conformation traps the domains in the "open" state; closure of the domains would result in steric clashes similar to what we observe in the $hUGDH_{\Delta 132}$ structure.[5]

The Thr131 loop in the E* structure is shifted 0.9 Å from the active conformation and toward the inactive state (Figure A4.C). In this position, the substrate and cofactor binding sites are partially occluded. The position of the Cα atom of Thr131 varies by only 0.1 Å among the three unique molecules in the asymmetric unit. The interface between adjacent dimers of the E* hexamer also adopts a conformation between the active and inactive states (Figure A4.D). To form the inactive state, the feedback inhibitor UDP-Xyl induces the α6 switch helix to translate ~1.43

Å and rotate 10.3°, and the χ1 torsion angle of Ser139 rotates 158° (Figure A4.D).[8] A compensatory repacking of Pro425 and loop 322–324 in the adjacent dimer accommodates the shift of the α6 switch helix. Most notably, the Cα atom of Phe323 translates 2.0 Å and the χ1 torsion angle rotates −60.6°. The interface in the E* structure is in an intermediate conformation in the transition from the active to the horseshoe hexamers. *(i)* The α6 switch helix has translated ∼0.7 Å and rotated by 9.0°. *(ii)* The χ1 torsion angle of Ser139 rotated by 114.0°. *(iii)* The χ1 torsion angle of Phe323 has rotated −8.6°, and the Cα atom has shifted 1.6 Å. On the basis of these observations, we believe that the role of the E* structure is to facilitate feedback inhibition by shifting the conformation of the hUGDH subunits toward the inactive state.

A5. Discussion

The formation of the hUGDH hexamer is under allosteric control.[5, 6, 8] Cofactor binding enhances the affinity among three weakly interacting dimers to induce formation of a stable 32 symmetry hexamer (Figure A1.A,B).[6] Similarly, the feedback inhibitor UDP-Xyl induces formation of a stable but inactive horseshoe-shaped hexamer.[8] Earlier work showed that hysteresis in hUGDH is not due to hUGDH dimers associating to form a more active hexamer; instead, a cofactor and substrate induced isomerization of hUGDH from a less active E* to a more active E conformation (Figure A1.D).[7] Here, we have investigated the role of the E* state in enzyme activity and feedback inhibition. First we show that the E* state is favored at physiological pH. The length of the hysteretic lag in progress curves increases as the pH decreases from 8.5 to 6.5, suggesting that E* is favored at the lower pH. Analysis of the v_i/v_{ss} ratio of the lag also predicts a larger fraction of E*/E at the lower pHs assuming E* is inactive (Table A2). However, even if E* has a low level of activity, the E*/E ratio is consistent with the stabilization of E* at lower pH values. Finally, sedimentation velocity studies show that the hexamer is stabilized at lower pH values

(Figures A1.C and A2.A). This is important because the E* conformation is an intrinsic property of the hexamer (Figure A1.C).[7]

Next, we determined the crystal structure of hUGDH under conditions that stabilize the E* complex. The v_i/v_{ss} ratio suggests that between 90 and 99.3% of hUGDH is in the E* conformation between pH 7.5 and 6.5 (Table A2). Thus, the stable hexamer observed at physiological pH in sedimentation velocity studies most likely represents the E* conformation (Figure A2.A). This is also good evidence that the crystal structure at pH 7.0 represents the authentic E* conformation (Figure A4). The crystal structure of E* closely resembles the $hUGDH_{\Delta 132}$ construct (PDB entry 3TF5), which was designed to mimic an intermediate in the allosteric transition between the active and inactive states.[5] In fact, at physiological pH, the Thr131 loop−α6 switch in hUGDH adopts a conformation that appears to be an intermediate in the transition from the active 32 symmetry hexamer to the inactive, horseshoe-shaped complex (Figure A4.C,D). To accommodate this conformation, the domains of hUGDH have rotated to the open conformation (Figure A4.B). The open domain conformation is essential for the Thr131 loop−α6 switch to change conformations.[5] One result of this intermediate conformation is that the active site is partially occluded, meaning that the Thr131 loop−α6 switch will have to move to bind cofactor and substrate (Figure A4.C,D). We believe that this intermediate conformation represents E* and that the transient observed in progress curves is due to the isomerization of the Thr131 loop−α6 allosteric switch and closing of the domains to adopt the active conformation. But what is the function of the E* conformation in hUGDH? We believe that the E* conformation facilitates feedback inhibition by shifting the Thr131 loop−α6 switch helix into a conformation closer to the inhibited state. In fact, hUGDH is 8.6-fold more sensitive to UDP-Xyl inhibition at pH 7.5 than at pH 8.5 (Figure A3 and Table A3). Previous studies have reported a similar pH dependence of the UDP-Xyl K_i.[23, 25] It is likely that

the hexamer restrains the enzyme in the E* conformation at physiological pH. In fact, the data show that the formation of the hexamer alters the activity of the enzyme. The pH optimum of the hUGDH hexamer is 8.6, but that of the dimer is 7.4 (Figure A2.B,E). The fact that the hexamer and dimer have the same active sites suggests that formation of the hexamer favors an enzyme conformation with an alkaline pH optimum. We also note that the alkaline pH shift in catalytic activity is correlated with a weaker hexameric complex (Figure A2.A). Thus, it would appear that the stable hexamer at physiological pH restrains the structure in the E* conformation.

The observation of a stable, E* conformation also explains accounts of substrate cooperativity only in the presence of UDP-Xyl (Figure A3 and Table A3).[3, 23] The sigmoidicity is consistent with a cooperative transition from the inactive, horseshoe-shaped complex to the active, 32 symmetry hexamer, assuming the former structure has a lower affinity for UDP-Glc. In fact, the conformation of the Thr131 loop−α6 allosteric switch in the active site of the horseshoe-shaped hexamer is specific for UDP-Xyl and introduces a steric clash with the C6′-hydroxyl of UDP-Glc.[8] As UDP-Glc outcompetes UDP-Xyl, the conformation of hUGDH would switch to the 32 symmetry hexamer, which would have a higher affinity for UDP-Glc, thus explaining the substrate cooperativity observed in the presence of UDP-Xyl. But if hUGDH is a cooperative enzyme, why are the saturation curves for UDP-Glc hyperbolic in the absence of UDP-Xyl? Studies have shown that allosteric enzymes whose active and inactive states bind substrate with equal affinity will have hyperbolic substrate saturation curves.[26] However, if a competitive inhibitor shows a distinct preference for binding to the inactive state, the substrate saturation curves will become sigmoidal in the presence of the inhibitor (Figure A3 and Table A3).[26] Thus, if the active and E* conformations of hUGDH both display equal affinity for UDP-Glc, the substrate saturation curves will be hyperbolic. The cooperativity will be observed only after UDP-Xyl triggers the formation

of the horseshoe complex with a reduced affinity for UDP-Glc (Figures A1.B and A3). Similar behavior has been observed in anabolic ornithine transcarbamoylase and l-alanine dehydrogenase, which show cooperativity in substrate saturation curves only in the presence of a competitive inhibitor that favors the inactive conformation of the enzyme.[26-28]

Finally, we turn our attention to the formation of the inactive horseshoe hexamer (Figure A1.A). Specifically, what is the source of the initial asymmetry that causes the 32 symmetry hexamer to dissociate preferentially along one interface to form the horseshoe complex? We believe that substoichiometric binding of UDP-Xyl triggers the asymmetry that produces the horseshoe hexamer (Figure A5). As UDP-Xyl binds to a specific subunit in E*, the Thr131 loop−α6 switch can be switched to the inactive state while the domains remain open. In fact, our earlier work showed that the interface between adjacent dimers in the domain "open" hexamer can adopt the horseshoe conformation without disrupting the 32 symmetry.[5] We have also shown that the interface of the UDP-Xyl-induced hexamer is more stable than the UDP-Glc interface.[5, 8] Thus, substoichiometric binding of UDP-Xyl can result in an "open" domain hexamer with a mix of strong and weak interfaces (Figure A5). The domains in such a hexamer can close in one of two ways. *(i)* To maintain 32 symmetry, the domains must disrupt the stable contacts in the UDP-Xyl-buttressed interfaces to rotate to the "closed" state, or *(ii)* to form the horseshoe hexamer, the weaker UDP-Xyl-free interface can dissociate to allow the dimers to rotate as rigid bodies and close the domains without disrupting the two remaining interfaces.

On the basis of the results we have presented here, we propose that at physiological pH, the hexameric structure of hUGDH restrains the enzyme in a less active conformation that favors feedback inhibition by UDP-Xyl. This conformational leveraging allows the small difference in the effector molecule (the absence of a C5′-hydroxymethyl) to induce a disproportionately larger

allosteric transition. This "conformational leveraging" by the formation of a higher-order quaternary structure is not without precedent. A similar behavior is observed in mammalian hemoglobin. The individual αβ hemoglobin dimers have a high affinity for oxygen.[29] However, formation of the $(\alpha\beta)_2$ tetramer restrains the subunits to a lower-affinity conformation, which is essential for oxygen delivery.

A6. Figures and Tables

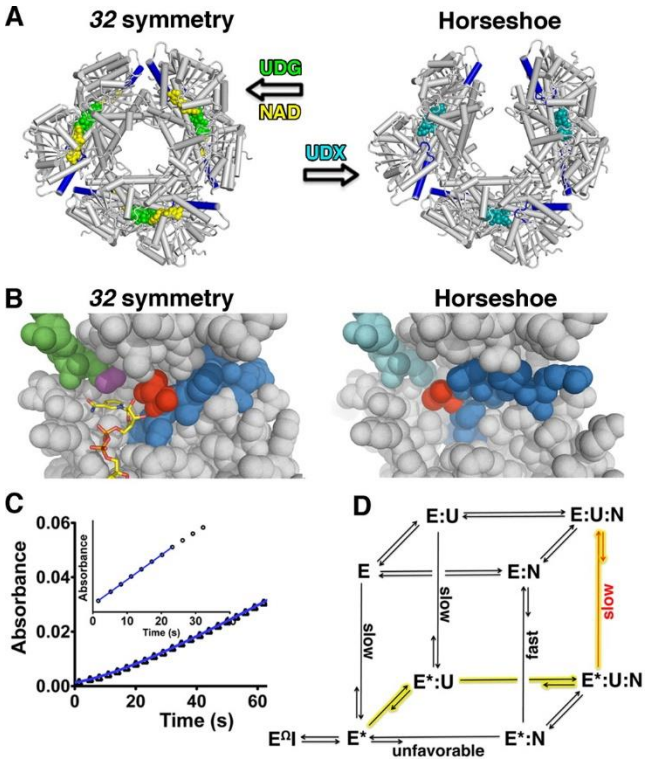


Figure A1: Allostery and hysteresis in hUGDH. (A) hUGDH hexamer depicted in the active, 32 symmetry and inactive horseshoe-shaped conformations. UDP-Glc (UDG, green), NAD^+ (yellow), and UDP-Xyl (UDX, cyan) are depicted as spheres. Changes in the Thr131 loop (blue loop) conformation will alter the affinity of the subunit interface via the α6 helix (blue rod).[8] For the sake of clarity, only the top trimer of dimers in the hexamer is depicted. **(B)** Active site showing the Thr131 loop (blue spheres with Thr131 colored red) in the 32 symmetry and horseshoe conformations. NAD^+ (yellow sticks) buttresses the Thr131 loop to increase the affinity between subunit interfaces and induce formation of the active hexamer.[6] The Thr131 loop switches conformation in response to UDP-Xyl (cyan) binding, to occlude the binding sites for NAD^+ and the C5′-hydroxylmethyl (magenta spheres) of UDP-Glc (green spheres).[8] The inhibited conformation also increases the affinity between dimers to induce formation of the horseshoe hexamer. For the sake of clarity, residues burying the active site and Thr131 loop have been removed. **(C)** hUGDH progress curves showing that hysteresis is a property of the hexamer and that there is no lag in the dimer (inset). The lag is fit (blue line) with eq 1 from Materials and Methods. This figure is modified from ref 7. **(D)** Ligand-induced structural transition (LIST) model[10] for hUGDH hysteresis based on our earlier work.[7] Inactive complexes are identified with an asterisk, and E* is the thermodynamically favored conformation of the apoenzyme. Unequal arrows identify favored directions supported by our earlier kinetic studies.[7] The binding of UDP-Glc (U) to E* is >104-fold faster than the binding of NAD^+ (N) and thus represents the kinetically preferred path of substrate addition (highlighted in yellow).[7] Subsequent binding of cofactor N to form the ternary E*:U:N complex induces a slow isomerization to the active, E:U:N Michaelis complex. This transition is observed as the lag in progress curves. Turnover of the complex produces E, which can rapidly bind substrate and cofactor to form the E:U:N complex or slowly

convert to the more stable E* at low substrate concentrations. It is possible that the E*:U:N complex is active, but turnover would necessarily be slow to cause hysteresis. The isomerization of the E*:N complex to the E:N complex is inferred from our previous work showing that binding of U to the saturated E:N complex does not display hysteresis.[7]

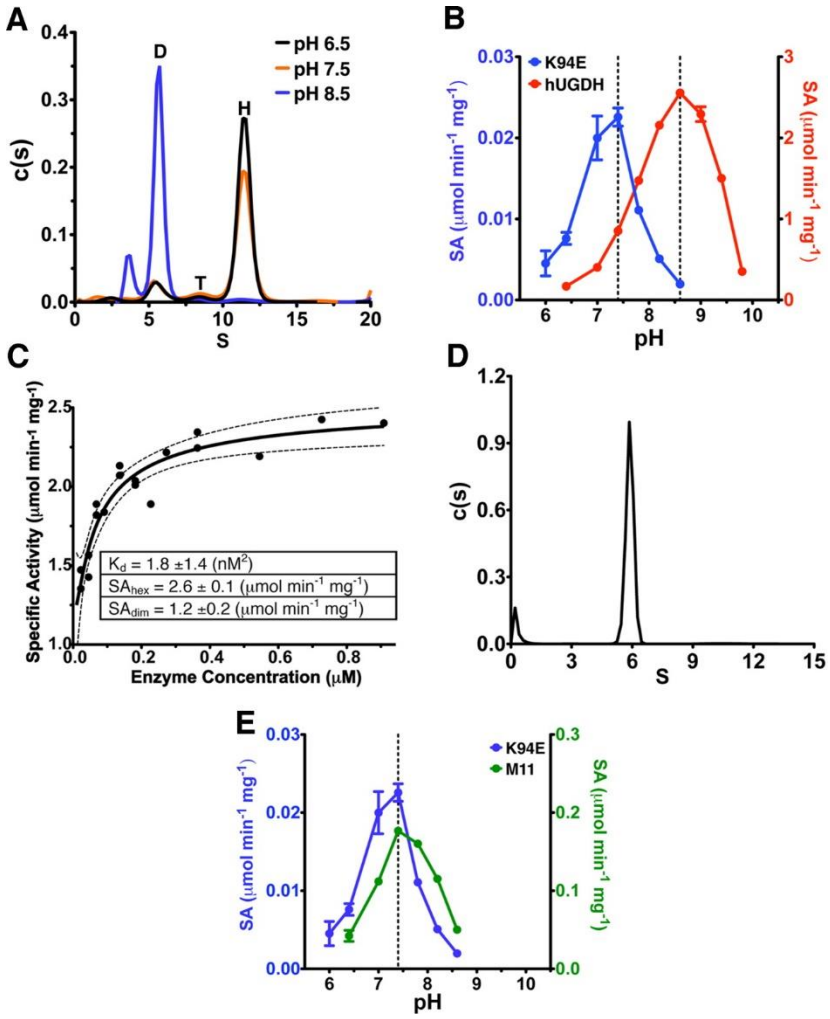


Figure A2: Hexamer stability and activity are pH sensitive. (A) The *c(s)* sedimentation velocity distributions show that hexameric hUGDH (H, 11.4 S) is stable at pH 6.5–7.5 while the dimer (D, 5.7 S) dominates at pH 8.5. A small amount of tetrameric hUGDH (T, 8.6 S) is also observed. **(B)** Wild-type hUGDH (red line) and dimeric $hUGDH_{K94E}$ (blue) have pH optima of 8.6 and 7.4, respectively. Specific activities (SAs) are plotted in corresponding colors on opposite Y-axes. **(C)** The specific activity of hUGDH shows a hyperbolic dependency on enzyme concentration at pH 8.4, consistent with the formation of a more active hexamer. The data were fit (black line) using eq 11. The dotted lines identify the 95% confidence intervals of the fit. The modeled specific activities for the hexamer (SA_{hex}) and dimer (SA_{dim}) are indicated, as well as the dissociation constant $K_d = 1/K_a$ (see the text). **(D)** The *c(s)* sedimentation velocity distribution of 9.1 μM hUGDHM11 at pH 7.5 reveals a single species at 5.7 S, confirming the dimer structure. (E) Dimeric $hUGDH_{K94E}$ (blue line) and $hUGDH_{M11}$ (green line) have the same pH optima of 7.4, despite different rates. Specific activities (SAs) are plotted in corresponding colors on opposite Y-axes.

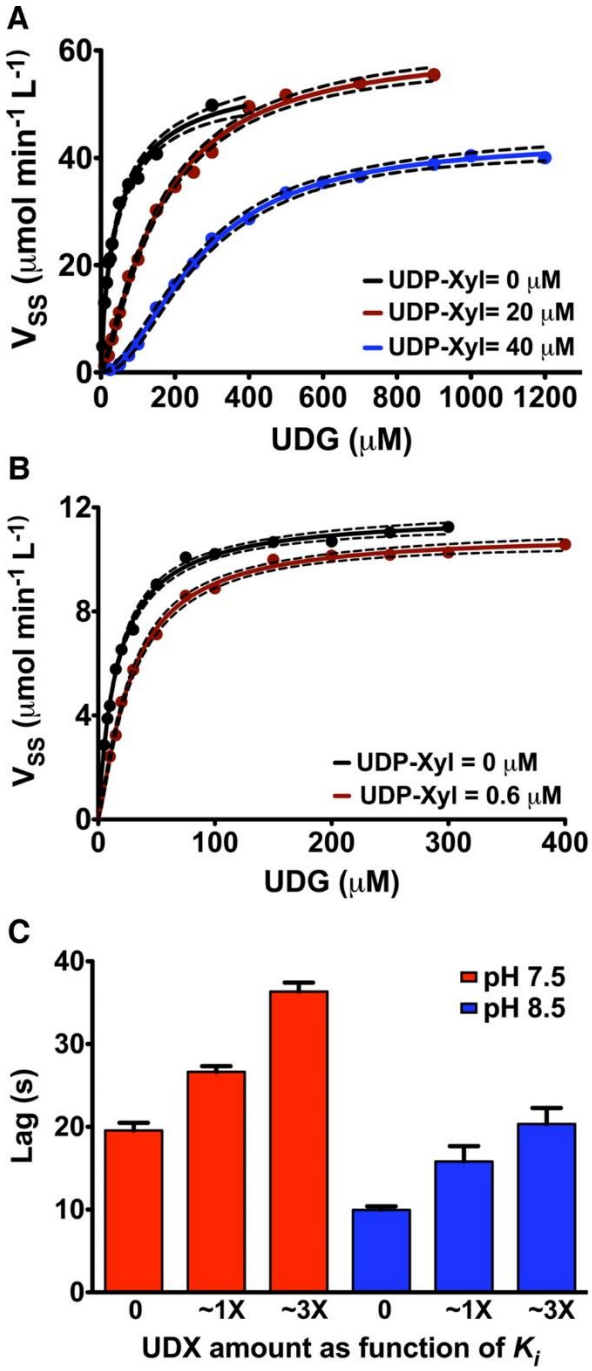


Figure A3: UDP-xylose inhibition is sensitive to pH. (A) At pH 8.5, steady-state analysis of hUGDH in the presence of 0 (black), 20 (red), and 40 μM (blue) UDP-Xyl (UDX) (see Table A3 for parameters). Steady-state velocities were determined with eq 1. Because of the increased lag and resulting product inhibition with 40 μM UDP-Xyl, only the 0 and 20 μM UDX data were globally fit to the equation for competitive inhibition, modified for cooperativity (see the text and Materials and Methods). Dotted lines indicate 95% confidence intervals. **(B)** Same as above, but at pH 7.5 with UDP-Xyl at 0 (black) and 0.6 μM (red). **(C)** The hUGDH hysteretic lag increases with UDP-Xyl concentration at pH 7.5 (red) and 8.5 (blue). UDP-Xyl concentrations are reported in units of K_i (0.81 and 6.98 μM for pH 7.5 and 8.5, respectively).

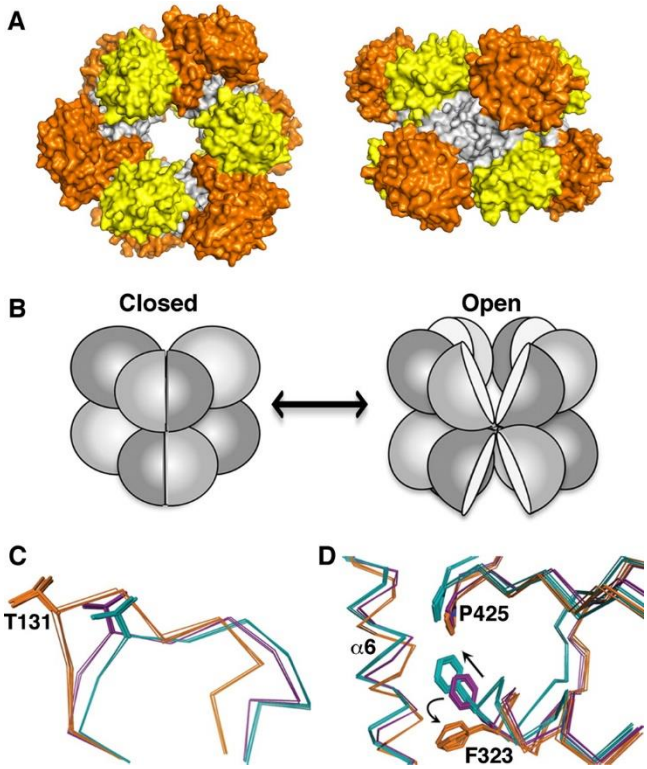


Figure A4: Crystal structure of the E* complex. (A) Top and side views of the E* hexamer. The NAD^+ (orange), nucleotide sugar binding domain (yellow) and dimerization domain (gray) are depicted. **(B)** Cartoon showing the closed and open conformations of the NAD^+ and nucleotide sugar binding domains. **(C)** Superposition of the Thr131 loop of the E* (magenta), active 32 symmetry hexamer (cyan, PDB entry 2Q3E) and the horseshoe-shaped hexamer (orange, PDB entry 3PTZ). To illustrate the small degree of variability in the position of the Thr131 loop, all of the unique chains contained in the asymmetric unit of each crystal structure are superimposed. **(D)** Hexamer building interface illustrated as in panel C. The α6 helix (α6), loop 322–324 (F323), and Pro425 (P425) are shown. Black arrows depict the motions that Phe323 undergoes to switch states.

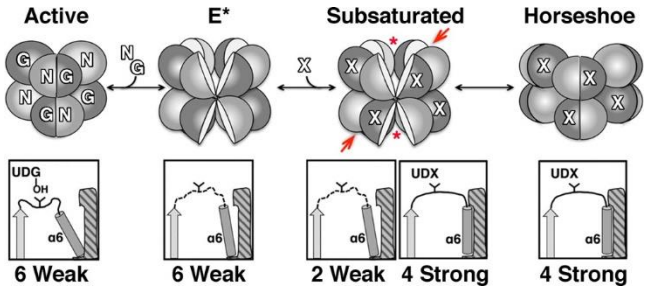


Figure A5: Model for the activation and UDP-Xyl inhibition of the E* hexamer. The hexamer is depicted in the active, E*, subsaturated, and horseshoe states. The insets below each state illustrate the conformation of Thr131 loop−α6 switch as a loop connecting a gray strand to the α6 helix (gray cylinder). The hexamer interface is shown as a hatched surface complementary to the α6 helix. The loop is depicted in ligand-bound (—) and intermediate (---) conformations. Thr131 (the "Y" structure in the loop) interacts with UDP-Glc (UDG) or UDP-Xyl (UDX). The strong and weak interfaces are identified. The binding of NAD^+ (N) and UDP-Glc (G) to E* generates the closed, active structure with a weak, flexible hexamer interface.[5] The binding of substoichiometric amounts of UDP-Xyl (X) produces a subsaturated complex with a mix of strong and weak interfaces. The UDP-Xyl-bound subunits rotate the α6 helix to create a strong interface. The red asterisk indicates the weaker interface between the two unbound subunits (red arrow). Dissociation along the weak interface will allow the subunits to undergo a rigid body rotation to close the domains without disrupting the UDP-Xyl-stabilized interfaces.

Table A1: Data Collection and Refinement Statistics

Data Collection	
space group	C2
unit cell dimensions (*a*, *b*, *c*, β)	176.98 Å, 113.96 Å, 97.12 Å, 116.9°
completeness (%)	99.3 (97.4)[a]
no. of reflections	350028
redundancy	7.4 (7.3)
$I/\sigma(I)$	19.6 (1.5)
R_{meas}[b] (%)	12.7 (162.5)
$CC_{1/2}$[c]	0.999 (0.655)
Refinement	
resolution (Å)	2.7
R_{work}/R_{free}[d]	0.18/0.23
no. of atoms	
protein	10830
ligand	4
water	8
B factor (Å^2)	
protein	86.8
ligand	105.7
water	62.7

stereochemical ideality	
bond lengths ($Å^2$)	0.009
bond angles (deg)	1.1
φ,ψ most favored (%)[e]	97.4
φ,ψ additionally allowed (%)[e]	2.6

[a] Values in parentheses are for the highest-resolution shell (2.7–2.86 Å).
[b] Rmeas is the redundancy-independent merging *R* factor of Diederichs and Karplus.[30]
[c] $CC_{1/2}$ is the correlation between intensities from random half-data sets.[31]
[d] An a posteriori R_{free} was obtained as explained in Materials and Methods.
[e] Values obtained from structure validation using MolProbity.[32]

Table A2: pH Dependence of Hysteresis

pH	v_i (μM/min)[a]	v_{ss} (μM/min)[a]	v_i/v_{ss} (%)	τ (s)[a]	k_{obs} (s^{-1})[b]	lag (s)[c]
6.5	0.047 ± 0.064	7.0 ± 0.05	0.67	70.5 ± 1.8	0.014	191.5 ± 4.8
7.5	2.7 ± 1.7	26.1 ± 0.07	10.3	7.9 ± 0.5	0.127	21.4 ± 1.4
8.5	34 ± 13.3	53.1 ± 1.5	64	3.67 ± 0.22	0.272	9.97 ± 0.6

[a] Data for the progress curves were collected in triplicate except for those at pH 8.5, which are duplicates. Data were fit to eq 1 using global analysis to calculate v_i, v_{ss}, and τ.
[b] $k_{obs} = 1/\tau$.
[c] Lag was calculated using the equation lag = eτ.

Table A3: Global Fit of Data from UDP-Xylose Inhibition Studies

pH	[UDP-Xyl] (μM)	[hUGDH] (μM)	UDP-Glc K_M (μM)[a]	k_{cat} (s^{-1})[a]	UDP-Xyl K_i (μM)[a]	Hill (h)[a]
8.5	0	0.23	43.4 ± 6.7	4.17	6.98 ± 1.6	0.87 ± 0.08
	20			4.62		1.3 ± 0.08
	40[b]			3.12		1.9 ± 0.07
7.5	0	0.10	16.1 ± 0.79	1.97	0.81 ± 0.12	1.02 ± 0.05
	0.6			1.83		1.24 ± 0.06

[a] The kinetic parameters, K_M, K_i, k_{cat}, and h, were calculated using a global analysis of the inhibition data using eqs 2 and 3.

[b] The 40 μM UDP-Xyl sample was not used in calculating K_i because of significant product inhibition (see the text).

A7 References

(1) Kalckar, H. M., Maxwell, E. S., and Strominger, J. L. (1956) Some properties of uridine diphosphoglucose dehydrogenase. *Arch. Biochem. Biophys*. 65, 2−10.

(2) Gainey, P. A., and Phelps, C. F. (1972) Uridine diphosphate glucuronic acid production and utilization in various tissues actively synthesizing glycosaminoglycans. *Biochem. J*. 128, 215−227.

(3) Balduini, C., Brovelli, A., De Luca, G., Galligani, L., and Castellani, A. A. (1973) Uridine diphosphate glucose dehydrogenase from cornea and epiphysial-plate cartilage. *Biochem. J*. 133, 243−249.

(4) Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U., and Nidetzky, B. (2010) UDP-glucose dehydrogenase: Structure and function of a potential drug target. *Biochem. Soc. Trans.* 38, 1378−1385.

(5) Sennett, N. C., Kadirvelraj, R., and Wood, Z. A. (2011) Conformational Flexibility in the Allosteric Regulation of Human UDP-α-D-Glucose 6-Dehydrogenase. *Biochemistry* 50, 9651−9663.

(6) Sennett, N. C., Kadirvelraj, R., and Wood, Z. A. (2012) Cofactor Binding Triggers a Molecular Switch To Allosterically Activate Human UDP-α-D-glucose 6-Dehydrogenase. *Biochemistry* 51, 9364−9374.

(7) Kadirvelraj, R., Sennett, N. C., Custer, G. S., Phillips, R. S., and Wood, Z. A. (2013) Hysteresis and negative cooperativity in human UDP-glucose dehydrogenase. *Biochemistry* 52, 1456−1465.

(8) Kadirvelraj, R., Sennett, N. C., Polizzi, S. J., Weitzel, S., and Wood, Z. A. (2011) Role of Packing Defects in the Evolution of Allostery and Induced Fit in Human UDP-Glucose Dehydrogenase. *Biochemistry* 50, 5780−5789.

(9) Dickinson, F. M. (1988) Studies on the unusual behaviour of bovine liver UDP-glucose dehydrogenase in assays at acid and neutral pH and on the presence of tightly bound nucleotide material in purified preparations of this enzyme. *Biochem. J*. 255, 775−780.

(10) Frieden, C. (1970) Kinetic aspects of regulation of metabolic processes. The hysteretic enzyme concept. *J. Biol. Chem*. 245, 5788−5799.

(11) Kabsch, W. (2010) Xds. *Acta Crystallogr*. D66, 125−132

(12) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., GrosseKunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr*. D66, 213−221.

(13) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr*. D66, 486−501.

(14) Urzhumtsev, A., Afonine, P. V., and Adams, P. D. (2013) TLS from fundamentals to practice. *Crystallogr. Rev*. 19, 230−270.

(15) Brunger, A. T. (1997) Free R Value: Cross-validation in Crystallography. *Methods Enzymol*. 277, 366−396.

(16) Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J*. 78, 1606−1619.

(17) Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical ultracentrifugation in biochemistry and polymer science* (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.) pp 90−125, The Royal Society of Chemistry, Cambridge, U.K.

(18) Cornish-Bowden, A. (2001) Detection of errors of interpretation in experiments in enzyme kinetics. *Methods* 24, 181−190.

(19) Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem*. 320, 104−124.

(20) Gainey, P. A., Pestell, T. C., and Phelps, C. F. (1972) A study of the subunit structure and the thiol reactivity of bovine liver uridine diphosphate glucose dehydrogenase. *Biochem. J*. 129, 821−830.

(21) Hyde, A. S., Farmer, E. L., Easley, K. E., van Lammeren, K., Christoffels, V. M., Barycki, J. J., Bakkers, J., and Simpson, M. A. (2012) UDP-glucose dehydrogenase polymorphisms from patients with congenital heart valve defects disrupt enzyme stability and quaternary assembly. *J. Biol. Chem*. 287, 32708−32716.

(22) Hyde, A. S., Thelen, A. M., Barycki, J. J., and Simpson, M. A. (2013) UDP-glucose dehydrogenase activity and optimal downstream cellular function require dynamic reorganization at the dimer-dimer subunit interfaces. *J. Biol. Chem*. 288, 35049−35057.

(23) Gainey, P. A., and Phelps, C. F. (1975) Interactions of urdine diphosphate glucose dehydrogenase with the inhibitor urdine diphosphate xylose. *Biochem. J*. 145, 129−134.

(24) Hayward, S., and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* 30, 144−154.

(25) Huang, Y. H., Roy-Burman, P., and Visser, D. W. (1971) Uridine diphosphate glucose dehydrogenase of calf liver. Properties and inhibition characteristics with uridine diphosphate xylose analogues. *Biochem. Pharmacol*. 20, 2447−2458.

(26) Kuo, L. C. (1983) Allosteric cofactor-mediated enzyme cooperativity: A theoretical treatment. *Proc. Natl. Acad. Sci.* U.S.A. 80, 5243−5247.

(27) Kuo, L. C., Lipscomb, W. N., and Kantrowitz, E. R. (1982) Zn(II)-induced cooperativity of Escherichia coli ornithine transcarbamoylase. *Proc. Natl. Acad. Sci.* U.S.A. 79, 2250−2254.

(28) Kim, S. J., Kim, Y. J., Seo, M. R., and Jhun, B. S. (2000) Regulatory mechanism of L-alanine dehydrogenase from Bacillus subtilis. *Bull. Korean Chem. Soc.* 21, 1217−1221.

(29) Hewitt, J. A., Kilmartin, J. V., Eyck, L. F., and Perutz, M. F. (1972) Noncooperativity of the dimer in the reaction of hemoglobin with oxygen (human-dissociation-equilibrium-sulfhydryl-absorption-Xray analysis). *Proc. Natl. Acad. Sci.* U.S.A. 69, 203−207.

(30) Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269−275 592 (Erratum).

(31) Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* 336, 1030−1033.

(32) Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr*. D66, 12−21.

APPENDIX B

ALLOSTERY AND HYSTERESIS ARE COUPLED IN HUMAN UDP-GLUCOSE DEHYDROGENASE

Reprinted here with permission from: Beattie, N.R., **Keul, N.D.**, Sidlo, A.M. and Wood, Z.A., Allostery and Hysteresis are coupled in human UDP-glucose dehydrogenase. *Biochemistry*, 56(1), pp.202-211 (2016).

Keul, N.D. conducted and analyzed sedimentation velocity experiments. In addition, Keul, N.D. contributed in manuscript revisions.

B1. Abstract

Human UDP-glucose dehydrogenase (hUGDH) is regulated by an atypical allosteric mechanism in which the feedback inhibitor UDP-xylose (UDP-Xyl) competes with the substrate for the active site. Binding of UDP-Xyl triggers the T131-loop/α6 allosteric switch, which converts the hexameric structure of hUGDH into an inactive, horseshoe-shaped complex (E^{Ω}). This allosteric transition buries residue A136 in the protein core to produce a subunit interface that favors the E^{Ω} structure. Here we use a methionine substitution to prevent the burial of A136 and trap the T131-loop/α6 switch in the active conformation. We show that hUGDHA136M does not exhibit substrate cooperativity, which is strong evidence that the methionine substitution prevents the formation of the low-UDP-Glc-affinity E^{Ω} state. In addition, the inhibitor affinity of $hUGDH_{A136M}$ is reduced 14-fold, which most likely represents the K_i for competitive inhibition in the absence of the allosteric transition to the higher-affinity E^{Ω} state. hUGDH also displays a lag in progress curves, which is caused by a slow, substrate-induced isomerization that activates the enzyme. Stopped-flow analysis shows that $hUGDH_{A136M}$ does not exhibit hysteresis, which suggests that the T131-loop/α6 switch is the source of the slow isomerization. This interpretation is supported by the 2.05 Å resolution crystal structure of $hUGDH_{A136M}$, which shows that the A136M substitution has stabilized the active conformation of the T131-loop/α6 allosteric switch. This work shows that the T131-loop/α6 allosteric switch couples allostery and hysteresis in hUGDH.

B2. Introduction

Glucuronidation is a major component of the phase II metabolism of drugs.[1-3] This pathway has been shown to be upregulated in some lung, breast, and colorectal cancer cells, where it functions as an intrinsic drug resistance mechanism.[4-8] Thus, controlling glucuronidation is a promising strategy for sensitizing this class of tumors to existing chemotherapeutics. It is possible to inhibit glucuronidation by limiting the availability of the essential substrate, UDP-glucuronic acid.[9-11] UDP-glucose dehydrogenase (UGDH) catalyzes the NAD^+-dependent oxidation of UDP-glucose (UDP-Glc) to produce UDP-glucuronic acid.[12-15] Understanding how UGDH activity is regulated is an important goal in developing new strategies for controlling glucuronidation-dependent drug resistance.

Human UGDH (hUGDH) forms a hexamer that is regulated by an atypical allosteric mechanism in which the feedback inhibitor UDP-xylose (UDP-Xyl) competes with the substrate for the active site.[16-19] Here, allostery arises from the distinct conformational changes that are induced by the binding of the substrate or inhibitor (Figure B1.A). The binding of UDP-Glc favors the formation of an active 32 symmetry hexamer called the E state. In contrast, UDP-Xyl binding produces an inactive, horseshoe-shaped hexamer (E^{Ω}). These effector specific transitions are controlled by a buried allosteric switch called the T131-loop/α6 helix (Figure B1.A–C). To select for a specific hexamer conformation, the NAD^+ binding domain (NB) of hUGDH must adopt an "open" conformation (the E* or E* bound state) that exposes the allosteric switch (Figure B1.D). Once exposed, the T131-loop and α6 helix are free to change conformation in response to UDP-Glc or UDP-Xyl binding to produce the E or E^{Ω} hexamer, respectively (Figure B1.D).[16-18] Thus, the active site of hUGDH also functions as an allosteric site in that it controls the structure of the hexamer-building interface and the affinity between subunits.[16, 17] To the best of our knowledge,

the only other enzyme known to have a similar bifunctional active/allosteric site is dCTP-deaminase.[20, 21]

hUGDH also displays hysteresis, which can be observed as a lag in progress curves.[16-19, 22, 23] In the absence of any effector, hUGDH favors the inactive E* conformation (Figure B1.D).[16] Hysteresis is caused by coenzyme and substrate binding, which induces the E* state to slowly isomerize to the active E conformation.[22] Still, the molecular basis of the hysteretic transition was not known. E* adopts an open domain structure that stabilizes the allosteric switch and the hexamer-building interfaces in an intermediate conformation between the E and E^{Ω} states[16, 18, 24](Figure B1.B–D). When an effector binds to E*, the T131-loop/α6 helix and eight surrounding residues repack into the appropriate E or E^{Ω} conformation (Figure B1.D).[16] We have proposed that the repacking of the allosteric switch and adjacent residues from the E* to E conformation is the source of hysteresis observed in hUGDH.[22] Here we have used an amino acid substitution (A136M) to stabilize the E state and test our hypothesis.

B3. Materials and Methods

Protein Expression, Crystallization, and Structure Solution

$hUGDH_{A136M}$ and wild-type hUDGH were recombinantly expressed in Escherichia coli as previously described.[16-19, 22] Following purification, the His tags were removed from both proteins using TEV protease. The purified protein was dialyzed into storage buffer [25 mM Tris (pH 8.0) and 50 mM NaCl] and concentrated using a Millipore Amicon Ultra-15 10K centrifugal filter unit to ~20 mg/mL. Proteins were quantified using an Agilent 8453 UV/vis instrument with an ε280 of 48360 M^{-1} cm^{-1} and an ε280 of 49850 M^{-1} cm^{-1} for hUGDH and $hUGDH_{A136M}$, respectively. The molar absorptivities for these proteins were calculated using the amino acid sequence and Protparam.[25] $hUGDH_{A136M}$ was crystallized at 20 °C using a hanging drop vapor diffusion method

with a 2 μL drop mixed in a 1:1 ratio of protein to reservoir (final protein concentration of 10 mg/mL). The reservoir was comprised of 0.2 M NaCl, 12% PEG 3350, and 0.1 M Tris buffer (pH 7.6). Crystals were cryoprotected using a solution matching the reservoir supplemented with 20% of a cryoprotectant mixture (a 1:1:1 dimethyl sulfoxide:ethylene glycol:glycerol ratio) and then cooled by being plunged into liquid nitrogen. A 2.05 Å resolution data set was collected on the 21-ID beamline (SER-CAT) at the Argonne National Laboratory (Argonne, IL) using a Rayonix MX300HS CCD detector. The data set was processed using XDS[26] and 5% of the data set aside for cross validation.[27] The data collection statistics are listed in Table B1.

The structure of $hUGDH_{A136M}$ was solved by molecular replacement using Protein Data Bank (PDB) entry 4RJT as a search model in the PHENIX software suite.[28] The hUGDHA136M model was subjected to iterative cycles of manual rebuilding using COOT[29] followed by automated refinement with NCS restraints. *B* factors were refined using TLS as implemented in PHENIX.[28, 30] We used the following rationale in modeling the occupancy of the copurified UDP-Glc. At resolutions lower than 1.5 Å, B factors and occupancy are strongly correlated, which makes it impossible to refine both of these parameters. Still, it is possible to approximate occupancy because ligands and interacting residues are similarly ordered. Briefly, the occupancy of the ligand is fixed at different values, and only the *B* factors are refined. The approximate occupancy is the value at which the ligand *B* factors converge to those of the interacting residues. Final model refinement statistics are listed in Table B1.

Sedimentation Velocity

hUGDH and $hUGDH_{A136M}$ were dialyzed into buffers containing 150 mM KCl and 25 mM buffer [HEPES (pH 7.5) or TRIS (pH 8.5)], quantified (as described above), and then diluted to a final protein concentration of 9 μM. Samples were loaded into 12 mm double-sector Epon centerpieces

equipped with quartz windows and equilibrated for 1 h at 20 °C in an An60 Ti rotor. Sedimentation velocity data were collected in an Optima XLA analytical ultracentrifuge using a rotor speed of 50000 rpm at 20 °C. Sedimentation data were recorded at 280 nm in radial step sizes of 0.003 cm. SEDNTERP[31] was used to estimate the partial specific volume of hUGDH (0.73840 mL/g) and $hUGDH_{A136M}$ (0.73842 mL/g), and the densities of both the pH 7.5 (1.00726 g/mL) and pH 8.5 (1.00603 g/mL) buffers. Viscosities for both pH 7.5 and 8.5 buffers were calculated to be 0.01018 and 0.01007 P, respectively. SEDFIT[32] was used to analyze raw sedimentation data. Data were modeled as a continuous sedimentation coefficient, *c(s)*, distribution and were fit using the baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Theoretical sedimentation coefficient (s) values were calculated from hUGDH atomic coordinates under standard conditions using HYDROPRO20 (www.bbri.org/RASMB).

Stopped-Flow Analysis of Hysteresis

NAD^+ and UDP-glucose were purchased from Sigma, and UDP-xylose was purchased from Carbosource (Complex Carbohydrate Research Center, University of Georgia, Athens, GA). Enzyme hysteresis was monitored at 25 °C using an Olis RSM 1000 Rapid-scanning Absorbance and Fluorescence Spectrophotometer with a stopped-flow assembly and a 0.4 mm path length. The 480 nM enzyme solution in 50 mM HEPES (pH 7.5), 50 mM NaCl, and 5 mM EDTA was rapidly mixed with an equal volume of 1 mM UDP-Glc and 10 mM NAD^+ in the same buffer [final concentrations of 240 nM enzyme, 0.5 mM UDP-Glc, 5 mM NAD^+, 50 mM HEPES (pH 7.5), 50 mM NaCl, and 5 mM EDTA]. Absorbance readings at 340 nm were taken every 0.1 s for a total of 120 s to monitor NADH production. hUGDH progress curves were fit to Frieden's equation[33] describing enzyme hysteresis:

$$P(t) = v_{ss}t - \tau(v_{ss} - v_i)(1 - e^{-t/\tau}) \qquad \text{(equation 1)}$$

where P is the concentration of the product at time t and τ is equal to $1/k_{obs}$, where k_{obs} is the apparent rate constant for the transition between the initial velocity (v_i) and the steady state velocity (v_{ss}). Here, the initial velocity, v_i, in eq 1 is the pre-steady state velocity and does not obey the steady state approximation. The length of the lag in seconds is calculated as the product of Euler's number and τ ($e\tau$). Data were fit using PRISM (GraphPad Software Inc., San Diego, CA), and all fits were analyzed using residual plot analysis.[34]

Steady State Kinetics

The assay conditions for the steady state analysis of hUGDH have been described previously.[16-19, 22] Briefly, assays contained 100 nM hUDGH or 230 nM $hUGDH_{A136M}$ in a standard reaction buffer of 50 mM HEPES (pH 7.5), 50 mM NaCl, and 5 mM EDTA with a saturating concentration of either 2.5 mM NAD^+ or 1 mM UDP-glucose. Prior to the reaction, the substrate and enzyme were separately preincubated at 25 °C for 5 min, and the reaction was initiated by rapidly mixing the enzyme and substrate solutions. Progress curves were measured by monitoring NADH production at 340 nm (molar absorptivity coefficient of 6220 M^{-1} cm^{-1}) with 0.5 s data points using an Agilent 8453 UV/vis spectrometer at 25 °C. All data were fit using nonlinear regression from PRISM. hUDGH steady state velocities were calculated by fitting progress curves to eq 1, as we have previously described.[16-19, 22] For a hysteretic enzyme, the initial velocity does not satisfy the steady state approximation. Thus, the v_{ss} prior to the depletion of 10% of the substrate represents the initial steady state velocity. $hUGDH_{A136M}$ did not display hysteresis; thus, the initial velocities satisfy the steady state approximation and were determined using the linear portion of the progress curves. Initial steady state velocities were fit to a sigmoidal rate equation and analyzed with residual analysis:[34]

$$v = \frac{k_{cat}[E_t][S]^h}{K_M{}^h + [S]^h} \qquad \text{(equation 2)}$$

where h is the Hill coefficient. Negative cooperativity in the NAD^+ saturation curve of hUGDH was evaluated using Kurganov's analysis of a concave-up Eadie–Hofstee plot.[35] Briefly, data were fit to the equation:

$$\frac{v}{[S]} = \frac{V_{max} - v}{K_m^{eff}}, \text{ where } K_m^{eff} = K_0 + (K_{lim} - K_0)(v/V_{max}) \quad \text{(equation 3)}$$

where K_0 is the estimated K_M for the high-affinity binding sites and K_{lim} is the average of the lower-affinity binding sites.

We determined the K_i for the inhibitor UDP-Xyl as previously described.[22] Briefly, UDP-Xyl competes with UDP-Glc for the active site and induces hUGDH to form the E^{Ω} state.[17, 22] Because the E and E^{Ω} states have different affinities for UDP-Glc and UDP-Xyl, the inhibition studies display cooperativity. To determine the K_i, the substrate saturation curves from different concentrations of UDP-Xyl were fit simultaneously to the equation for competitive inhibition corrected for substrate cooperativity using global analysis in PRISM:

$$v_o = \frac{k_{cat}[E_t][S]^h}{(K_M^{app})^h + [S]^h}, \text{ where } K_M^{app} = K_M\left(1 + \frac{[I]}{K_i}\right) \quad \text{(equation 4)}$$

B4. Results

Design and Rationale for the A136M Substitution

The A136M substitution was designed to prevent the isomerization of the T131-loop/α6 allosteric switch. In the active conformation of hUGDH, the Cβ atom of A136 is solvent-exposed (Figure B2.A). To form the inhibited E^{Ω} state, the α6 helix tilts ~10.6° and then rotates about the helix axis ~12°, burying A136 near a cavity in the core of the protein (Figure B2.B).[17] We hypothesized that an amino acid substitution could prevent the burial of A136 and effectively lock the α6 helix in the active conformation. We excluded polar and charged residues from consideration because A136 is located in the hexamer-building interface, and the burial of an

unsatisfied electrostatic group might destabilize the hexamer. We also excluded β-branched amino acids based on a modeling experiment with valine at position 136; each preferred rotamer of valine introduced at least one steric clash between a γ-methyl and the surrounding residues of the A136 pocket (Figure B2.C). Similarly, leucine, phenylalanine, and tryptophan are also expected to introduce steric clashes. The structure of the A136 surface pocket permits only an unbranched amino acid substitution with a χ1 torsion angle of 180° (Figure B2.C). On the basis of these observations, we chose the A136M substitution; methionine not only satisfies the criteria mentioned above but also is flexible enough to conform to the local packing constraints in the A136 pocket.

Structural Analysis of hUGDH$_{A136M}$

The crystal structure of hUDGH$_{A136M}$ was solved in space group P1 and refined to a resolution of 2.05 Å (Table B1). The unit cell contains a single 32 symmetry hexamer (Figure B3.A). The residues in loop 385–387 and C-terminal residues 467–494 are disordered in each monomer and are not included in the final model. All 466 Cα atoms in each of the six monomers superimpose with root-mean-square deviations ranging from 0.344 to 0.834 Å. The largest structural differences between the chains involve variability in the amount of hinge-bending motion of the NB (residues 1–212) and SB (residues 323–466) domains about an axis located between residues 219 and 220 (Figure B3.B). A DynDom[36] comparison of hUDGH$_{A136M}$ to the closed conformation of hUGDH (represented by PDB entry 2Q3E) shows that the NB and SB domains of chains A–F are rotated open by 11.2°, 9.4°, 9.1°, 10°, 12.4°, and 5.9°, respectively. In the SB domain of chain C, we observed electron density consistent with a weakly ordered nucleotide sugar that has copurified with the enzyme (Figure B3.C). We have modeled the density as a UDP-Glc based on the observation that it binds in the same position and orientation as the

substrate in the abortive ternary complex of UGDH (PDB entry 2Q3E). The occupancy of the UDP-Glc was adjusted to 50% to match the *B* factors of the interacting residues in the active site (see Methods). In the other five chains of $hUDGH_{A136M}$, the corresponding density is weaker and supports only our modeling of the diphosphate in UDP-Glc (not shown). The quality of the nucleotide sugar electron density appears to correlate with the average *B* factors of the chains; chain C has the lowest average *B* factor (44 Å2) and the best ordered UDP-Glc, while the average B factors for the remaining chains range from 51 to 60 Å2 with much weaker density for the nucleotide sugar. This is not the first observation of a nucleotide sugar copurifying with UGDH. Dickinson reported that bovine UGDH purified from liver contained a nucleotide sugar that he tentatively identified as UDP-Glc or UDP-Xyl.[23] We chose to model UDP-Glc based on the fact that *E. coli* does not produce UDP-Xyl.

The A136M substitution is well-ordered and fills the A136 surface pocket in the hexamer-building interface without making any van der Waals contacts with the adjacent monomer in the hexamer (Figure B4.A,B). As expected, M136 adopts the trans rotamer with the following angles: $\chi1 = 179.3°$, $\chi2 = 73.4°$, and $\chi3 = -111.0°$. The A136M substitution does not interfere with the lattice contacts in either of the two UDP-Xyl-bound E^{Ω} crystal forms we have previously described.[17] Still, our attempts to obtain X-ray diffraction quality crystals of $hUDGH_{A136M}$ using both previously published conditions failed. The fact that $hUDGH_{A136M}$ will not crystallize in the E^{Ω} state suggests that the substitution prevents the allosteric transition.

As in previous studies,[16] we analyzed the conformation of the allosteric switch by superimposing the Cα atoms of residues 1–124 in $hUDGH_{A136M}$ onto the active E, E*, and UDP-Xyl-inhibited E^{Ω} structures of hUGDH [PDB entries 2Q3E, 4QEJ, and 3PTZ, respectively (Figure B4.C)]. The T131-loops in all six monomers of $hUDGH_{A136M}$ are in a conformation between the

active E and E* states, with the T131 Cα atom of $hUGDH_{A136M}$ being displaced 0.5–0.7 and 0.4–0.6 Å from the corresponding atoms in the E and E* structures, respectively. In contrast, the α6 helix and the hexamer building interface of $hUGDH_{A136M}$ are most similar to the E conformation of hUGDH (PDB entry 2Q3E) (Figure B4.D). Specifically, F323 in the hexamer-building interface adopts the same rotamer observed in the active state, with the Cα atoms of F323 in all chains of $hUDGH_{A136M}$ positioned 0.4–0.8 Å from the E state compared to being 1.5–1.9 Å from the E* state.

The A136M Substitution Stabilizes the Hexamer

We have previously shown that the ligand-free hUGDH hexamer is relatively unstable in solution and dissociates into a concentration-dependent distribution of dimers, tetramers, and hexamers in rapid equilibrium.[19] Because of its location in the hexamer-building interface, it is possible that the A136M substitution could alter the stability of the hexamer. We used sedimentation velocity studies to analyze the oligomeric structure of $hUGDH_{A136M}$ in solution. The *c(s)* distribution of 9 μM $hUGDH_{A136M}$ at pH 7.5 is dominated (94.1%) by an 11.5 S species corresponding to the hexamer, and a smaller (5.9%) amount of the 5.2 S dimer (Figure B5.A). For comparison, the c(s) distribution of 9 μM hUGDH reveals an 11.4 S hexamer (78.2%), an 8.3 S tetramer (4.2%), and a 5.5 S dimer (13.1%) (Figure B5.B). The distribution also contains a 3.2 S species that is most likely a small amount (4.5%) of misfolded monomer. The slight differences in S values for corresponding species in the hUDGH and $hUGDH_{A136M}$ distributions are not significant as S values obtained from a system in rapid equilibrium are inherently biased by the mean of the *c(s)* distribution.[32, 37]

Next, we examined the oligomeric state of $hUGDH_{A136M}$ at the more alkaline pH of 8.5, which we have previously shown to weaken the hUGDH hexamer.[16] Again, the main sedimenting

species in the hUGDH$_{A136M}$ *c(s)* distribution is an 11.85 S hexamer (83.7%), followed by a 5.49 S dimer (13.4%) and a 7.9 S tetramer (3.0%) (Figure B5.C). In contrast, the *c(s)* distribution of hUGDH reveals significantly less 10.6 S hexamer (64.4%) and a complementary increase in the 6.5 S dimer concentration (34.0%) (Figure B5.D). The broadening of the hexamer and dimer peaks in hUGDH is a characteristic of the rapid equilibrium between species.[16, 19, 38] This likely explains why we do not observe a quantifiable peak for the tetramer, which represents a transient in the formation and dissociation of the hexamer.[16] These results show that the A136M substitution favors the hexameric state of hUGDH.

Hysteresis Is Not Observed in hUGDH$_{A136M}$

Hysteresis in hUGDH and hUDGH$_{A136M}$ was analyzed using stopped-flow absorbance spectroscopy. hUGDH was injected into a rapid mixing cell with saturating substrate and NAD^+, and the characteristic lag in enzyme activity was measured by fitting the data to eq 1 as described in Methods (Table B2). The relaxation of hUGDH has a τ of 16.30 ± 0.24 s, which corresponds to a lag of 44.3 ± 0.66 s before reaching the steady state velocity of 1.0 ± 0.01 nM NADH s^{-1} (Figure B6.A). In contrast, we observed no lag for hUDGH$_{A136M}$ under these conditions (Figure B6.B). In fact, hUGDH$_{A136M}$ does not display hysteresis even at saturating concentrations of the feedback inhibitor UDP-Xyl (Table B2). This is not true of hUGDH, where the lag has been shown[16] to increase with UDP-Xyl concentration (Table B2). Finally, the steady state velocity of hUDGH$_{A136M}$ appears to be 30% faster than that of hUGDH, but this is most likely due to negative cooperativity in the latter enzyme (discussed below) (Table B2).

The A136M Substitution Disrupts NAD^+-Induced Negative Cooperativity

The NAD^+ saturation curve of hUGDH displays negative cooperativity (Hill coefficient of 0.73 ± 0.07), which we have previously shown[22] to be an intrinsic property of the substrate-bound

complex (in the absence of UDP-Glc, the binding of NAD^+ is noncooperative) (Figure B6.C and Table B3). In contrast, the NAD+ saturation curve of $hUGDH_{A136M}$ is hyperbolic (Figure B6.D and Table B3). Negative cooperativity indicates an asymmetry in an enzyme that results in a mixture of high-affinity and low-affinity binding sites. It is possible to estimate the K_M for high-affinity and low-affinity NAD^+ binding (K_0 and K_{lim}, respectively) using Kurgonov's methodology[35] (see Methods) (Figure B6.E). This analysis of hUGDH NAD^+ saturation curves yields a K_0 of 88 ± 20 μM, which is similar to the K_M (90 ± 6 μM) observed for $hUGDH_{A136M}$ (Figure B6.D and Table B3).

With respect to turnover, the k_{cat} values for the NAD^+ saturation curves of both hUGDH and $hUGDH_{A136M}$ are comparable at 0.76 ± 0.05 and 0.83 ± 0.02 s^{-1}, respectively (Table B3). The UDP-Glc saturation curves for hUGDH and $hUGDH_{A136M}$ are both hyperbolic, with similar K_M values of 9.7 ± 0.8 and 7.26 ± 0.67 μM, respectively (Table B3) However, there is a significant discrepancy in the k_{cat} values observed in the UDP-Glc saturation curves; while the $hUGDH_{A136M}$ k_{cat} (0.76 ± 0.02 s^{-1}) is similar to that measured in the NAD^+ saturation curves, the k_{cat} for hUGDH is significantly lower (0.55 ± 0.01 s^{-1}) (Table B3). The difference in turnover numbers is due to the coenzyme-dependent negative cooperativity of hUGDH, which makes it difficult to saturate the enzyme with NAD^+ (K_{lim} = 1800 μM) to achieve the pseudo-first-order conditions necessary for steady state analysis of UDP-Glc kinetics.

$hUDGH_{A136M}$ Binding of UDP-Xylose Is Noncooperative

Next we examined the impact of the A136M substitution on the allosteric inhibition mechanism of hUGDH. UDP-Xyl binding stabilizes the E^{Ω} conformation of hUGDH (Figure B1).[16-18] Competition with UDP-Glc induces the low-substrate-affinity E^{Ω} conformation of hUGDH to undergo a cooperative conformational change to the high-affinity E state.[16, 39] The

allosteric transition can be observed as sigmoidicity in the UDP-Glc saturation curves during steady state analysis of UDP-Xyl inhibition (Figure B6.F). Global analysis of the hUGDH steady state inhibition curves yields a K_i of 0.32 ± 0.05 μM for UDP-Xyl, which is consistent with our earlier work.[16] The UDP-Glc saturation curve in the presence of UDP-Xyl at a concentration that is ~10 times the K_i is strongly cooperative, with a Hill coefficient of 2.1 ± 0.2 (Figure B6.F and Table B4). This agrees with our previous work, which showed that sigmoidicity increases with UDP-Xyl concentration.[16] In contrast, $hUGDH_{A136M}$ has a much lower affinity for UDP-Xyl (Ki of 4.2 ± 0.4 μM) and displays no cooperativity (Figure B6.G and Table B4). In fact, the UDP-Glc saturation curves are hyperbolic even at inhibitor concentrations 14 times the K_i (60 μM UDP-Xyl), which is strong evidence that $hUGDH_{A136M}$ does not undergo a conformational change between inhibited and uninhibited states (Figure B6.F).

B5. Discussion

We have previously shown that binding of the feedback inhibitor UDP-Xyl to hUGDH induces the T131-loop/α6 allosteric switch to repack and form the inactive E^{Ω} state (Figure B1).[17] This remarkable allosteric transition requires the α6 helix to rotate, which buries A136 in the protein core (Figure B2.A,B). Here we have tested the allosteric mechanism by trapping the α6 helix in the active conformation using the A136M substitution (Figure B4). The burial of the methionine would be unlikely, as it would require a significant local unfolding of the protein and repacking of the core to accommodate the bulky side chain (Figure B2). This rationale is supported by the substrate saturation kinetics of the UDP-Xyl-inhibited enzyme (Table B4). The E^{Ω} state has a lower affinity for UDP-Glc, which results in positive cooperativity in substrate saturation curves of hUGDH (Figure B6.F).[16] The fact that UDP-Xyl-saturated $hUGDH_{A136M}$ does not display substrate cooperativity is strong evidence that the A136M substitution prevents the formation of

the E^{Ω} state (Figure B6.F and Table B4). This interpretation implies that the lower affinity for UDP-Xyl (K_i = 4.2 μM) observed in $hUGDH_{A136M}$ corresponds to the competitive inhibition constant for the E state only (Figure B6.G and Table B4). This is the first report of the E state affinity constant for UDP-Xyl and suggests the higher affinity (K_i = 0.32 μM) observed for hUGDH is a property of the E^{Ω} state (Table B4).

The specific activity of hUGDH shows a hyperbolic dependency on protein concentration that can be modeled as three low-activity dimers associating to form a higher-activity hexamer.[16] It is possible for enzyme association to result in hysteresis,[40, 41] but this is not the case for hUGDH. Instead, hysteresis in hUGDH is caused by the slow isomerization from the inactive E* state to the active E state upon binding of NAD^+ and substrate.[19, 22] While the increased stability of the $hUGDH_{A136M}$ hexamer does not explain the absence of a lag in progress curves, it does focus attention on the conformational flexibility of the allosteric switch in the E* state (Figures B5 and B6.A). The intermediate conformation T131-loop/α6 allosteric switch and the hexamer-building interface in E* led us to hypothesize that the hysteresis was the result of NAD^+ and substrate inducing the allosteric switch and surrounding residues to slowly repack into the E conformation[16, 18] (Figure B1.B,C). The A136M substitution allows us to test this model. The crystal structure of unliganded $hUGDH_{A136M}$ shows that both the allosteric switch and the hexamer-building interface favor the E state, despite the open domain conformation of the enzyme (Figure B4.C,D). By stabilizing the allosteric switch in the E state, the A136M substitution would prevent repacking of the core and interface. This model accurately predicts the absence of a lag in the $hUGDH_{A136M}$ progress curves (Figure B6.B). Additional evidence comes from UDP-Xyl inhibition studies. We have previously shown[16] that the lag in hUGDH increases with UDP-Xyl concentration (Table B2). This increase is most likely caused by the inhibitor stabilizing the E^{Ω} conformation, which

must first convert to E*Bound before it can slowly isomerize into the active E state (Figure B1.D). Thus, the absence of hysteresis in UDP-Xyl-saturated $hUGDH_{A136M}$ is strong evidence that the A136M substitution prevents the formation of both E* and E^{Ω} states (Table B2). Together, these results identify the isomerization of the allosteric switch as the source of the hysteresis observed in hUGDH.

Negative cooperativity can arise from enzyme heterogeneity that results in a mixed population with different affinity constants, but this model is not supported by ligand binding studies of hUGDH.[22] Briefly, the binding of UDP-Glc or NAD^+ to hUGDH is noncooperative, indicating that the enzyme is homogeneous, with single affinity constants for substrate and coenzyme.[22] However, the binding of NAD^+ to the binary hUGDH–UDP-Glc complex displays negativity cooperativity.[22] The simplest interpretation of these results is that the binding of UDP-Glc induces an asymmetry in the enzyme that produces two distinct affinities for the coenzyme. The molecular basis of negative cooperativity in hUGDH is not known, but the fact that the T131-loop contributes to both substrate and cofactor binding[17] suggests that it may play a role. This assumption is supported by the observation that $hUGDH_{A136M}$ constrains the T131-loop and does not display negative cooperativity (Figure B6.C,D and Table B3). A comparison of hUGDH and $hUGDH_{A136M}$ suggests that the A136M substitution selects for the high-affinity conformation of the NAD^+ binding sites. Using Kurgonov's approach,[35] we estimated the high- and low-affinity K_M values for negative cooperativity in binding of the coenzyme to hUGDH (Figure B6.E and Table B3). The apparent high affinity K_M (K_0) for hUGDH is very similar to the noncooperative K_M we observe for $hUGDH_{A136M}$ (Table B3). This suggests that the A136M substitution prevents the asymmetry induced by UDP-Glc binding. This suggests that the flexibility of the allosteric switch is linked to the negative cooperativity observed in NAD^+ substrate saturation curves.

This work shows that targeting the allosteric switch is a practical method of selecting specific conformations of hUGDH and, by extension, regulating enzyme activity. Regulating hUGDH is an important milestone in our long-term goal of controlling the pharmacokinetics of chemotherapeutics. Future work will focus on inhibiting hUGDH by trapping the allosteric switch in the E^{Ω} conformation.

B6. Figures and Tables

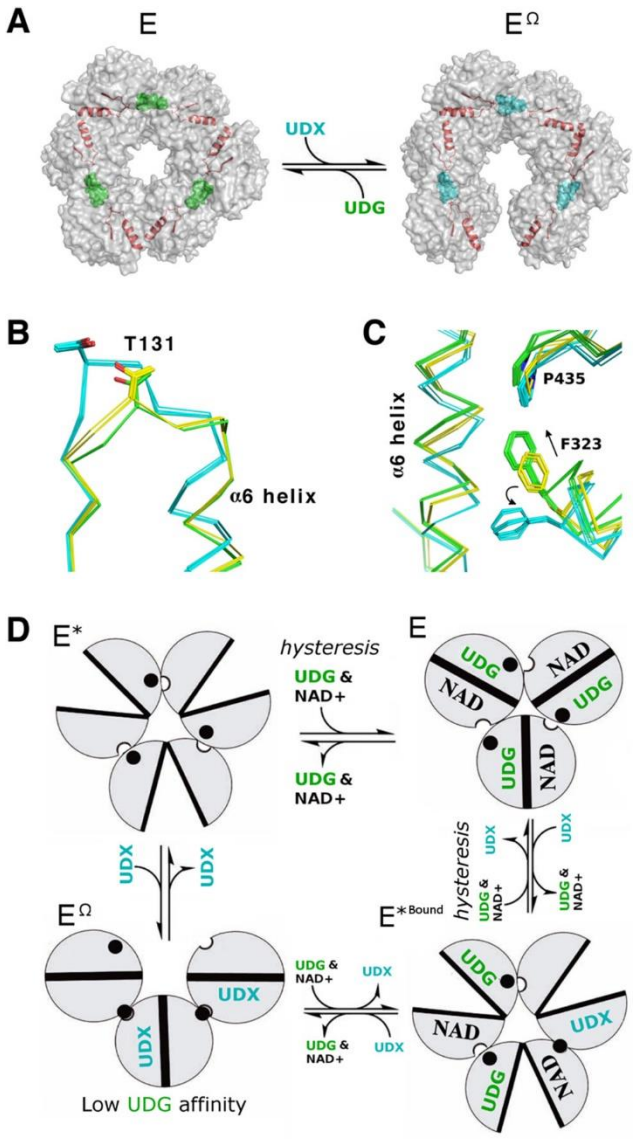


Figure B1: hUDGH undergoes conformational changes during allostery and hysteresis. (A) Semitransparent surface rendering showing the buried T131-loop/α6 helix (red) in the E and E^{Ω} states. UDP-Glc (green spheres, UDG) binds to the active site and favors the 32 symmetry E state. UDP-Xyl (teal spheres, UDX) competes for the active site and triggers the allosteric switch to rotate the α6 helix and favor the E^{Ω} state. **(B)** Position of the T131-loop/α6 helix in the E (green), E* (yellow), and E^{Ω} (cyan) states. All of the unique chains in the crystal structures of E, E*, and E^{Ω} are superimposed (Protein Data Bank entries 2Q3E, 4RJT, and 3PTZ, respectively). **(C)** Hexamer-building interfaces of the E, E*, and E^{Ω} states depicted as in panel B. Panels B and C are adapted from ref 16. **(D)** Relationship between domain rotations (clamshell circles) and the α6 helix (black filled circle) conformation in the E (active), E^{Ω} (inhibited), E* (inactive), and E*Bound (inactive, inhibitor- and substrate-bound) states. In the absence of ligand, the inactive E* state is favored. The binding of the substrate and NAD^+ to E* induces a slow isomerization

(hysteresis) to produce the active E state.[22] The E*Bound state forms as UDG and UDX compete to induce the transition between the E and E^{Ω} states (the ligand stoichiometry that induces the transition is unknown). UDP-Xyl binding shifts the α6 helix to strengthen the packing interactions between adjacent subunits (semicircle cutouts) and form the E^{Ω} hexamer-building interface.[16,17] The E^{Ω} conformation of the allosteric switch reduces the affinity for the substrate. For the sake of simplicity, only one trimer of the hUGDH hexamer is depicted.

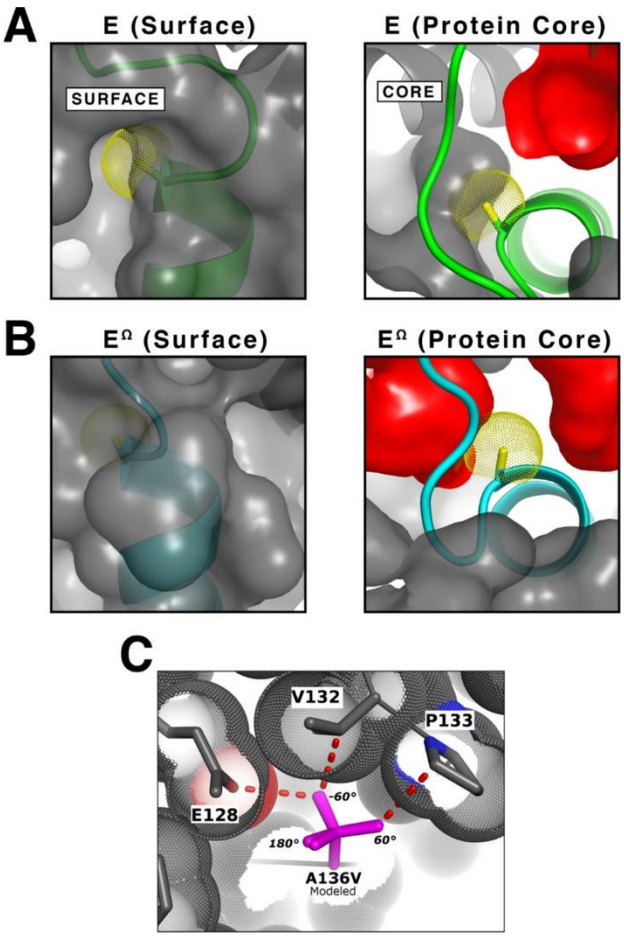


Figure B2: Rationale for the design of the A136M substitution. (A) The left panel is a semitransparent surface rendering of the E state (PDB entry 2Q3E) showing the α6 helix (green) with the Cβ atom of A136 (yellow van der Waals dots) in a solvent accessible pocket. The right panel depicts a "cutaway" of the protein surface looking down the axis of the α6 helix to show conformational changes that take place in the protein core, specifically, the position of the α6 helix and the solvent-exposed A136 in the E state. Cavities buried in the protein core (red surfaces) were identified with a 1.2 Å radius probe. The contribution of the A136 Cβ atom to the surface was removed prior to rendering. **(B)** The E^{Ω} conformation (cyan) of hUGDH rotates the α6 helix to bury A136 in the protein core (PDB entry 3PTZ). The right panel depicts the rotation of the α6 helix to bury A136 in the E^{Ω} state. The left and right panels are illustrated as in panel A. Note that the cavities in the core change shape and number because of the repacking of the allosteric switch. **(C)** To investigate the steric constraints (dotted van der Waals surfaces) of a β-branched amino acid substitution in the A136 pocket, we modeled A136V with its three preferred rotamers superimposed (purple sticks). Each rotamer introduces a bad contact (red dashed lines) between a γ-methyl and adjacent amino acids. Thus, the pocket will accommodate only an unbranched amino acid with a 180° χ1 torsion angle (60°, 180°, and −60° χ1 torsion angles are labeled).

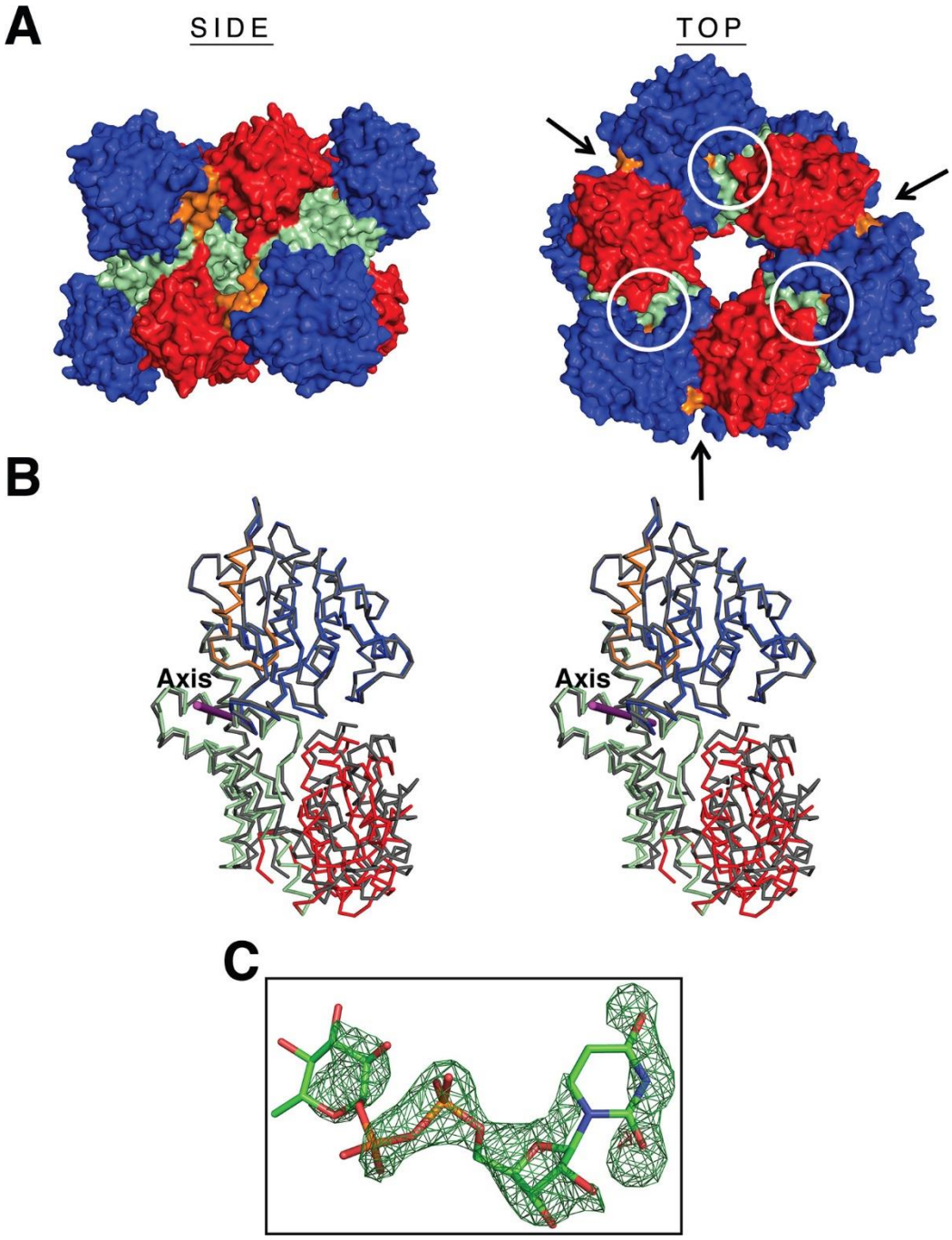


Figure B3: Crystal structure of hUGDH$_{A136M}$ in the open domain conformation. (A) Side and top views of the UDGH$_{A136M}$ 32 symmetry hexamer. The surfaces of the six monomers are colored by domain: NB (blue), dimerization (light green), and SB (red). The allosteric switch is colored orange. The active sites and hexamer-building interfaces are identified with circles and arrows, respectively. **(B)** Stereoview of a Cα trace of hUDGH$_{A136M}$ chain C (colored as in panel A) superimposed onto hUGDH in the closed conformation (gray, PDB entry 2Q3E) to illustrate the open domain conformation. The hinge-bending axis is depicted as a purple rod. **(C)** Difference (F_o – F_c) electron density map contoured at 3σ for the partially occupied UDP-Glc (sticks) calculated after the nucleotide sugar was omitted and the resulting model subjected to simulated annealing.

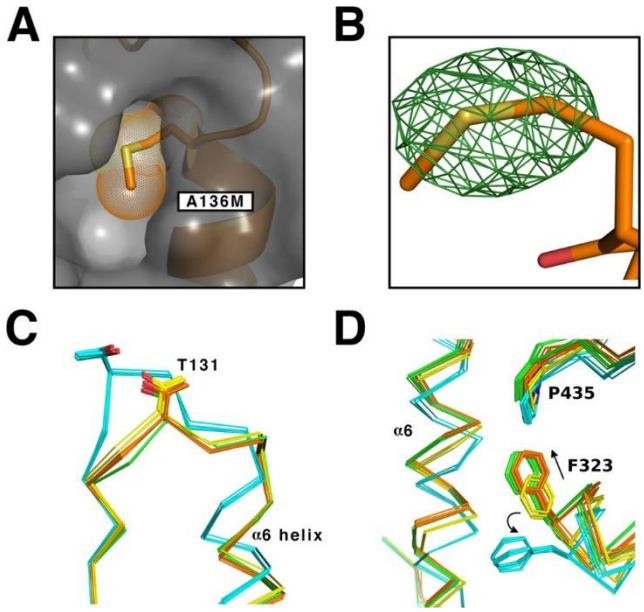


Figure B4: A136M substitution that stabilizes the E state of hUGDH. (A) The difference (F_o – F_c) electron density map contoured at 4σ for A136 calculated prior to modeling the methionine substitution. **(B)** Surface rendering (gray) of $hUDGH_{A136M}$ showing the A136M substitution (orange stick with van der Waals surface dots) and the α6 helix (orange). As expected, the A136M substitution fills the A136 pocket. **(C)** Position of the T131 loop in the inactive (cyan), active (green), E* (yellow), and $hUDGH_{A136M}$ (orange) states. **(D)** Hexamer-building interfaces of the inactive (cyan), active (green), E* (yellow), and $hUDGH_{A136M}$ (orange) states. Key residues are labeled with arrows used to depict the direction of motion between states.

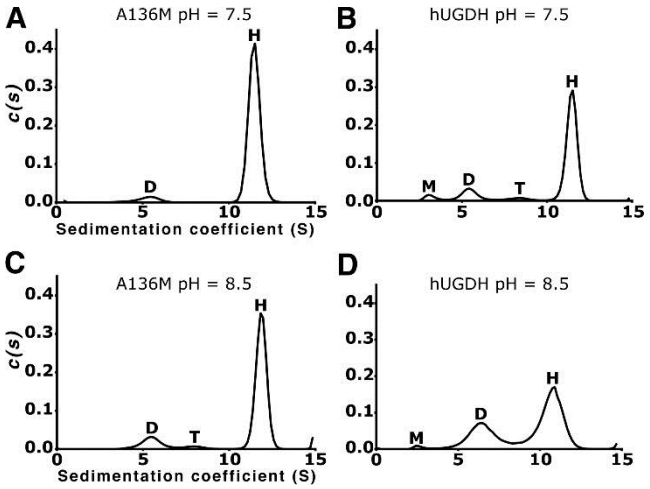


Figure B5: A136M substitution that stabilizes the hexamer. (A) Sedimentation velocity c(s) distribution of hUDGH$_{A136M}$ at pH 7.5 showing the hexamer (H) and dimer (D). **(B)** hUDGH c(s) distribution at pH 7.5 showing the hexamer (H), tetramer (T), dimer (D), and a peak that we interpret as misfolded monomer (M). **(C)** hUDGH$_{A136M}$ at pH 8.5. **(D)** hUDGH at pH 8.5.

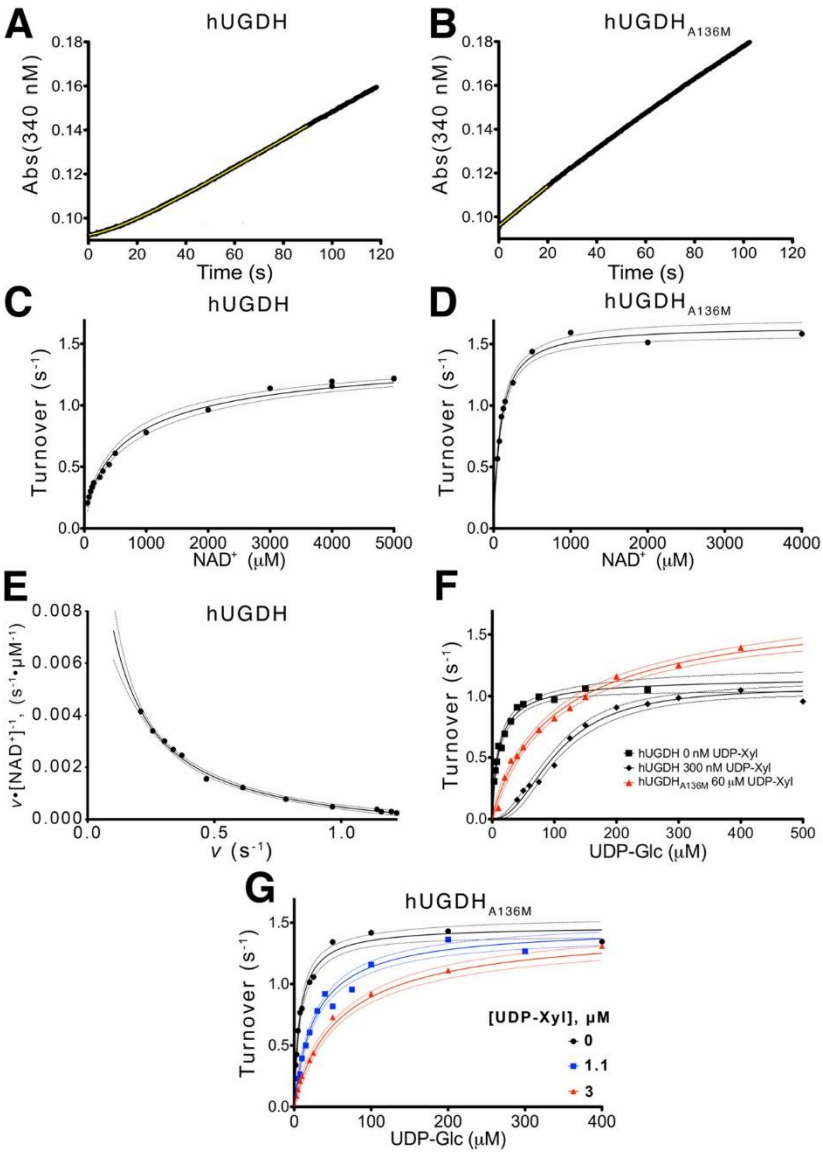


Figure B6: A136M substitution that disrupts hysteresis and cooperativity. (A) Stopped-flow analysis of hysteresis in the hUDGH progress curve under saturating concentrations of NAD^+ (5 mM) and UDP-Glc (0.5 mM). The data points (black line) are fit (yellow line) to eq 1 as described in Methods. **(B)** The $hUDGH_{A136M}$ progress curve was analyzed as described for panel A, shows no lag, and is fit to a linear equation (yellow line) to obtain the initial steady state velocity. The yellow line represents a linear fit for the first 20 s of the reaction. **(C and D)** NAD^+ saturation curves for hUGDH and $hUGDH_{A136M}$, respectively, fit to eq 3. The rates are normalized to turnover (nanomolar NADH per nanomolar enzyme per second). Dashed lines represent the 95% confidence interval of the fit in all panels. **(E)** An Eadie–Hofstee plot of the hUGDH NAD+ saturation data in panel C is concave-up (negative cooperativity) and was fit to eq 3 to determine the K_0 and K_{lim} values in Table 3. Velocity (v) was normalized to turnover (inverse seconds) as in panels C and D. **(F)** hUGDH UDP-glucose saturation curves (black lines) with 0 μM (■) and 3 μM (◆) UDP-xylose were globally fit to eq 4 to determine the K_i and the Hill coefficient. For comparison, data from a $hUGDH_{A136M}$ UDP-glucose saturation curve with 60 μM UDP-Xyl (red triangles) are plotted. **(G)** $hUGDH_{A136M}$ UDP-glucose saturation curves with 0 μM (black circles),

11 μM (blue squares), and 30 μM (red triangles) UDP-xylose were globally fit to eq 4 to determine K_i.

Table B1: Data Collection and Refinement Statistics for hUDGH$_{A136M}$

Data Collection	
Protein Data Bank entry	5TJH
space group	*P*1
unit cell dimensions	
a, *b*, *c* (Å)	92.35, 104.59, 107.71
α, β, γ (deg)	64.82, 68.35, 73.72
completeness (%)	96.6 (97.1)[a] [91.3][b]
redundancy	3.9 (4.0) [3.8]
no. of reflections	987044
$I/\sigma(I)$	7.82 (1.80) [1.03]
$CC_{1/2}$[c]	99.6 (70.3) [45.6]
R_{meas}[d] (%)	11.7 (92.8) [154.1]
Refinement	
resolution (Å)	2.05
R_{work}/R_{free}	0.171/0.203
R_{free} in the highest-resolution shell	0.345
no. of atoms	
protein	21736
ligand	89
water	799

B factor (Å^2)	
protein	53.5
ligand	58.6
water	47.7
Stereochemical Ideality	
bond lengths (Å^2)	0.008
bond angles (deg)	0.89
φ, ψ preferred region (%)	98.3
φ, ψ additionally allowed region (%)	1.7
φ, ψ disallowed region (%)	0.0

a Values in parentheses are for the highest-resolution shell (2.16–2.22 Å) based on an I/σ cutoff of 1.8.
b Values in brackets are for the highest-resolution shell (2.05–2.10 Å) based on the $CC_{1/2}$ cutoff defined by Diederichs and Karplus.[42]
c CC1/2 is the percentage of correlation between intensities from random half-data sets.[42]
d R_{meas} is the redundancy-independent merging *R* factor.[43]

Table B2: Effect of UDP-Xyl on Hysteresis

enzyme	[UDP-Xyl] (μM)	τ (s^{-1})	lag (s)	v_i	v_{ss}
hUGDH	0	16.3 ± 0.24	44.3 ± 0.66	0.43 ± 0.02[a]	1.0 ± 0.01[a]
	3[b]	38.3 ± 2.7	104 ± 7.3	0.35 ± 0.03	1.1 ± 0.01
$hUGDH_{A136M}$	0	ND[c]	ND[c]	ND[c]	1.3 ± 0.03
	60[b]	ND[c]	ND[c]	ND[c]	1.4 ± 0.03

[a] Units of nanomolar NADH per nanomolar enzyme per second.
[b] These concentrations of UDP-Xyl are ≥10 times the K_i listed in Table B4.
[c] Not detected.

Table B3: Steady State Parameters

enzyme	ligand	K_M (μM)	Hill	k_{cat}[a] (s^{-1})
hUGDH	UDP-Glc	9.7 ± 0.8	1	0.55 ± 0.01
	NAD^+	780 ± 180	0.73 ± 0.07	0.76 ± 0.05
$hUGDH_{A136M}$	UDP-Glc	7.26 ± 0.67	1	0.73 ± 0.02
	NAD^+	90 ± 6	1	0.83 ± 0.02

[a] One complete catalytic turnover produces two molecules of NADH.

Table B4: Global Analysis of Competitive Inhibition

enzyme	[UDP-Xyl] (μM)	UDP-Glc K_M (μM)	k_{cat}[a] (s^{-1})	UDP-Xyl K_i (μM)	Hill
hUGDH	0	9.5 ± 1.0[b]	0.55 ± 0.02[b]	0.30 ± 0.03	1
	3				2.1 ± 0.2
$hUGDH_{A136M}$	0	7.34 ± 0.80[b]	0.74 ± 0.02[b]	4.2 ± 0.4	1
	11				1
	30				1
$hUGDH_{A136M}$[c]	60	6.3 ± 0.5	0.85 ± 0.02		1

[a] Turnover produces two molecules of NADH per cycle.
[b] For global analysis, this parameter was refined as a shared value for all UDP-Xyl concentrations.
[c] The 60 μM UDP-Xyl experiment was conducted separately and excluded from the global fit.

B7. References

(1) Guillemette, C. (2003) Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J*. 3, 136−158.

(2) Tukey, R. H., and Strassburg, C. P. (2000) Human UDP-glucuronosyltransferases: Metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol*. 40, 581−616.

(3) Lin, J. H., and Lu, A. Y. H. (1997) Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol. Rev.* 49, 403−449.

(4) Cummings, J., Boyd, G., Ethell, B. T., Macpherson, J. S., Burchell, B., Smyth, J. F., and Jodrell, D. I. (2002) Enhanced clearance of topoisomerase I inhibitors from human colon cancer cells by glucuronidation. *Biochem. Pharmacol*. 63, 607−613.

(5) Cummings, J., Ethell, B. T., Jardine, L., Boyd, G., Macpherson, J. S., Burchell, B., Smyth, J. F., and Jodrell, D. I. (2003) Glucuronidation as a mechanism of intrinsic drug resistance in human colon cancer: Reversal of resistance by food additives. *Cancer Res*. 63, 8443−8450.

(6) Cummings, J., Zelcer, N., Allen, J. D., Yao, D. G., Boyd, G., Maliepaard, M., Friedberg, T. H., Smyth, J. F., and Jodrell, D. I. (2004) Glucuronidation as a mechanism of intrinsic drug resistance in colon cancer cells: contribution of drug transport proteins. *Biochem. Pharmacol*. 67, 31−39.

(7) de Almagro, M. C., Selga, E., Thibaut, R., Porte, C., Noe, V., and Ciudad, C. J. (2011) UDP-glucuronosyltransferase 1A6 overexpression in breast cancer cells resistant to methotrexate. *Biochem. Pharmacol*. 81, 60−70.

(8) Oguri, T., Takahashi, T., Miyazaki, M., Isobe, T., Kohn, N., Mackenzie, P. I., and Fujiwara, Y. (2004) UGT1A10 is responsible for SN-38 glucuronidation and its expression in human lung cancers. *Anticancer Res*. 24, 2893−2896.

(9) Wei, Q., Galbenus, R., Raza, A., Cerny, R. L., and Simpson, M. A. (2009) Androgen-stimulated UDP-glucose dehydrogenase expression limits prostate androgen availability without impacting hyaluronan levels. *Cancer Res*. 69, 2332−2339.

(10) Kultti, A., Pasonen-Seppanen, S., Jauhiainen, M., Rilla, K. J., Karna, R., Pyoria, E., Tammi, R. H., and Tammi, M. I. (2009) 4-Methylumbelliferone inhibits hyaluronan synthesis by depletion of cellular UDP-glucuronic acid and downregulation of hyaluronan synthase 2 and 3. *Exp. Cell Res*. 315, 1914−1923.

(11) Kakizaki, I., Kojima, K., Takagaki, K., Endo, M., Kannagi, R., Ito, M., Maruo, Y., Sato, H., Yasuda, T., Mita, S., Kimata, K., and Itano, N. (2004) A novel mechanism for the inhibition of hyaluronan biosynthesis by 4-methylumbelliferone. *J. Biol. Chem*. 279, 33281−33289.

(12) Balduini, C., Brovelli, A., Deluca, G., Galligani, L., and Castellani, A. A. (1973) Uridine Diphosphate Glucose Dehydrogenase from Cornea and Epiphyseal-Plate Cartilage. *Biochem. J.* 133, 243−249.

(13) Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U., and Nidetzky, B. (2010) UDP-glucose dehydrogenase: structure and function of a potential drug target. *Biochem. Soc. Trans*. 38, 1378−1385.

(14) Gainey, P. A., and Phelps, C. F. (1972) Uridine Diphophate Glucuronic Acid Production and Utilization in Various Tissues Actively Synthesizing Glycosaminoglycans. *Biochem. J.* 128, 215−227.

(15) Maxwell, E. S., Kalckar, H. M., and Strominger, J. L. (1956) Some Properties of Uridine Diphosphoglucose Dehydrogenase. *Arch. Biochem. Biophys*. 65, 2−10.

(16) Kadirvelraj, R., Custer, G. S., Keul, N. D., Sennett, N. C., Sidlo, A. M., Walsh, R. M., and Wood, Z. A. (2014) Hysteresis in Human UDP-Glucose Dehydrogenase Is Due to a Restrained Hexameric Structure That Favors Feedback Inhibition. *Biochemistry* 53, 8043−8051.

(17) Kadirvelraj, R., Sennett, N. C., Polizzi, S. J., Weitzel, S., and Wood, Z. A. (2011) Role of Packing Defects in the Evolution of Allostery and Induced Fit in Human UDP-Glucose Dehydrogenase. *Biochemistry* 50, 5780−5789.

(18) Sennett, N. C., Kadirvelraj, R., and Wood, Z. A. (2011) Conformational Flexibility in the Allosteric Regulation of Human UDP-alpha-D-Glucose 6-Dehydrogenase. *Biochemistry* 50, 9651−9663.

(19) Sennett, N. C., Kadirvelraj, R., and Wood, Z. A. (2012) Cofactor Binding Triggers a Molecular Switch To Allosterically Activate Human UDP-alpha-D-glucose 6-Dehydrogenase. *Biochemistry* 51, 9364−9374.

(20) Helt, S. S., Thymark, M., Harris, P., Aagaard, C., Dietrich, J., Larsen, S., and Willemoes, M. (2008) Mechanism of dTTP inhibition of the bifunctional dCTP deaminase: dUTPase encoded by Mycobacterium tuberculosis. *J. Mol. Biol*. 376, 554−569.

(21) Johansson, E., Thymark, M., Bynck, J. H., Fanø, M., Larsen, S., and Willemoes, M. (2007) Regulation of dCTP deaminase from Escherichia coli by nonallosteric dTTP binding to an inactive form of the enzyme. *FEBS J*. 274, 4188−4198.

(22) Kadirvelraj, R., Sennett, N. C., Custer, G. S., Phillips, R. S., and Wood, Z. A. (2013) Hysteresis and Negative Cooperativity in Human UDP-Glucose Dehydrogenase. *Biochemistry* 52, 1456−1465.

(23) Dickinson, F. M. (1988) Studies on the Unusual Behavior of Bovine Liver UDP-Glucose Dehydrogenase in Assays at Acid and Neutral pH and on the Presence of Tightly Bound Nucleotide Material in Purified Preparations of this Enzyme. *Biochem. J.* 255, 775−780.

(24) Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U., and Nidetzky, B. (2011) Structure and Mechanism of Human UDPglucose 6-Dehydrogenase. *J. Biol. Chem.* 286, 23877−23887.

(25) Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.*, 571−607.

(26) Kabsch, W. (2010) XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 125−132.

(27) Brunger, A. T. (1997) Free R value: Cross-validation in crystallography. In *Macromolecular Crystallography*, Part B (Carter, C.W., and Sweet, R. M., Eds.) pp 366−396, Academic Press, San Diego.

(28) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., GrosseKunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 213−221.

(29) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 486−501.

(30) Urzhumtsev, A., Afonine, P. V., and Adams, P. D. (2013) TLS from fundamentals to practice. *Crystallogr. Rev.* 19, 230−270.

(31) Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Analytical Ultracentrifuge in Biochemistry and Polymer Science. *R. Soc. Chem.*, 90−125.

(32) Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem*. 320, 104−124.

(33) Frieden, C. (1970) Kinetic Aspects of Regulation of Metabolic Processes- Hysteretic Enzyme Concept. *J. Biol. Chem*. 245, 5788.

(34) Cornish-Bowden, A. (2001) Detection of errors of interpretation in experiments in enzyme kinetics. *Methods 24*, 181−190.

(35) Kurganov, B. I. (2000) Analysis of negative cooperativity for glutamate dehydrogenase. *Biophys. Chem*. 87, 185−199.

(36) Hayward, S., and Berendsen, H. J. C. (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins: Struct., Funct., Genet*. 30, 144−154.

(37) Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J*. 78, 1606−1619.

(38) Zhao, H., Balbo, A., Brown, P. H., and Schuck, P. (2011) The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity. *Methods 54*, 16−30.

(39) Kuo, L. C. (1983) Allosteric Cofactor-Mediated Enzyme Cooperativity- A Theoretical Treatment. *Proc. Natl. Acad. Sci. U. S. A. 80*, 5243−5247.

(40) Kurganov, B. I., Dorozhko, A. K., Kagan, Z. S., and Yakovlev, V. A. (1976) The theoretical analysis of kinetic behaviour of kinetic behaviour of ″hysteretic″ allosteric enzymes. III. Dissociating and associating enzyme systems in which the rate of installation of equilibrium between the oligomeric forms in comparable to that of enzymatic reaction. *J. Theor. Biol. 60*, 287−299.

(41) Kurganov, B. I., Dorozhko, A. I., Kagan, Z. S., and Yakovlev, V. A. (1976) The theoretical analysis of kinetic behaviour of ″hysteretic″ allosteric enzymes. II. The dissociating and associating enzymic systems in which the rate of installation of equilibrium between the oligomeric forms is small in comparison with that of enzymatic reaction. *J. Theor. Biol. 60*, 271−286.

(42) Karplus, P. A., and Diederichs, K. (2012) Linking Crystallographic Model and Data Quality. *Science 336*, 1030−1033.

(43) Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol. 4*, 269−275.

APPENDIX C

HYSTERESIS AND ALLOSTERY IN HUMAN UDP-GLUCOSE DEHYDROGENASE REQUIRE A FLEXIBLE PROTEIN CORE

Keul, N.D. conducted and analyzed sedimentation velocity experiments. In addition, Keul, N.D. contributed in manuscript revisions.

C1. Abstract

Human UDP-glucose dehydrogenase (hUGDH) oxidizes UDP-glucose to UDP-glucuronic acid, an essential substrate in the phase II metabolism of drugs. The activity of hUGDH is regulated by the conformation of a buried allosteric switch (T131 loop/α6 helix). Substrate binding induces the allosteric switch to slowly isomerize from an inactive E* conformation to the active E state, which can be observed as enzyme hysteresis. When the feedback inhibitor UDP-xylose binds, the allosteric switch and surrounding residues in the protein core repack, converting the hexamer into an inactive, horseshoe-shaped complex (E^{Ω}). This allosteric transition is facilitated by large cavities and declivities in the protein core that provide the space required to accommodate the alternate packing arrangements. Here, we have used the A104L substitution to fill a cavity in the E state and sterically prevent repacking of the core into the E^{Ω} state. Steady state analysis shows that $hUGDH_{A104L}$ binds UDP-xylose with lower affinity and that the inhibition is no longer cooperative. This means that the allosteric transition to the high-UDP-xylose affinity E^{Ω} state is blocked by the substitution. The crystal structures of $hUGDH_{A104L}$ show that the allosteric switch still adopts the E and E* states, albeit with a more rigid protein core. However, the progress curves of $hUGDH_{A104L}$ do not show hysteresis, which suggests that the E* and E states are now in rapid equilibrium. Our data suggest that hysteresis in native hUGDH originates from the conformational entropy of the E* state protein core.

C2. Introduction

UDP-glucuronic acid (UDP-GlcA) is the essential substrate in glucuronidation, a major component of phase II metabolism in mammalian cells.[1–3] Glucuronosyltransferases catalyze the nucleophilic addition of glucuronic acid to drugs or toxins, which are then rapidly excreted from the body.[4] Glucuronidation is exploited by some cancers as a mechanism of chemotherapeutic resistance.[5–8] Thus, controlling glucuronidation is a potential strategy for combating drug resistance in these cancers. Because there are 28 glucuronosyltransferase isozymes, each with broad substrate specificity, designing specific inhibitors is challenging.[4] Alternatively, it is possible to inhibit glucuronidation by reducing the availability of the essential substrate UDP-GlcA.[9–11] UDP-GlcA is produced by UDP-glucose dehydrogenase (UGDH), which uses two molecules of NAD^+ to oxidize the C6 hydroxyl of UDP-glucose (UDP-Glc).[12] The downstream metabolite UDP-xylose (UDP-Xyl) acts as an allosteric feedback inhibitor to control the activity of hUGDH.[13,14] Understanding the allosteric mechanism of UGDH is an important goal in developing strategies to control glucuronidation.

Human UGDH (hUGDH) forms a hexamer, with each chain containing an NAD^+ binding (NB), dimerization (DD), and a nucleotide–sugar binding (SB) domain (Figure C1.A). The activity of hUGDH is controlled by the feedback inhibitor UDP-Xyl, which converts the enzyme into an inactive horseshoe-shaped hexamer (E^{Ω}) (Figure C1.A).[14,15] This conformational change is controlled by a buried allosteric switch that links the active site to the α6 helix in the hexamer-building interface (Figure C1.B,C).[15,16] The allosteric switch (T131 loop/α6 helix) can adopt one of three distinct conformations: the active E state, the allosterically inhibited E^{Ω}, and an inactive E* conformation that represents a structural intermediate between the E and E^{Ω} states (Figure C1.C).[15–19] In the absence of a ligand, the enzyme favors the E* state, which resembles the E state

with the exception that the NAD^+ binding domains are rotated into an "open" conformation that exposes the T131 loop and surrounding residues to the solvent (Figure C1.B).[18,20] When UDP-Glc binds, the allosteric switch and surrounding residues in the protein core adopt the active E conformation (Figure C1.B,C). The activity of hUGDH is regulated by an atypical allosteric mechanism. Briefly, UDP-Xyl competes with the substrate for the active site and upon binding induces the allosteric switch to adopt the E^{Ω} conformation (Figure C1.A,C).[15,18] This remarkable transition occurs because UDP-Xyl lacks the C5 hydroxymethyl of UDP-Glc, which allows the allosteric switch to move ~5 Å into the active site (Figure C1.B).[15] What makes this allosteric is that the binding of UDP-Xyl in the active site increases the affinity between subunits in the E^{Ω} hexamer.[21] In the inhibited state, the T131 loop occludes the coenzyme binding site and reduces the affinity for UDP-Glc (Figure C1.B).[15,18] Because the E^{Ω} and E states have different affinities for UDP-Glc, the allosteric transition can be observed as positive cooperativity in steady state studies of UDP-Xyl inhibition.[15,18]

The substrate-induced transition from the inactive E* to the active E hexamer is slow and can be observed as hysteresis in progress curves.[18,19,22] It has been proposed that the source of this slow enzyme isomerization is the repacking of the allosteric switch and surrounding residues into the E state.[18,19,23] This hypothesis is supported by a previous study that showed trapping the allosteric switch in the active conformation abolishes hysteresis.[23] The remarkable plasticity exhibited by the hUGDH protein core likely originates from large cavities and deep surface pockets (declivities) near the allosteric switch (Figure C1.D).[15] Here, we have examined how the flexibility of the hUGDH protein core contributes to both hysteresis and allostery. Using a small to large amino acid substitution (A104L), we have partially filled a cavity to stabilize the E state. Our results show that these packing defects are potential targets for regulating hUGDH activity.

C3. Materials and Methods

Cavity Characterization

Caver Analysis 1.0[24] was used to characterize the packing defects observed in crystal structures of hUGDH (UniProtKB entry O60701). Protein Data Bank (PDB) entries 2Q3E, 4RJT, and 3PTZ were used as the models for the E, E*, and E^{Ω} states, respectively. Each chain of the structures was converted to an individual PDB file prior to analysis, which includes 2Q3E chains A–F, 4RJT chains A–C, and 3PTZ chains A–F. The volumes of cavities were quantified using an inner probe radius of 1.2 Å and an outer probe radius of 1.9 Å. The inner probe radius was selected to approximate the radius of water. The outer probe radius was selected to maximize the volume while preventing the cavity from opening to the surface. Declivities were defined using an inner probe radius of 1.2 Å and an outer probe radius of 2.3 Å. For the chains of 3PTZ, the outer probe radius was varied from 1.9 to 2.3 Å to minimize extraneous surface inclusion in the declivity volume calculation. The volumes were calculated after maximizing tessellation by using the "increase precision" option that comes as part of Caver Analysis 1.0. Because the size and locations of the packing defects are nearly identical in all chains, the contributing atoms of each packing defect were identified using chain "A" of each hUGDH structure.

RosettaVIP Redesign

Chain A of hUGDH in the E state (PDB entry 2Q3E) was used as the template for RosettaVIP[25] core redesign. Because the initial relaxation step in the RosettaVIP protocol introduced small shifts that decreased the size of the packing defects observed in the native crystal structures, we did not include it in the redesign effort. Packing defects were detected using a probe with a radius of 1.2 Å. See results for residues targeted in the redesign. The protein structure was subject to iterative rounds of substitution, fixed backbone relaxation, and then scoring.

Protein Expression, Crystallization, and Structure Solution

Both hUGDH and $hUGDH_{A104L}$ were recombinantly expressed in Escherichia coli and purified as previously described.[15,18−21,23] Following purification, TEV protease was used to remove the His tag, and another round of purification was conducted to remove the TEV. Proteins were quantified using an Agilent 8453 UV/vis instrument and an ε280 of 49850 M^{-1} cm^{-1} for hUGDH and $hUGDH_{A104L}$ (calculated using Protparam[26]). The proteins were dialyzed into a storage buffer [25 mM Tris (pH 8.0) and 50 mM NaCl] and concentrated to ∼20 mg/mL using a Millipore Amicon Ultra-15 10K centrifugal filter unit. Aliquots of protein were flash-frozen in liquid nitrogen and stored at −80 °C. All proteins were crystallized at 20 °C using the hanging drop vapor diffusion method, with a 2 μL drop consisting of a 1:1 mixture of protein (10 mg/mL) and the reservoir solution. The E complex of $hUGDH_{A104L}$ ($hUGDH_{A104L}$:UDP-Glc:NADH) was formed with 3 mM UDP-Glc and 5 mM NADH in the protein solution over a reservoir of 16% polyethylene glycol 3350, 0.2 M ammonium sulfate, 0.3 M NaCl, 1% 1,2-propanediol, and 0.1 M Tris (pH 8.6). Unliganded $hUGDH_{A104L}$ (E*) was crystallized at 20 °C over a reservoir containing 16% polyethylene glycol 3350, 0.5 M NaCl, and 0.1 M Tris (pH 7.6). Crystals were cryoprotected using the reservoir conditions supplemented with 15% EDG (1:1:1 ethylene glycol:dimethyl sulfoxide:glycerol ratio) and then rapidly plunged into liquid nitrogen. All X-ray diffraction data were collected using the SER-CAT beamlines at the Advanced Photon Source (Argonne, IL). For the E complex ($hUGDH_{A104L}$:UDP-Glc:NADH), a 2.5 Å resolution data set was collected on the 22-ID beamline using a Rayonix MX300HS CCD detector. For the E* structure, a 2.65 Å resolution data set was collected on the 22-BM beamline using a MAR MX300 CCD detector. XDS[27] was used for data processing, and 5% of the data was set aside for cross validation.[28] The structure of the E complex ($hUGDH_{A104L}$:UDP-Glc:NADH) was determined by molecular

replacement using PDB entry 2Q3E as a search model in Phaser-MR of the PHENIX[29] software package. Because the hUGDH$_{A104L}$ E* crystal was isomorphous to the native E* structure (PDB entry 4RJT), the native R_{free} flags were preserved. Both structures were subjected to iterative rounds of manual rebuilding in COOT[30] followed by automated refinement in PHENIX.[29] The data collection and refinement statistics are listed in Table C1.

Normalized B-Factor Analysis

The *B*-factors of the isomorphous hUGDH$_{A104L}$ and native hUGDH (PDB entry 4RJT) E* crystal structures were set to 20 and then refined until convergence (approximately seven cycles) without TLS[33] or NCS restraints using PHENIX.[29] Because the NAD domain (residues 1–212), dimerization domain (residues 213–322), and sugar binding domain (residues 323–466) of each chain behave as separate rigid bodies in the crystal, the *B*-factor of each Cα atom was normalized relative to the other Cα atoms within the same domain using the equation[34]:

$$B_{normalized} = (B - B_{mean})/B_{stdev} \qquad \text{(equation 1)}$$

where B is the *B*-factor of the specific Cα atom, B_{mean} is the average *B*-factor of the domain in question, and B_{stdev} is the standard deviation of the Cα atom *B*-factors within the domain. Next, the normalized *B*-factors for each domain in a crystal were averaged. Because the crystals are isomorphous, the differences in the average normalized *B*-factors of hUGDH$_{A104L}$ and native hUGDH will reflect only changes in flexibility due to the A104L substitution.

Sedimentation Velocity

hUGDH and hUGDH$_{A104L}$ were dialyzed for 24 h into 25 mM HEPES (pH 7.5) and 150 mM KCl and then quantified as described above. The samples were then diluted with the dialysis buffer to a final protein concentration of 9 μM, loaded into 12 mm double-sector Epon centerpieces, and equilibrated for 1 h at 20 °C. Sedimentation velocity data were collected at 50000

rpm and 20 °C in an Optima XLA analytical ultracentrifuge. Data were recorded at 280 nm in radial step sizes of 0.003 cm. SEDNTERP[35] was used to estimate the partial specific volumes of hUGDH (0.73840 mL/g) and $hUGDH_{A104L}$ (0.73872 mL/g) and the density (1.00726 g/mL) and viscosity (0.01018) of the buffer. SEDFIT[36] was used to analyze the raw sedimentation data. Data were modeled as a continuous sedimentation coefficient distribution [*c(s)*] and were fit using the baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Fit data for both experiments had root-mean-square deviations (rmsd's) of <0.004 AU.

Progress Curve Analysis of Hysteresis

UDP-glucose and NAD^+ were purchased from Sigma, and UDP-xylose was purchased from Carbosource (Complex Carbohydrate Research Center, University of Georgia, Athens, GA). Reactions were monitored at 25 °C using an Olis RSM 1000 spectrophotometer equipped with a stopped-flow assembly and a 0.4 mm path length. Separate enzyme (460 nM) and substrate (1 mM UDP-glucose and 10 or 5 mM NAD^+) solutions were made in reaction buffer [50 mM HEPES (pH 7.5), 50 mM NaCl, and 5 mM EDTA] and equilibrated at 25 °C for 5 min. We used the 5 mM NAD^+ concentration for $hUGDH_{A104L}$ because the A104L substitution increased the apparent affinity for the coenzyme (Table C3). Equal volumes of enzyme and substrate solutions were then rapidly mixed using the stopped flow. Absorbance readings at 340 nm were taken every 0.1 s for a total of 200 s for native hUGDH and 90 s for $hUGDH_{A104L}$ to monitor NADH production. Native hUGDH progress curves were fit to Frieden's equation for enzyme hysteresis:[37]

$$P(t) = v_{ss}t - \tau(v_{ss} - v_i)(1 - e^{-t/\tau}) \qquad \text{(equation 2)}$$

where P is the concentration of the product at time t and τ is equal to $1/k_{obs}$, where k_{obs} is the apparent rate constant for the transition between the initial rate (v_i) and the steady state rate (v_{ss}). For hysteretic enzymes, the initial velocity defined by eq 2 is the pre-steady state rate and does not

follow the steady state approximation. v_{ss} represents the initial steady state velocity and is calculated before 10% of the substrate is converted to product. The length of the lag is calculated as eτ. The $hUGDH_{A104L}$ progress curves did not display hysteresis, and the initial steady state velocity was derived from the linear portion of the progress curves prior to depleting 10% of the substrate. Data were fit using PRISM (GraphPad Software Inc., San Diego, CA), and all fits were analyzed using residual plots.[38]

Steady State Kinetics

The assay conditions were as previously described,[15,18–21,23] with the exception that the enzyme dilution buffer was changed to 50 mM HEPES (pH 7.0), 50 mM NaCl, and 5 mM EDTA. Briefly, for NAD^+ saturation curves, the assays contained 100 nM enzyme in 50 mM HEPES (pH 7.5), 50 mM NaCl, and 5 mM EDTA with a saturating concentration of 1 mM UDP-glucose. The same conditions were used for UDP-Glc saturation curves, except that the NAD^+ concentration was fixed at 5 mM. Because hUGDH displays negative cooperativity with respect to NAD^+ binding, it is difficult to saturate the low-affinity sites with the coenzyme (see Results).[19] $hUGDH_{A104L}$ does not exhibit negative cooperativity with respect to NAD^+ binding and is saturated with 2.5 mM coenzyme. Prior to being mixed, the enzyme and substrate were incubated separately at 25 °C for 5 min. The reactions were initiated by rapidly mixing enzyme and substrate solutions, and the progress curves were measured by continuously monitoring the change in 340 nm absorbance every 0.5 s using an Agilent 8453 UV/vis spectrometer. The NADH concentration was determined at 340 nm using a molar absorptivity coefficient of 6220 M^{-1} cm^{-1}. Because native hUGDH is hysteretic, the steady state velocities were calculated by fitting progress curves to eq 2 as described above.[15,18–21,23] $hUGDH_{A104L}$ does not display hysteresis; the steady state velocity was determined by fitting the initial liner portion of the progress curves to a straight line. Initial steady

state velocities from progress curves were fit to a sigmoidal rate equation and analyzed with residual plots:[38]

$$v = \frac{k_{cat}[E_t][S]^h}{K_M{}^h + [S]^h} \qquad \text{(equation 3)}$$

where *h* is the Hill coefficient. The negative cooperativity observed in hUGDH NAD^+ substrate saturation curves was evaluated using Kurganov's analysis as described previously.[23,39] Briefly, the data were fit to the equation:

$$\frac{v}{[S]} = \frac{V_{max} - v}{K_M^{eff}}, \text{ where } K_M^{eff} = K_0 + (K_{lim} - K_0)(v/V_{max}) \qquad \text{(equation 4)}$$

where K_0 is the approximate K_M for the high-affinity binding sites and K_{lim} is the average apparent K_M for the low-affinity sites.[39]

Inhibition Kinetics

The K_i for the inhibitor UDP-Xyl was determined as previously described.[15,18–21,23] Briefly, UDP-Xyl competes with UDP-Glc for the active site and induces a conformational change in hUGDH to the inhibited E^Ω state. Because the E and E^Ω states have different affinities for UDP-Glc and UDP-Xyl, the inhibition studies show cooperativity.[18] To determine the K_i, the substrate saturation curves at different concentrations of UDP-Xyl were fit simultaneously to an equation for competitive inhibition corrected for cooperativity using global analysis in Prism:

$$v_o = \frac{k_{cat}[E_t][S]^h}{(K_M^{app})^h + [S]^h}, \text{ where } K_M^{app} = K_M\left(1 + \frac{[I]}{K_i}\right) \qquad \text{(equation 5)}$$

C4. Results

Rational Design of the A104L Cavity Filling Substitution

The cavities and declivities in the E state protein core are believed to provide space and accommodate the repacking of the allosteric switch and surrounding residues into the E^Ω conformation (Figure C1.D). Our goal was to design a cavity filling amino acid substitution that

would stabilize the E state and prevent the repacking of the protein core to the E^{Ω} state. We chose cavity C-I (109.9 ± 6.0 $Å^3$) because it includes the Cβ atom of A104, which is a good target for small-to-large amino acid substitutions (Figure C2.A and Table C.S1). In fact, A104 is conserved as a phenylalanine in the nonallosteric UGDH from Streptococcus pyogenes, which does not have a cavity corresponding to C-I.[15] The C-I cavity in hUGDH will not accommodate an A104F substitution without significant changes in the protein core. In fact, we are not able to express soluble hUGDH with the A104F substitution. This is most likely due to the low level of sequence identity (23%) between SpUGDH and hUGDH, which has resulted in numerous changes in the protein core packing.[15] As an alternative, we used RosettaVIP to predict cavity filling amino acid substitutions that would preserve the integrity of the protein core. The following cavity C-I residues were included in the automated redesign routine of RosettaVIP: T91, T93, R102, A103, A104, L106, V132, K279, D280, and N283. Cavity residues P92 and P133 were excluded because prolines reduce the configurational entropy of the peptide backbone and are often important for protein folding and stability.[40] Additionally, P133 is conserved in nonallosteric UGDH from S. pyogenes, which suggests that it may be significant to the structure or function of the enzyme.[15] T131 was also excluded from the redesign because it contributes to the catalytic mechanism by coordinating a nucleophilic water.[41] RosettaVIP identified A104L as the optimal cavity filling substitution; no other substitution was identified in multiple (<10) stochastic design simulations. The predicted A104L substitution fills approximately 61% of cavity C-I and will favor the E state by introducing steric clashes into the core of the E^{Ω} state (Figure C2.B,C).

We also examined the A104L model for unsatisfied polar groups contributing to cavity C-I. Native cavity C-I contains three ordered water molecules that form several hydrogen bonds with polar groups in the protein core (Figure C2.A). The first water (W1) forms hydrogen bonds with

the carbonyl oxygens of R102 and K279 as well as Oγ of T93 and the second water molecule (W2) (Figure C2.A). Water (W2) also forms hydrogen bonds with the carbonyl oxygen of T91 and the third water molecule (W3), which also accepts a hydrogen from Nδ of N283 (Figure C2.A). The RosettaVIP A104L model conserves a small cavity that can accommodate W1 and its hydrogen bonds; however, the leucine substitution will displace W2 and W3 (Figure C2.A,B). In the absence of W2 and W3, the carbonyl oxygen of T91 can still be satisfied by Nζ of K279, and the caboxamide of N283 can be exposed to solvent by a small rotation about the χ1 torsion angle.

The $hUGDH_{A104L}$ Crystal Structure Reveals a Reduced Cavity C-I Volume

hUGDH with the A104L substitution ($hUGDH_{A104L}$) was co-crystallized with UDP-Glc and NADH to obtain the E state conformation ($hUGDH_{A104L}$:UDP-Glc:NADH). The structure was determined to a resolution of 2.53 Å and revealed two 32 symmetry hexamers in the asymmetric unit (Table C1). All residues are modeled with the exception of loop 385–387 and the C-terminus (residues 467–494), which are disordered in all 12 chains. Residues 23–30, 34–54, and 69–72 in the NB domain of chain A were poorly ordered and not modeled. Each monomer contains unambiguous electron density for the bound NADH and UDP-Glc (not shown). In all chains, the allosteric switch and hexamer-building interface are in the E state. The largest structural difference between the chains involves the rotation of the NAD^+ domain; chains G and J adopt the closed conformation, while the 10 remaining chains have NAD^+ domains that are rotated between 6.8° and 10.6° toward the open conformation. In chains G and J, the allosteric switch and hexamer-building interface are in the E state (Figure C.S1). Because the open domain conformation exposes the C-I cavity to solvent, we will limit our analysis of the A104L-induced changes in the cavity volume to the closed domain conformations in chains G and J.

The crystal structure of $hUGDH_{A104L}$ agrees well with the RosettaVIP model (Figure C2.B). The A104L substitution has split cavity C-I into two, smaller cavities with volumes of 20.7 ± 2.2 and 33.3 ± 2.9 $Å^3$ (Figure C2.B). The smaller cavity contains weak electron density for a water corresponding to W1 in chains B–F, K, and L. The electron density in the remaining chains is weak and did not support modeling the water. Because the absence of water W1 would leave unsatisfied polar groups, it is likely that there is a poorly ordered water in the remaining cavities (Figure C2.B).[42,43] As expected, the L104 side chain has displaced waters W2 and W3 (Figure C2.B). As a result, the Cε and Nζ atoms of K293 are disordered, which exposes the carbonyl oxygen of T91 to solvent. To compensate for the loss of the hydrogen bond with W3, the χ2 torsion angle of N283 has rotated from −10.3 ± 6° to the more favorable rotomer value of 49.4 ± 0.5°, which exposes the carboxamide to solvent. The only significant difference between the $hUGDH_{A104L}$ structure and the RosettaVIP model involves the χ2 torsion angle of N283, which is rotated 180° to flip the amide and carbonyl oxygen of the carboxamide group. Because it is not possible to assign the absolute conformation of the carboxamide group in the absence of known hydrogen bond acceptors or donors at this resolution, we chose to model the χ2 rotamer in the conformation originally reported in the 2Q3E crystal structure.[16]

To see if the A104L substitution destabilized the hexamer, we examined the oligomeric structure of native hUGDH and $hUGDH_{A104L}$ using sedimentation velocity. Consistent with previous reports,[15,18,20,21,23] the sedimentation of 9 μM native hUGDH reveals an 11.4 S hexamer (78.0%), an 8.2 S tetramer (3.8%), a 5.5 S dimer (13.3%), and a 3.1 S peak that likely represents a monomer (4.9%) (Figure C2.D). Similarly, the *c(s)* distribution of 9 μM $hUGDH_{A104L}$ consists of the 11.5 S hexamer (83.5%), a small amount of tetramer (<1%), the 5.5 S dimer (9.6%), and the 3.0 S monomer (6.9%) (Figure C2.D). The *c(s)* distribution of native hUGDH is in rapid

equilibrium, with the tetramer representing a transient between the more stable hexamer and dimer states.[18,21,23,44] Because of this, it was difficult to assign an accurate S value for the tetramer in the $hUGDH_{A104L}$ distribution.

$hUGDH_{A104L}$ Adopts the E State*

$hUGDH_{A104L}$ was also crystallized under conditions known to stabilize the native E* state.[18] The resulting crystals diffract to a resolution of 2.65 Å and are isomorphous with the native E* structure (PDB entry 4RJT) (Table C1). Briefly, the asymmetric unit contains three chains, with the complete hexamer being formed by crystallographic symmetry. All residues are modeled except for a surface loop (residues 385–387) and the C-terminus (residues 467–494), which are disordered in every chain. Additionally, there was incomplete main chain electron density for residue 72 in chain A as well as residues 13, 14, 49, and 154 in chain C, so they were not modeled. As in the native E* crystal structure,[18] the NB domains of E* $hUGDH_{A104L}$ are rotated to the open conformation. The only significant conformational change in the hexamer involves the NB domain of $hUGDH_{A104L}$ chain C, which is rotated ~0.9° less open than the corresponding chain in the native E* structure. The allosteric switch and hexamer-building interface of $hUGDH_{A104L}$ are also in the E* conformation (Figure C3.A). In native hUGDH, four residues in the protein core adopt χ1 torsion angles that that are unique to the E* state (Figure C3.B,C). Residues S88 and E161 adopt unique χ1 angles in the open domain E* state that would introduce steric clashes if the domain rotated to the closed conformation (Figure C3.B,C). Residue E128 changes conformation with the allosteric switch and adopts a unique rotamer conformation in the E, E*, and EΩ states (Figure C3.C). Finally, S120 adopts a χ1 rotamer angle of 67 ± 1°, which is observed only in the E* conformation. S120 is located at the C-terminus of the α5 helix, which is adjacent to the allosteric α6 helix in the hexamer-building interface (Figure C3.B,C). It is likely that the change

in the S120 χ1 rotamer is correlated with the movement of the α6 helix. In both χ1 rotamers, Oγ of S120 caps the carboxy terminus of the α5 helix by contributing a hydrogen bond to the carbonyl oxygen of I116. In the E* conformation of $hUGDH_{A104L}$, the χ1 rotamers of S88, E128, and E161 are conserved. However, Oγ of S120 is disordered and could not be modeled unambiguously.

The open domain conformation of E* exposes the allosteric switch and surrounding residues to solvent, which relaxes the steric packing constraints that would otherwise resist the transition between the E and E^{Ω} states.[20] A comparison of the $hUGDH_{A104L}$ and native E* normalized B-factors shows that the A104L substitution decreases the flexibility of several residues in cavity C-I (Figure C4.A). These changes originate from the A104L substitution, which packs against the allosteric switch residues V132 and P133, and the carboxamide of N283 (Figure C4.B,C). These stabilizing interactions are propagated to the surrounding second-shell residues V285, Y286, and L287, which become significantly more rigid (Figure C4.A,B). In contrast, the residues that are unique to cavity C-2 and declivities D-1 and D-2 show no significant change in flexibility (Figure C4.A).

hUGDHA104L Does Not Display Hysteresis or Negative Cooperativity

Unlike those of native hUGDH, the progress curves of $hUGDH_{A104L}$ do not exhibit hysteresis (Figure C5.A,B and Table C2). In fact, we observe no lag in the progress curves of $hUGDH_{A104L}$ even in the presence of UDP-Xyl, which is known[18,19] to increase the lag in native hUGDH progress curves (Table C2). Next, we examined the impact of the A104L substitution on steady state kinetics. The NAD^+ saturation curves of $hUGDH_{A104L}$ are hyperbolic (Hill = 1), with a K_M of 48 ± 3 μM (Figure C5.C and Table C3). In contrast, the NAD^+ saturation curves of native hUGDH are known to display negative cooperativity[18,19,23] (Figure C5.D and Table C3). Earlier work showed that UDP-Glc binding induces an asymmetry that results in a mixture of high- and

low-affinity NAD^+ binding sites.[19] Using Kurganov's analysis[39] (eq 4), the high (K_o) and low (K_{lim}) NAD^+ binding affinities for native hUGDH can be estimated (157 ± 13 and 2300 ± 500 μM, respectively) (Figure C5.E and Table C3). Both native hUGDH and $hUGDH_{A104L}$ have similar k_{cat} values with respect to NAD^+ (0.94 ± 0.07 and 0.79 ± 0.01 s^{-1}, respectively). The UDP-Glc saturation curves for both $hUGDH_{A104L}$ and native hUGDH are hyperbolic, with similar values for K_M (9.4 ± 0.5 and 7.6 ± 0.6 μM, respectively) (Table C3). The k_{cat} values observed in the UDP-Glc and NAD^+ saturation curves of $hUGDH_{A104L}$ are also similar (0.82 ± 0.01 and 0.79 ± 0.01 s^{-1}, respectively) (Table C3). The difference between the kcat values measured in the UDP-Glc and NAD^+ saturation curves of native hUGDH (0.59 ± 0.01 and 0.94 ± 0.07 s^{-1}, respectively) is a consequence of the negative cooperativity in NAD^+ binding, which makes it difficult to saturate hUGDH with coenzyme during the pseudo-first-order steady state analysis.[15]

$hUGDH_{A104L}$ Does Not Form the E^{Ω} State

Because the A104L substitution will introduce steric clashes into the E^{Ω} state, $hUGDH_{A104L}$ should favor the E and E* states. Attempts to crystallize $hUGDH_{A104L}$ in the E^{Ω} state using both of the previously published conditions[15] were not successful. However, the allosteric transition from the E^{Ω} to the E conformation can be observed in steady state inhibition studies of native hUGDH.[18,23] Briefly, UDP-Xyl binding favors the E^{Ω} state, which has a low affinity for UDP-Glc. As UDP-Glc outcompetes UDP-Xyl, the allosteric transition to the high-UDP-Glc affinity E state can be observed as substrate cooperativity (Figure C5.F,G and Table C4). As previously shown,[18] the degree of cooperativity increases with inhibitor concentration (Figure C5.F,G and Table C4). Global analysis of $hUGDH_{A104L}$ steady state inhibition curves shows that the UDP-Xyl affinity is 13-fold lower (K_i = 2.3 ± 0.2 μM) than that of native hUGDH (K_i = 0.18 ± 0.01 μM) (Table C4). In addition, none of the steady state inhibition curves for $hUGDH_{A104L}$ show evidence of substrate

cooperativity (Figure C5.H). Even at a saturating UDP-Xyl concentration, 15-fold higher than the K_i, the substrate saturation curves of $hUGDH_{A104L}$ are hyperbolic (Figure C5.I). This suggests that $hUGDH_{A104L}$ does not undergo the transition to the low-substrate affinity E^{Ω} state upon inhibitor binding.

The A104G Substitution Has No Significant Effect on Hysteresis or Allostery

The A104L substitution reduces the volume of cavity C-I and produces a more rigid protein core (Figures C2 and C4). We also investigated the effect of increasing the cavity volume and core flexibility using the large-to-small A104G substitution. The $hUGDH_{A104G}$ construct is expressed at a level (~2.5 mg/L) much lower than that of the native enzyme (~20 mg/L), which suggests that the substitution has negatively impacted folding. A decrease in stability may also explain the higher UDP-Glc K_M observed in $hUGDH_{A104G}$ (17.2 ± 1.1 compared to 7.4 ± 0.6 in the native form) (Table C4). Global analysis of UDP-Xyl inhibition shows that the UDP-Xyl affinity (K_i) of $hUGDH_{A104G}$ and substrate cooperativity are similar to that of the native enzyme (Figure C5.J and Table C4). Finally, the A104G substitution does not appear to have a significant effect on enzyme hysteresis; the progress curves of $hUGDH_{A104G}$ display lags that are similar to that observed in native hUGDH (Figure C5.K and Table C2).

C5. Discussion

The allosteric inhibition mechanism in hUGDH involves the ligand-induced isomerization of the T131 loop from the E or E* state to the E^{Ω} conformation (Figure C1.A–C).[15,18,20] In the absence of ligand, the NB domain is rotated open to expose the allosteric switch and surrounding core residues to solvent (Figure C6.Ai). This open conformation relaxes the core packing constraints, which allows the allosteric switch and some core residues to adopt the inactive E* state (Figure C3.A,C). The binding of UDP-Glc and NAD^+ triggers the allosteric switch and core

residues to adopt the active E state conformation, which allows the domains to rotate closed (Figure C6.Aii,iii). In contrast, UDP-Xyl binding induces the allosteric switch and the protein core to repack into the E^{Ω} state (Figure C6.Aiv,v). This repacking is facilitated by the remarkable plasticity of the protein core, which originates from the large cavities and declivities surrounding the T131 loop (Figures C1.D and C6.A).[15] In previous work,[23] we tested the hypothesis that the repacking of the allosteric switch is the source of hysteresis using the A136M substitution ($hUGDH_{A136M}$). We showed that A136M traps the allosteric switch in the active E state and abolished ligand-induced hysteresis (Figure C6.Bi → ii). However, it was not clear if the hysteresis was a feature of the inactive E* state, the inhibited E^{Ω} complex, or the repacking of the protein core.

Here, we have examined the contribution of protein core flexibility to both allostery and hysteresis by filling cavity C-I with the A104L substitution (Figures C1.D and C2.A). This cavity is present only in the E state, and our modeling studies showed that the A104L substitution would sterically prevent the repacking of the protein core into the E^{Ω} conformation (Figures C2.C and C6.C). We tested this using steady state inhibition kinetics; the allosteric transition in native hUGDH can be observed as substrate-induced cooperativity in the presence of a saturating level of UDP-Xyl[18] (Figure C5.F,G and Table C4). Thus, the absence of cooperativity in UDP-Xyl inhibition studies is strong evidence that the A104L substitution prevents the formation of the E^{Ω} state (Figure C5.H,I and Table C4). This interpretation is also consistent with the reduced UDP-Xyl affinity of $hUGDH_{A104L}$, which has been shown[23] to be a property of a constitutive E state $hUGDH_{A136M}$ (Table C4).

The progress curves of $hUGDH_{A104L}$ do not display hysteresis, despite the fact that the crystal structures show that the enzyme can adopt the E and E* states (Figures C3.A, C5.A, and C6.C, i). This is the first evidence that the source of hysteresis in hUGDH is not simply the

movement of the allosteric switch from the E* to E conformation.[15,19,21] Our data suggest that the conformational entropy of the native hUGDH protein core is an important factor in enzyme hysteresis.[19] In this model, the large packing defects allow residues in the protein core to sample nonproductive conformations as they converge to the correct E state rotamers. By filling cavity C-I with the A104L substitution, we sterically constrain the conformational freedom of the core, which should increase the rate at which the core residues converge to the E state conformation (Figure C6.C, i → ii). This interpretation is supported by our normalized *B*-factor analysis of the native and $hUGDH_{A104L}$ crystal structures, which shows $hUGDH_{A104L}$ has a more rigid protein core surrounding the substitution (Figure C4.A). To see if increasing the flexibility of the protein core would increase hysteresis, we introduced the A104G substitution. $hUGDH_{A104G}$ displays a rate of hysteresis and an inhibitor affinity that are similar to those of the native enzyme (Tables C2 and C4). This suggests that the main chain flexibility of A104 does not contribute to hysteresis. It is also likely that the small increase in cavity C-I volume due to the A104G substitution does not significantly increase the conformational entropy of the protein core residues. Taken together, our data show for the first time that the packing defects in the core of hUGDH play a significant role in both the allosteric and hysteretic mechanisms of hUGDH.

C6. Figures and Tables

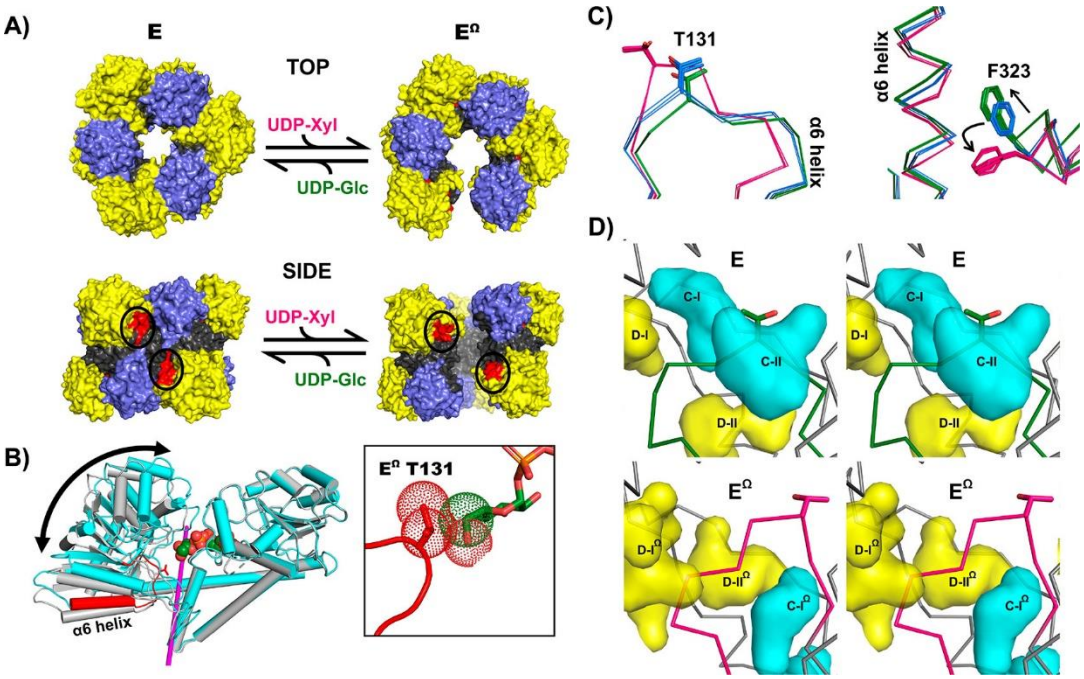


Figure C1: Allostery and hysteresis in hUGDH. (A) Active E and allosterically inhibited E^{Ω} states of hUGDH. The allosteric switch (red), the NB domain (yellow), the dimerization domain (gray), and the SB domain (purple) are identified. The α6 helix of the allosteric switch helix (red) in the hexamer-building interface is identified with black circles. **(B)** Cartoon depicting the rotation of the NB domain (curved arrow) about the hinge bending axis (magenta rod) from the open (gray) to the closed (cyan) state. The allosteric switch (red) and UDP-Glc (spheres) in the active site are shown. The inset box shows a close-up of the active site and depicts the van der Waals clash (dots) that would occur between T131 of the allosteric switch in the E^{Ω} state (red) and the C5 hydroxymethyl of UDP-Glc (green). UDP-Xyl lacks the C5 hydroxymethyl. **(C)** The left panel depicts the E (green), E* (blue), and E^{Ω} (pink) conformations of the T131 loop. The right panel shows the corresponding conformations of the α6 helix and F323 of the adjacent subunit in the hexamer-building interface. The arrows depict the directions that F323 in the E* state (blue) moves to adopt the E (green) or E^{Ω} state (pink). **(D)** Stereo diagrams of the allosteric switch in the E (top panel, green sticks) and E^{Ω} (bottom panel, pink sticks) states. Cavities (blue surfaces) and declivities (yellow surface) are identified. Surfaces were calculated using a 1.2 Å probe. The residues lining each cavity are listed in Table S1.

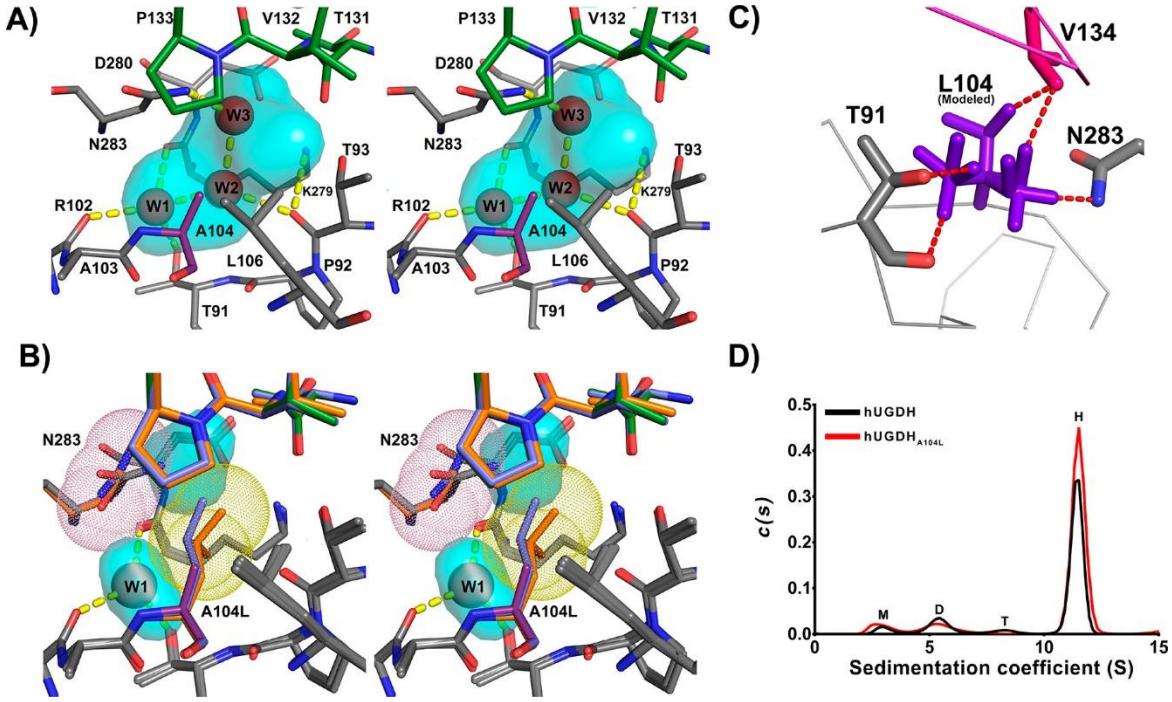


Figure C2: RosettaVIP design of the A10L cavity filling substitution in the E state. (A) Stereoview of the residues that line E state cavity C-I (cyan surface) and share hydrogen bonds (yellow dashes) with three buried waters (red spheres). A104L and the allosteric switch residues are colored purple and green, respectively. **(B)** Residues that line cavity C-I in the native hUGDH structure superimposed onto the RosettaVIP model and the $hUGDH_{A104L}$ crystal structure. The native A104 is colored purple (native), and all other residues are colored gray. The A104L substitution and allosteric switch are colored slate (RosettaVIP model), orange ($hUGDH_{A104L}$), or green (native hUGDH). The residual surfaces (cyan) of cavity C-I in $hUGDH_{A104L}$ and a single buried water molecule (gray sphere) are also shown. Surfaces were calculated with a 1.2 Å probe. The van der Waals radii of the $hUGDH_{A104L}$ L104 and the N283 carboxamide are depicted as yellow and pink dots, respectively. **(C)** Modeling L104 into the E protein core shows the substitution will introduce steric clashes in all rotameric states. In the E^{Ω} state (PDB entry 3PTZ), all preferred rotamers of L104 (purple) introduce steric clashes (red dashes). Even strained rotamers are sterically excluded (not shown). The E^{Ω} state of the allosteric switch is colored pink. **(D)** *c(s)* distribution of $hUGDH_{A104L}$ (red) and native hUGDH (black) as determined by sedimentation velocity analysis. M, D, T, and H correspond to the predicted S value of the monomer, dimer, tetramer, and hexamer, respectively.

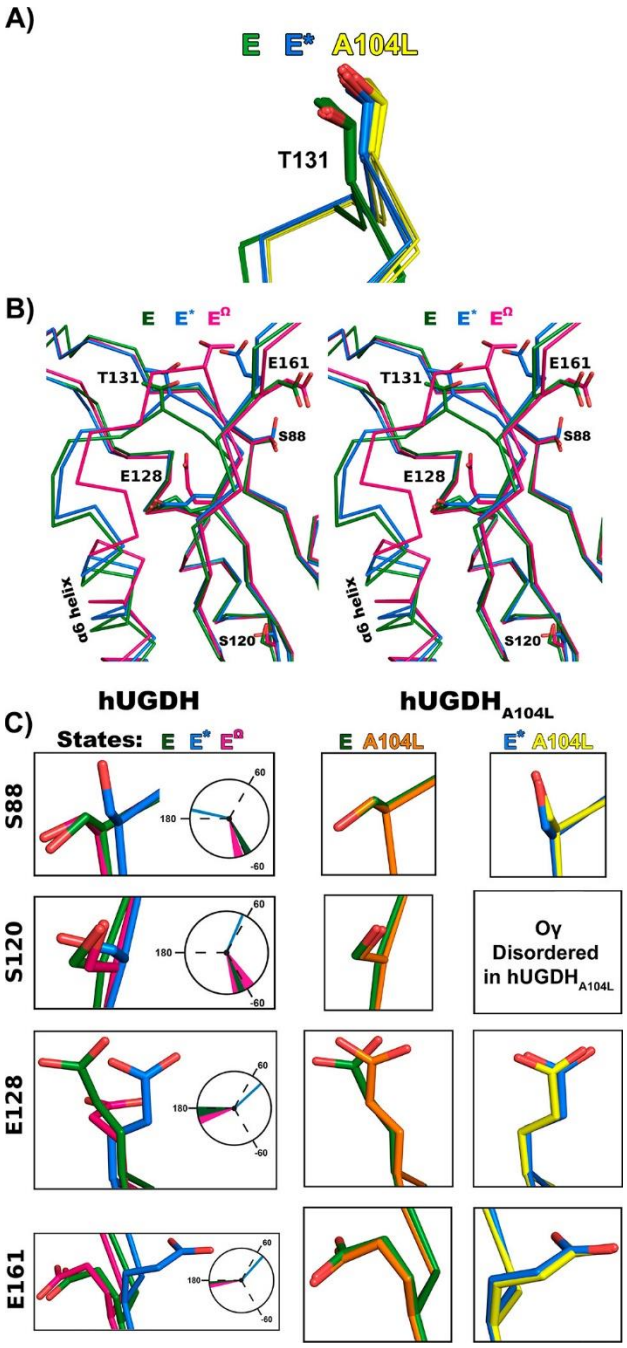


Figure C3: E* state of hUGDH$_{A104L}$. (A) The E* state of the hUGDHA104L allosteric switch (yellow) superimposes onto the native E (green) and E*(blue) conformations. **(B)** Stereodiagram of the Cα traces of the hUGDH protein core surrounding the allosteric switch in the E (green), E* (blue), and E^{Ω} (pink) states. Residue T131 is shown as sticks, as are residues that adopt rotamers unique to the E* state. The figure was created by superimposing residues 1–124. **(C)** Residues that adopt rotamers unique to the E* state. The column labeled hUGDH shows representative rotamers from the E (green), E* (blue), and E^{Ω} (pink) states as sticks with the corresponding χ1 angles observed in all chains in the native structures represented in a Newman projection. The last two

columns labeled $hUGDH_{A104L}$ show the E (orange) and E* (yellow) conformations superimposed onto the corresponding native conformation.

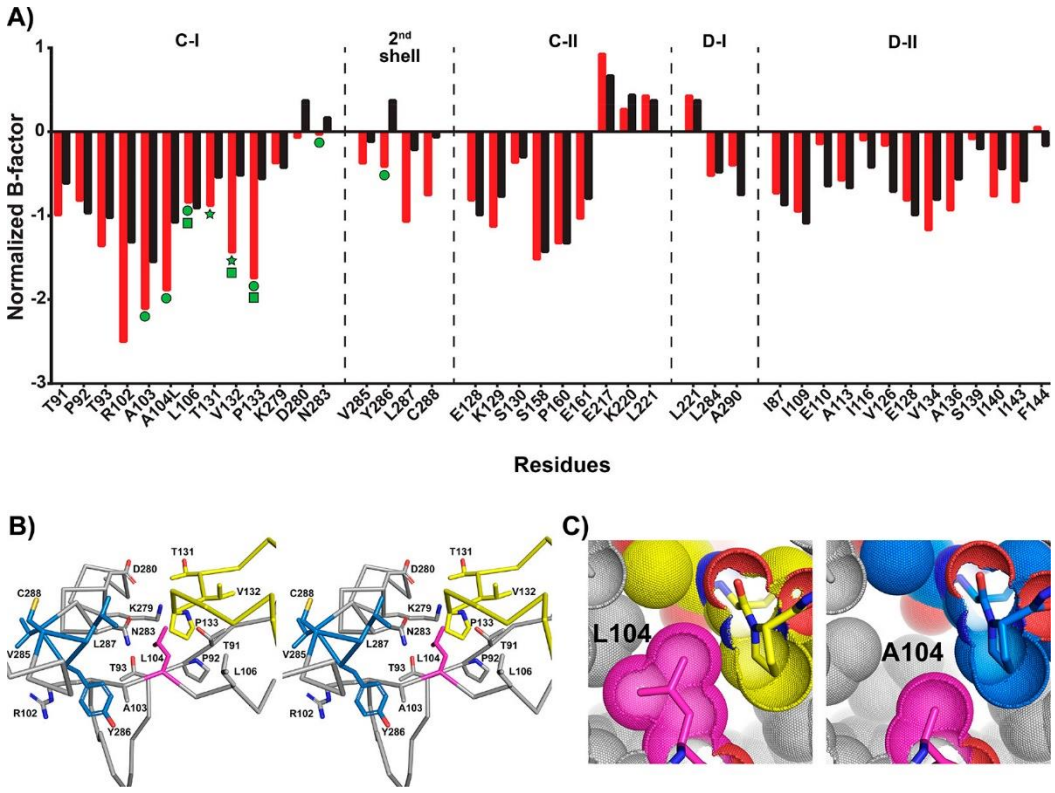


Figure C4: A104L substitution decreases the flexibility of cavity C-I and surrounding residues. (A) Histogram of the normalized *B*-factors (eq 1) for the Cα atoms of residues lining the packing defects (C-I, C-II, D-I, and D-II) of native hUGDH (black) and $hUGDH_{A104L}$ (red) structures. The second shell identifies residues that are adjacent to cavity C-I that show a significant decrease in flexibility. Residues that are shared between cavity C-I and the other packing defects are identified in the C-I plot with green symbols (stars, circles, and squares for C-II, D-I, and D-II, respectively). **(B)** Stereoview of the A104L form (magenta), allosteric switch (yellow), remaining cavity C-I residues (gray), and second-shell residues (blue). **(C)** The left panel depicts new van der Waals contacts between A104L (magenta) and the E* allosteric switch (yellow). The right panel is the same view with the native A104 (magenta) and the native E* allosteric switch (blue).

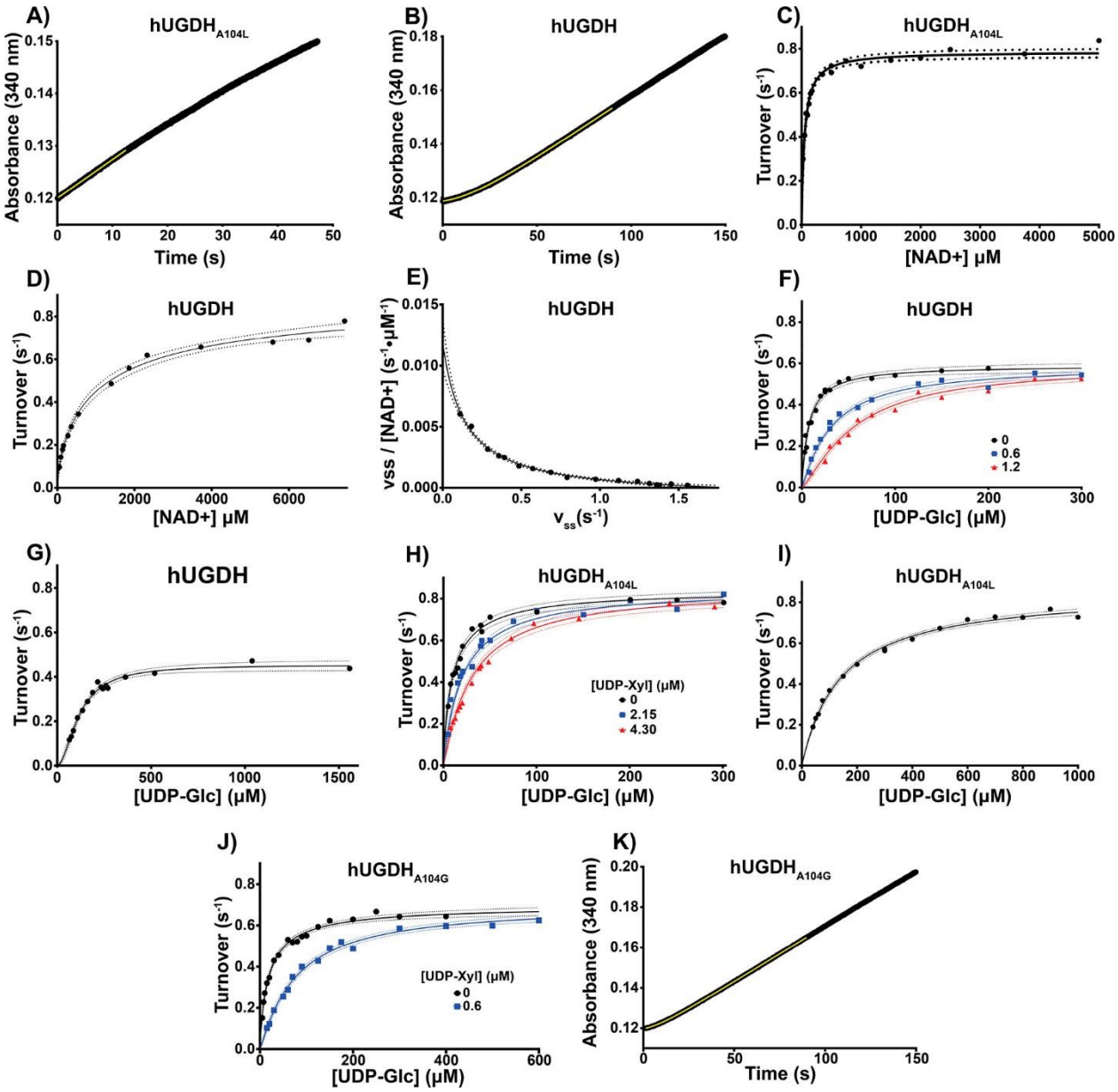


Figure C5: Kinetic analysis of native hUGDH and hUGDH$_{A104L}$. (A) Stopped-flow absorbance spectroscopy of an hUGDH$_{A104L}$ progress curve with the saturating substrate fit to a straight line (yellow). **(B)** Same as panel A but of a native hUGDH fit (yellow line) to Frieden's equation for hysteresis (eq 2). **(C)** NAD^+ substrate saturation curve with hUGDH$_{A104L}$ fit to eq 3. Rates are normalized by enzyme concentration to give turnover, and dashed lines represent the 95% confidence interval of the fit in all panels. **(D)** NAD^+ substrate saturation curve with native hUGDH as in panel C. For native hUGDH, the initial steady state velocities were obtained by fitting progress curves to Frieden's equation for hysteresis (eq 2), as described in Methods. **(E)** Eadie–Hofstee plot of the hUGDH NAD^+ saturation data that is concave up (negative cooperativity) and was fit to eq 4 to determine the K_0 and K_{lim} values in Table C4 (described in Methods). **(F)** hUGDH UDP-glucose saturation curves with 0 µM (black circles), 0.6 µM (blue squares), and 1.2 µM (red triangles) UDP-Xyl were globally fit to eq 5. **(G)** hUGDH UDP-Glc

saturation curves with 2.7 μM UDP-Xyl were fit to eq 3. **(H)** $hUGDH_{A104L}$ UDP-Glc saturation curves with 0 μM (black circles), 2.2 μM (blue squares), and 4.3 μM (red triangles) UDP-Xyl were globally fit to eq 5. **(I)** $hUGDH_{A104L}$ UDP-Glc saturation curves with 34 μM UDP-Xyl were fit to eq 3. **(J)** $hUGDH_{A104G}$ UDP-glucose saturation curves with 0 μM (black circles) and 0.6 μM (blue squares) UDP-Xyl were globally fit to eq 5. **(K)** The $hUGDH_{A104G}$ progress curve was fit to eq 2 (yellow).

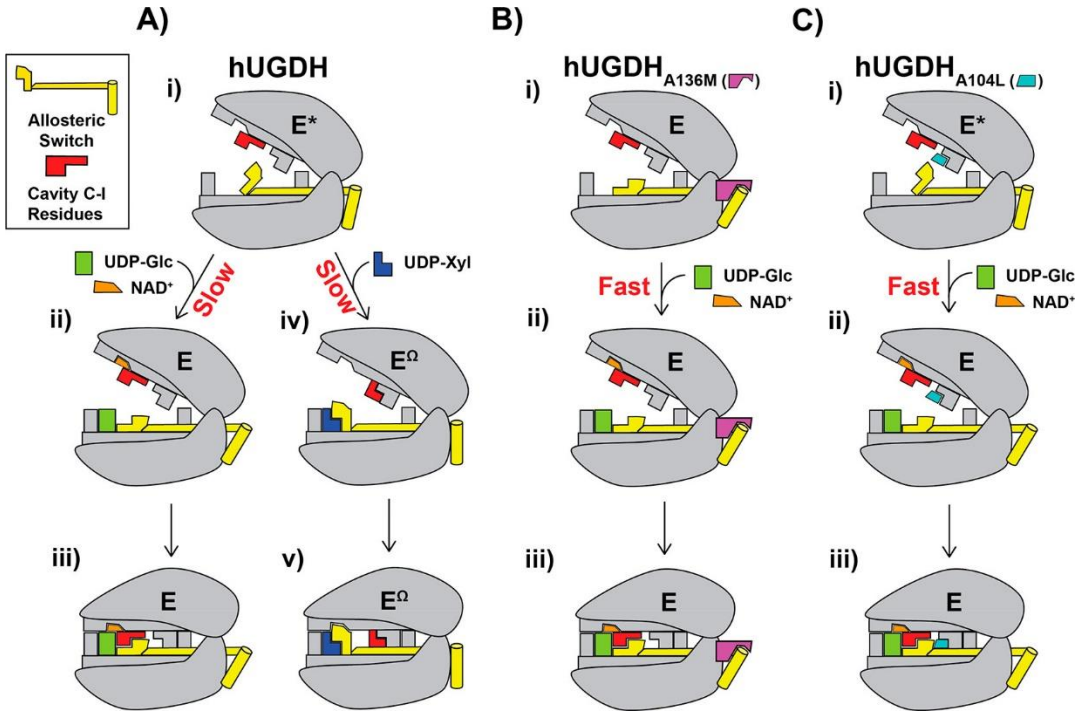


Figure C6: Proposed mechanism for allostery and hysteresis in hUGDH. (A) (i) A single monomer of hUGDH is represented by a gray clamshell with the allosteric switch (yellow) in the E* state and cavity C-I residues depicted as a red L-shape. The cavity C-I residues are depicted in the E state in the open domain conformation (as observed in the crystal structures), but it is likely that they can sample alternate rotamers, including those of the E^{Ω} state. (ii) Binding of UDP-Glc (green rectangle) and NAD^+ (orange trapezium) triggers the allosteric switch to adopt the active E state, and (iii) the domains rotate closed with the cavities and declivities represented as white spaces between the domains. In contrast, (iv) binding of UDP-Xyl (blue L-shape) triggers an allosteric switch; the cavity C-I residues adopt the E^{Ω} conformation, and then, (v) the domains rotate closed. In the E^{Ω} conformation, the allosteric switch occludes the NAD^+ binding site. Previously, we proposed[15] that the slow isomerization of E* to E (i → ii) or EΩ (i → iv) originated from the repacking of the allosteric switch and core residues. **(B)** (i) The A136M substitution (purple polygon) traps the allosteric switch in the E state. (ii) Here, binding of UDP-Glc and NAD^+ does not induce a slow isomerization, and (iii) the domains rotate closed. **(C)** (i) The A104L substitution (cyan trapezoid) prevents the cavity C-I residues from sampling the EΩ state. However, the allosteric switch is free to adopt the E* state, as seen in Figure C3.A. (ii) Upon binding of UDP-Glc and NAD^+, the allosteric switch rapidly adopts the E state and (iii) the domains rotate closed.

Table C1: Data collection and refinement statistics

	$hUGDH_{A104L}$:UDP-Glc:NADH	**$hUGDH_{A104L}$**
Protein Data Bank entry	6C4J	6C4K
Data collection		
Space Group	C121	C121
Unit cell dimensions (*a, b, c, β*)	403.5, 112.6, 184.7, 98.5°	178.5, 114.5, 96.9, 116.5°
Resolution Range	(48.63-2.53)	(48.16-2.65)
Completeness (%)	99.9 (99.9)[a]	99.7 (97.8)[a]
Redundancy	7.5 (7.5)	7.7 (7.7)
No. of reflections	2056278	389870
$I / \sigma(I)$	9.8 (1.0)	13.7 (1.1)
$CC_{1/2}$[b]	99.8 (53.3)	99.9 (52.8)
R_{meas}[c] (%)	14.8 (222.2)	12.5 (228.5)
Refinement		
Resolution (Å)	2.53 (2.56-2.53)	2.65 (2.71-2.65)
R_{work} / R_{free}	0.181/0.214	0.222/0.255
R_{free} in highest resolution shell	40.33	0.425
No. atoms Protein / Ligand / Water	43090 / 1043 / 594	10810 / 16 / 69
Wilson *B*-factor	62.3	71.5
B-factors (Å^2) Protein / Ligand / Water	73.2 / 72.6 / 59.1	105.3 / 77.5 / 63.5
Stereochemical Ideality		
Bond lengths (Å^2)	0.004	0.002
Bond angles (°)	0.64°	0.43°
φ,ψ Preferred region (%)	97.6	96.2
φ,ψ Additionally allowed (%)	2.4	3.8
φ,ψ Disallowed region (%)	0.0	0.0

[a] Values in parentheses are for the highest-resolution shell (2.559-2.530 Å and 2.712-2.650 Å for 6C4J and 6C4K, respectively).
[b] $CC_{1/2}$ is correlation between intensities from random half-data sets.[31]
[c] R_{meas} is the redundancy-independent merging *R* factor.[32]

Table C2: Effect of UDP-Xyl on Hysteresis

Enzyme	UDP-Xyl (μM)	τ (s^{-1})	Lag (*s*)
hUGDH	0	17.7 ± 1.4	48.2 ± 3.7
	1[a]	23.9 ± 2.0	65.0 ± 5.3
$hUGDH_{A104L}$	0	N.D.[b]	N.D.
	11.3[a]	N.D.	N.D.
$hUGDH_{A104G}$	0	16.2 ± 2.0	44.0 ±5.3
	1[a]	22.9 ± 2.2	62.2 ± 5.7

[a] These concentrations of UDP-Xyl are 5x K_i listed in Table 4
[b]N.D.= not detected

Table C3: Steady State Parameters

Enzyme	Substrate	K_M (μM)	Hill	k_{cat} (s^{-1})[a]	K_0[b]	K_{lim}[b]
hUGDH	UDP-Glc	7.6 ± 0.6	1	0.59±0.01[c]		
	NAD^+	1160 ± 290 [d]	0.69	0.94±0.07	157±13[b]	2300±500 [b]
A104L	UDP-Glc	9.4 ± 0.5	1	0.82±0.01		
	NAD^+	48 ± 3	1	0.79±0.01		

[a] One complete catalytic turnover produces 2 NADH.

[b] K_0 and K_{lim} represent the high and low affinity sites derived from Kurganov analysis[39419], respectively.

[c] The k_{cat} for the UDP-Glc saturation curve is underestimated due to the negative cooperativity of NAD^+ binding (see text).

[d] Because of negative cooperativity, this value represents the apparent half-saturation of the ES complex $K_{0.5}$.

Table C4: Global analysis of competitive inhibition

Enzyme	UDP-Xyl (μM)	UDP-Glc K_M (μM)	k_{cat} (s^{-1})[a]	K_i UDP-Xyl (μM)	Hill
hUGDH	0	7.4±0.6 [b]	0.59±0.01 [b]	0.18±0.01	1
	0.6				1.1±0.1
	1.2				1.3±0.1
	2.7[c]	7.4±0.1	0.45±0.01		1.85±0.14
$hUGDH_{A104L}$	0	10.3±0.6[b]	0.84±0.01 [b]	2.3±0.2	1
	2.2				1
	4.3				1
	34[c]	8.8±0.4	0.85±0.01		1
$hUGDH_{A104G}$	0	17.2 ± 1.1	0.67 ± 0.02	0.19±0.2	1
	0.6				1.2±0.1

[a] One complete catalytic turnover produces 2 NADH.

[b] For global analysis, this parameter was refined as a shared value.

[c] This experiment was excluded from global analysis because at long lags induced at high UDP-Xyl concentrations, native hUGDH is subject to NADH inhibition before the enzyme reaches steady state velocity.

C7. Supplementary Figures and Tables

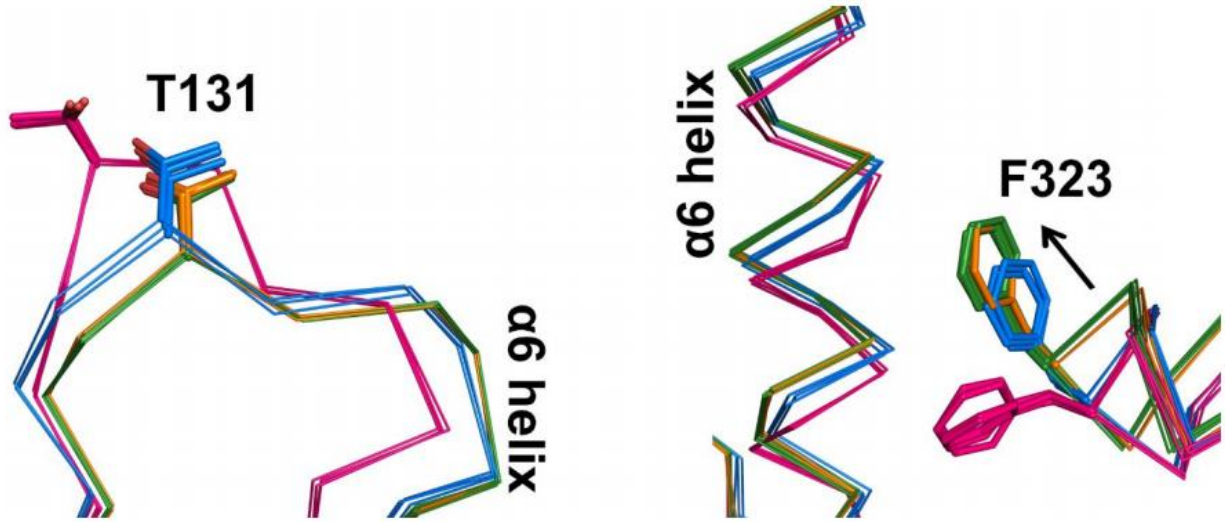


Figure C.S1: When substrate is bound hUGDHA104L adopts the E state. The $hUDGH_{A104L}$ (orange) structure superimposed onto the native structures in the E (green), E* (blue) and E^{Ω} (pink) conformations of the T131-loop (left). The α6 helix in the hexamer building interface and corresponding conformations of the F323 in the adjacent subunit (right). The arrows depict the directions that the F323 in the E*state moves to adopt the E or E^{Ω} states.

Table C.S1 Packing Defects and Associated Volumes

Packing Defect	Volume (Å^3)[a]	Waters[b]	Residue (Atoms)
C-1	109.9 ± 6.0	3	T91(O,Cβ,Cγ2,Oγ1), P92(C), T93(N,Cα,Oγ1), R102(O), A103(C,O), A104(N,Cα,Cβ), L106(Cδ2), T131(O,Cγ2), V132(Cα), P133(Cδ), K279(C,O,Cβ,Cγ,Cδ,Nζ), D280(Cα,Oδ1), N283(Cβ,Nδ2)
C-2	157.5 ± 16.8	4-5	E128(O,Cγ,Oε1), K129(O), S130(O), T131(Cα,Cγ2,Oγ1), V132(N,O,Cγ2), S158(Oγ), P160(N,Cα,Cβ,Cγ,Cδ) E161(N,O,Cβ,Cγ), E217(Cα,O,Cβ,Cγ), K220(Cβ,Cγ,Cδ,Cε), L221(N,Cγ,Cδ)
D-1	69.2 ± 23.6	4-6	A103(Cα,Cβ), A104(N,O,Cβ), L106(Cγ,Cδ1), P133(Cα,Cβ,Cγ), V134(N), L221(Cδ1), N283(Cα,O,Oδ1), L284(Cα), Y286(Cβ,Cδ,Cε2), A290(Cβ)
D-2	103.0 ± 26.4	1-4	I87(Cδ1), L106(Cβ,Cδ1,Cδ2), I109(C,O,Cβ,Cγ1,Cγ2), E110(Cα,Oε1) A113(N,Cα,Cβ), I116(Cδ1), V126(Cγ1), E128(Oε2), V132(Cγ1), P133(O,Cβ,Cγ,Cδ), V134(O), A136(N,Cα,O,Cβ), S139(C,Cβ,Oγ), I140(N,Cα,Cγ1,Cγ2,Cδ) I143(Cδ1), F144(Cζ)
C-1$^{\Omega}$	106.6 ± 10.6	1-2	I87(CβCγ1Cδ1), V126(C,O,Cγ1), T127(C), E128(N,Cα,O,Cβ,Cε1), S130(N,Cα,O,Cβ), T131(Cα,C,Oγ1), V132(Cγ1), P133(Cβ,Cγ), A137(Cα,Cβ), I140(Cγ2,Cδ2), V156(Cγ1,Cγ2), L157(O), S158(Cα,Cβ,Oγ), N159(N), P160(N,Cα,Cγ,Cδ), E217(Cγ, Oε2)
D-1$^{\Omega}$	130.3± 25.6	1-5	P133(O), V134(Cα,Cγ1,Cγ2), R135(N,Cα,C,Cγ,Cδ,Cε), E138(N,Cβ,Oε1), W214(Cγ,Cε,Cε,Cζ,Cη), E217(Cβ,Oε1), L221(Cδ1,Cδ2), N283(Cα,C,O,Cβ,Cδ), L284(N,Cα,Cδ1), Y286(O,Cβ,Cδ), L287(Cα,Cβ,Cδ1,Cδ2) A290(Cβ)
D-2$^{\Omega}$	157 ± 14.4	2-3	I87(C,O,Cγ2), S88(Cα), V89(N,O,Cγ2), N90(Cα,C,O,Cβ), T91(N,C,O,Cβ,Cγ2,Oγ1), A104(Cβ), L106(Cδ2), I109(Cγ1,Oδ1), E128(Oε2), K129(O), S130(Cβ,Oγ), T131(Oγ1), V132(N,O,Cβ,Cγ2), P133(Cα), V134(N,O,Cβ,Cγ2), A136(Cβ), S275(O), K279(Cε,Nζ)
C-1*	147.4 ± 8.6	0	E128(O,Cβ,Cγ,Cδ,Oε1), K129(C,O), S130(C,O), T131(N,Cα,Oγ1,Cγ2), V132(N,O,Cγ1,Cγ2), S158(Oγ), P160(N,Cα,Cβ,Cγ,Cδ), E161(N,O,Cγ,Cδ,Oε2), S216(O), E217(Cα,O,Cγ), K220(Cβ,Cγ,Cδ,Cε), L221(Cγ,Cδ2)

D-1*	100.5 ± 10.9	0	A103(Cα,Cβ), A104(N,O,Cβ), P133(Cα,Cβ,Cγ), V134(N,O,Cγ2) R135(Cζ,Nη1,Nη1), L221(Cδ2),N283(O,Cγ,Oδ1,Nδ2), L284(Cα), Y286(Cδ2,Cε2,Oη), L287(N,Cα,O,Cβ,Cδ2), A290(Cβ), L291(Cγ)
D-2*	136.0 ± 22.1	0	T91(O,Oγ1), P92(C), T93(N,Cα,Oγ1,Cγ2), G101(O), R102(Cα,O), A103(C,O), A104(N,Cα,Cβ), V132(C,O), P133(N,Cδ), L221(Cδ2), K279(Cα,Cγ,Cε,Nζ) D280(Oδ1), L282(Cβ,Cδ1), T131(O,Cγ2), N283(Cβ,Cγ,Oδ1,Nδ2)
D-3*	195.4 ± 30.5	0	I87(O,Cβ,Cγ1,Cγ2,Cδ1), L106(C,O,Cβ,Cδ1,Cδ2), K107(N,Cα,O,Cβ), I109(O,Cγ1,Cγ2,Cδ), A113(Cβ),V126(C,O,Cγ1), T127(C), E128(N,Cα,Cβ,Cγ,Cδ,Oε1), S130(Oγ),V132(Cγ1), E110(N,Cα,Cβ,Cγ,Cδ,Oε1) A136(C,O,Cβ), S139(C,Cβ,Oγ), I140(N,Cα,Cγ1,Cγ2,Cδ1), I143(Cδ1), V156(Cγ1,Cγ2), L157(O)

a calculated with w probe radius of 1.2 Å

b the number of waters is listed as a range where there is variability per chain in the observed ordered waters.

C8. References

(1) Lin, J. H., and Lu, A. Y. (1997) Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol. Rev. 49*, 403−449.

(2) Tukey, R. H., and Strassburg, C. P. (2000) Human UDP-glucuronosyltransferases: Metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol. 40*, 581−616.

(3) Guillemette, C. (2003) Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J.* 3, 136−158.

(4) Ritter, J. K. (2000) Roles of glucuronidation and UDP-glucuronosyltransferases in xenobiotic bioactivation reactions. *Chem.-Biol. Interact. 129*, 171−193.

(5) Cummings, J., Boyd, G., Ethell, B. T., Macpherson, J. S., Burchell, B., Smyth, J. F., and Jodrell, D. I. (2002) Enhanced clearance of topoisomerase I inhibitors from human colon cancer cells by glucuronidation. *Biochem. Pharmacol. 63*, 607−613.

(6) Cummings, J., Ethell, B. T., Jardine, L., Boyd, G., Macpherson, J. S., Burchell, B., Smyth, J. F., and Jodrell, D. I. (2003) Glucuronidation as a mechanism of intrinsic drug resistance in human colon cancer: Reversal of resistance by food additives. *Cancer Res. 63*, 8443−8450.

(7) de Almagro, M. C., Selga, E., Thibaut, R., Porte, C., Noe, V., and Ciudad, C. J. (2011) UDP-glucuronosyltransferase 1A6 overexpression in breast cancer cells resistant to methotrexate. *Biochem. Pharmacol. 81*, 60−70.

(8) Oguri, T., Takahashi, T., Miyazaki, M., Isobe, T., Kohn, N., Mackenzie, P. I., and Fujiwara, Y. (2004) UGT1A10 is responsible for SN-38 glucuronidation and its expression in human lung cancers. *Anticancer Res. 24*, 2893−2896.

(9) Kakizaki, I., Kojima, K., Takagaki, K., Endo, M., Kannagi, R., Ito, M., Maruo, Y., Sato, H., Yasuda, T., Mita, S., Kimata, K., and Itano, N. (2004) A novel mechanism for the inhibition of hyaluronan biosynthesis by 4-methylumbelliferone. *J. Biol. Chem. 279*, 33281−33289.

(10) Kultti, A., Pasonen-Seppanen, S., Jauhiainen, M., Rilla, K. J., Karna, R., Pyoria, E., Tammi, R. H., and Tammi, M. I. (2009) 4-Methylumbelliferone inhibits hyaluronan synthesis by depletion of cellular UDP-glucuronic acid and downregulation of hyaluronan synthase 2 and 3. *Exp. Cell Res. 315*, 1914−1923.

(11) Wei, Q., Galbenus, R., Raza, A., Cerny, R. L., and Simpson, M. A. (2009) Androgen-stimulated UDP-glucose dehydrogenase expression limits prostate androgen availability without impacting hyaluronan levels. *Cancer Res. 69*, 2332−2339.

(12) Kalckar, H. M., Maxwell, E. S., and Strominger, J. L. (1956) Some properties of uridine diphosphoglucose dehydrogenase. *Arch. Biochem. Biophys. 65*, 2−10.

(13) Bakker, H., Oka, T., Ashikov, A., Yadav, A., Berger, M., Rana, N. A., Bai, X., Jigami, Y., Haltiwanger, R. S., Esko, J. D., and GerardySchahn, R. (2009) Functional UDP-xylose transport across the endoplasmic reticulum/Golgi membrane in a Chinese hamster ovary cell mutant defective in UDP-xylose Synthase. *J. Biol. Chem. 284*, 2576−2583.

(14) Neufeld, E. F., and Hall, C. W. (1965) Inhibition of Udp-DGlucose Dehydrogenase by Udp-D-Xylose: A Possible Regulatory Mechanism. Biochem. *Biophys. Res. Commun. 19*, 456−461.

(15) Kadirvelraj, R., Sennett, N. C., Polizzi, S. J., Weitzel, S., and Wood, Z. A. (2011) Role of Packing Defects in the Evolution of Allostery and Induced Fit in Human UDP-Glucose Dehydrogenase. *Biochemistry 50*, 5780−5789.

(16) Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U., and Nidetzky, B. (2010) UDP-glucose dehydrogenase: structure and function of a potential drug target. *Biochem. Soc. Trans. 38*, 1378−1385.

(17) Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U., and Nidetzky, B. (2011) Structure and Mechanism of Human UDP-glucose 6-Dehydrogenase. *J. Biol. Chem. 286*, 23877−23887.

(18) Kadirvelraj, R., Custer, G. S., Keul, N. D., Sennett, N. C., Sidlo, A. M., Walsh, R. M., and Wood, Z. A. (2014) Hysteresis in Human UDP-Glucose Dehydrogenase Is Due to a Restrained Hexameric Structure That Favors Feedback Inhibition. *Biochemistry 53*, 8043−8051.

(19) Kadirvelraj, R., Sennett, N. C., Custer, G. S., Phillips, R. S., and Wood, Z. A. (2013) Hysteresis and Negative Cooperativity in Human UDP-Glucose Dehydrogenase. *Biochemistry 52*, 1456−1465.

(20) Sennett, N. C., Kadirvelraj, R., and Wood, Z. A. (2011) Conformational Flexibility in the Allosteric Regulation of Human UDP-alpha-D-Glucose 6-Dehydrogenase. *Biochemistry 50*, 9651−9663.

(21) Sennett, N. C., Kadirvelraj, R., and Wood, Z. A. (2012) Cofactor Binding Triggers a Molecular Switch To Allosterically Activate Human UDP-alpha-D-glucose 6-Dehydrogenase. *Biochemistry 51*, 9364−9374.

(22) Dickinson, F. M. (1988) Studies on the Unusual Behavior of Bovine Liver UDP-Glucose Dehydrogenase in Assays at Acid and Neutral pH and on the Presence of Tightly Bound Nucleotide Material in Purified Preparations of this Enzyme. *Biochem. J.* 255, 775−780.

(23) Beattie, N. R., Keul, N. D., Sidlo, A. M., and Wood, Z. A. (2017) Allostery and Hysteresis Are Coupled in Human UDP-Glucose Dehydrogenase. *Biochemistry 56*, 202−211.

(24) Kozlikova, B., Sebestova, E., Sustr, V., Brezovsky, J., Strnad, O., Daniel, L., Bednar, D., Pavelka, A., Manak, M., Bezdeka, M., Benes, P., Kotry, M., Gora, A., Damborsky, J., and Sochor, J. (2014) CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics 30*, 2684−2685.

(25) Borgo, B., and Havranek, J. J. (2012) Automated selection of stabilizing mutations in designed and natural proteins. *Proc. Natl. Acad. Sci. U. S. A. 109*, 1494−1499.

(26) Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, pp 571−607, Springer.

(27) Kabsch, W. (2010) XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 125−132.

(28) Brunger, A. T. (1997) Free R value: Cross-validation in crystallography. In *Macromolecular Crystallography, Part B (Carter, C. W., and Sweet, R. M., Eds.) pp 366−396*, Elsevier.

(29) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., GrosseKunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 213−221.

(30) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 486−501.

(31) Karplus, P. A., and Diederichs, K. (2012) Linking Crystallographic Model and Data Quality. *Science 336*, 1030−1033.

(32) Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol. 4*, 269−275.

(33) Hayward, S., and Berendsen, H. J.C. (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins: Struct., Funct., Genet. 30*, 144−154.

(34) Carugo, O., and Argos, P. (1997) Correlation between side chain mobility and conformation in protein structures. *Protein Eng., Des. Sel. 10*, 777−787.

(35) Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Analytical Ultracentrifugation in Biochemistry and Polymer Science. *R. Soc. Chem.*, 90−125.

(36) Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem. 320*, 104−124.

(37) Frieden, C. (1970) Kinetic aspects of regulation of metabolic processes. The hysteretic enzyme concept. *J. Biol. Chem. 245*, 5788−5799.

(38) Cornish-Bowden, A. (2001) Detection of errors of interpretation in experiments in enzyme kinetics. *Methods 24*, 181−190.

(39) Kurganov, B. I. (2000) Analysis of negative cooperativity for glutamate dehydrogenase. *Biophys. Chem. 87*, 185−199.

(40) Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. U. S. A. 84*, 6663−6667.

(41) Campbell, R. E., Mosimann, S. C., van De Rijn, I., Tanner, M. E., and Strynadka, N. C. (2000) The first structure of UDP-glucose dehydrogenase reveals the catalytic residues necessary for the two-fold oxidation. *Biochemistry 39*, 7012−7023.

(42) Levitt, M., and Park, B. H. (1993) Water: now you see it, now you don't. *Structure 1*, 223−226.

(43) Matthews, B. W., and Liu, L. (2009) A review about nothing: are apolar cavities in proteins really empty? *Protein Sci. 18*, 494−502.

(44) Zhao, H., Balbo, A., Brown, P. H., and Schuck, P. (2011) The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity. *Methods 54*, 16−30.