Fall 2016

Ligand-associated conformational changes of a flexible enzyme captured by harnessing the power of allostery

Sondra Faye Dean
University of Iowa

LIGAND-ASSOCIATED CONFORMATIONAL CHANGES OF A FLEXIBLE ENZYME CAPTURED BY HARNESSING THE POWER OF ALLOSTERY

by

Sondra Faye Dean

A thesis submitted in partial fulfillment
of the requirements for the Master of Science
degree in Pharmacy (Medicinal & Natural Products Chemistry) in the
Graduate College of
The University of Iowa

December 2016

Thesis Supervisor: Associate Professor M. Ashley Spies

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

MASTER'S THESIS

_____

This is to certify that the Master's thesis of

Sondra Faye Dean

has been approved by the Examining Committee for
the thesis requirement for the Master of Science degree
in Pharmacy (Medicinal & Natural Products Chemistry) at the December 2016
graduation.

Thesis Committee: _____
M. Ashley Spies, Thesis Supervisor

_____
Jonathan A. Doorn

_____
Michael W. Duffel

“Science is a way of thinking much more than it is a body of knowledge.”

Carl Sagan

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Ashley Spies. It has been a privilege to work for you. You have been great mentor and I have appreciated your insights, guidance, and support throughout my studies.

I am also grateful for my labmates: Nick Vance, Quinn Li, Laura Folly da Silva Constantino, and Dr. Katie Corum. I have appreciated your mentorship, collaboration, enthusiasm for science, and lively discussions.

I would like to thank the rest of my committee, Professor Jonathan Doorn and Professor Michael Duffel. Thank you for your time, effort, support, and willingness to help during this process.

I am also thankful for my parents. Thank you for being my cheerleaders and for your unceasing encouragement. Most importantly, thank you for being a constant source of love and support.

Finally, I would like to thank Sam. You are my number one. I appreciate you, your support, your willingness to listen, your calming presence, your clever wit, and the tremendous sacrifice you made letting me borrow your laptop to write this.

ABSTRACT

Flexible enzymes are notoriously a bane to structure-based drug design and discovery efforts. This is because no single structure can accurately capture the vast array of conformations that exist in solution and many are subject to ligand-associated structural changes that are difficult to predict. Glutamate racemase (GR) – an antibiotic drug discovery target involved in cell wall biosynthesis – is one such enzyme that has eluded basic structure-based drug design and discovery efforts due to these flexibility issues. In this study, our focus is on overcoming the impediment of unpredictable ligand-associated structural changes in GR drug discovery campaigns. The flexibility of the GR active site is such that it is capable of accommodating ligands with very different structures. Though these ligands may bind to the same pocket, they may associate with quite dissimilar conformations where some are more favorable for complexation than others. Knowledge of these changes is invaluable in guiding drug discovery efforts, indicating which compounds selectively associate with more favorable conformations and are therefore better suited for optimization and providing starting structures to guide structure-based drug design optimization efforts. In this study, we develop a mutant GR possessing a genetically encoded non-natural fluorescent amino acid in a region remote from the active site whose movement has been previously observed to correlate with active site changes. With this mutant GR, we observe a differential fluorescence pattern upon binding of two structurally distinct competitive inhibitors known to associate with unique GR conformations – one to a favorable conformation with a smaller, less solvated active site and the other to an unfavorable conformation with a larger, more solvated active site. A concomitant computational study ascribes the source of this differential

fluorescence pattern to ligand-associated conformational changes resulting in changes to the local environment of the fluorescent residue. Therefore, this mutant permits the elucidation of valuable structural information with relative ease by simply monitoring the fluorescence pattern resulting from ligand binding, which indicates whether the ligand has bound to a favorable or unfavorable conformation and offers insight into the general structure of this conformation.

PUBLIC ABSTRACT

Drug discovery efforts often rely on detailed structural information of proteins that play an important role in some biological process whose disruption or alteration could result in a desired outcome for improved health (i.e., proteins that make for promising drug targets). However, this structural information can be difficult to attain in some cases due to the fact that some proteins are quite flexible such that they adopt an array of structural variants that are difficult to capture. This is problematic because structure-based drug development efforts rely on accurate and precise structural information to examine (a) whether a compound will bind and (b) what a protein–compound structure looks like. This information may be examined virtually but must be validated experimentally. Experimental assessment of binding is often a rather facile process, but experimental assessment of the structure of the protein–compound complex can be quite difficult to determine. In this study, we have developed a method of assessing both binding and quality of binding of compounds to a flexible protein that is a promising antibiotic drug target. The protein of interest has been mutated such that it fluoresces. Different fluorescence patterns are observed upon complexation with two compounds known to associate with distinct structural variants, indicating its capacity to identify these structural variants, one of which is more favorable for compound binding than the other. The described method may be applied to high-throughput screening campaigns to identify favorable protein-compound complexes and guide optimization efforts following the discovery of an initial hit.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Glutamate racemase (GR) is a bacterial cell wall enzyme of great interest for antibiotic drug development. This enzyme – found in both Gram-positive and Gram-negative bacteria - is responsible for the turnover of L- to D-glutamate,[1] which is then incorporated into the cross-linking side chains of the peptidoglycan.[2] The activity of this enzyme is essential,[3] making it a promising candidate for an antibiotic drug target. Its essentiality was first shown in *E. coli*, where it was found that disrupted GR activity produced cellular lysis.[4] Its essentiality has since been shown to be ubiquitous amongst bacteria, even in Gram-positive bacteria possessing D-amino acid transaminase (D-AAT), which is capable of producing D-glutamate.[3,5-10] Additionally, there is no human isozyme, reducing the likelihood of off-target effects of any GR targeting drug and further enhancing its attractiveness as an antibiotic drug target.[2,11,12]

Given that roughly 60% of all antibiotics on the market target bacterial cell wall biosynthesis pathways,[3] there seems to be a high probability of success in developing a GR inhibitor into an effective antibiotic. In fact, some GR inhibitors have already been found to possess antimicrobial activity.[13,14] Despite the great promise of GR as a novel antibiotic drug target, it is in fact an exceedingly difficult target.

Flexibility is vital to the catalytic activity of this enzyme. A study of GRs encoded by the murI gene (MurI enzymes) showed that each monomer of these GR structures contains two domains that move as a hinge relative to one another about a single axis. At the interface of these domains, catalysis occurs. Each domain possesses one half of the residues necessary for catalysis – one domain possesses residues responsible for deprotonation of the substrate and the opposite domain possesses residues responsible for reprotonation to produce the isomer of the substrate.[15] Another study performed on *B.*

subtilis GR (RacE) – the enzyme examined in Chapter 2 – showed similar structural features as the MurI isozymes and demonstrated a high degree of flexibility, as well. Here, the substrate was found to bind only after an extensive conformational change due to the deep, highly buried nature of the substrate binding pocket.[11] A study conducted by Whalen et al[16] on *B. subtilis* GR examined substrate unbinding revealing that considerable conformational changes must occur in this process as well, further supporting that this is indeed a highly flexible enzyme.

Due to the high degree of flexibility of GR,[11,15,17-19] it is extraordinarily difficult to predict conformational changes that will occur upon ligand binding. This creates considerable challenges in the drug discovery and development process, particularly from a structure-based drug design and development point of view. It raises questions such as: to which conformation will any given ligand associate with? Will optimization of a ligand of interest result in association with a different conformation than the original hit? Will ligands that bind to the same pocket associate with similar conformations? Will ligands with similar properties associate with similar conformations? How can these conformational changes be readily elucidated? Some structural information may be predicted *a priori* via *in silico* methods but experimental validation is necessary and not always a facile endeavor.

Fluorescent labeling such as L-(7-hydroxycoumarin-4-yl) ethylglycine (7HC) is one method that has shown success in elucidating structural information with relative ease. Compared to other fluorescent labels, 7HC has the capacity to be incorporated into nearly any location on a protein, possesses a large fluorescence quantum yield, has an increased Stoke's shift, and is quite sensitive to the environment – including pH and

polarity of the solvent.[20,21] This probe has been successfully used to assess unfolding of sperm whale myoglobin induced by urea,[20] to monitor nucleotide binding to *E. coli* ATCase,[22] and follow the phosphorylation of signal transducer and activator of transcription 3 (STAT3).[23] This fluorescent probe has also been successfully applied to a number of FRET studies.[24-26]

In the study described in Chapter 2, we monitor ligand-associated structural changes of *Bacillus subtilis* GR (RacE) labelled with 7HC. The location selected for 7HC incorporation is Tyr53, whose movement has been identified as correlating with active site changes. The ligands we study are croconate and glucoronate, two competitive inhibitors of GR with distinct chemotypes that were previously observed to preferentially associate with different GR conformations.[16] This phenomenon of ligand-associated conformational changes is not unique to GR and has been observed in numerous proteins and drug targets.[27] We incorporate 7HC into the Tyr53 position to create the mutant $GR^{Y53/7HC}$ and assess whether fluorescence changes of the 7HC residue capture the anticipated ligand-associated structural changes upon titration with croconate and glucoronate. We also perform a concomitant computational study to closely investigate the ligand-associated conformational changes occurring upon binding of croconate and glucoronate and the specific changes occurring in the local environment of 7HC during these events.

2. BIOSYNTHESIS OF A NOVEL GLUTAMATE RACEMASE CONTAINING A SITE-SPECIFIC 7-HYDROXYCOUMARIN AMINO ACID: ENZYME-LIGAND PROMISCUITY REVEALED AT THE ATOMISTIC LEVEL

Glutamate racemase (GR) catalyzes the cofactor-independent stereoinversion of L- to D-glutamate for biosynthesis of bacterial cell walls. Because of its essential nature, this enzyme is under intense scrutiny as a drug target for the design of novel antimicrobial agents. However, the flexibility of the enzyme has made inhibitor design challenging. Previous steered molecular dynamics (MD), docking, and experimental studies have suggested that the enzyme forms highly varied complexes with different competitive inhibitor scaffolds. The current study employs a mutant orthogonal tRNA/aminoacyl-tRNA synthetase pair to genetically encode a non-natural fluorescent amino acid, L-(7-hydroxycoumarin-4-yl) ethylglycine (7HC), into a region (Tyr53) remote from the active site (previously identified by MD studies as undergoing ligand-associated changes) to generate an active mutant enzyme ($GR^{Y53/7HC}$). The $GR^{Y53/7HC}$ enzyme is an active racemase, which permitted us to examine the nature of these idiosyncratic ligand-associated phenomena. One type of competitive inhibitor resulted in a dose-dependent quenching of the fluorescence of $GR^{Y53/7HC}$, while another type of competitive inhibitor resulted in a dose-dependent increase in fluorescence of $GR^{Y53/7HC}$. In order to investigate the environmental changes of the 7HC ring system that are distinctly associated with each of the $GR^{Y53/7HC}$–ligand complexes, and thus the source of the disparate quenching phenomena, a parallel computational study is described, which includes essential dynamics, ensemble docking and MD simulations of the relevant $GR^{Y53/7HC}$–ligand complexes. The changes in the solvent exposure of the 7HC ring system due to ligand-associated GR changes are consistent with the experimentally observed quenching

phenomena. This study describes an approach for rationally predicting global protein allostery resulting from enzyme ligation to distinctive inhibitor scaffolds. The implications for fragment-based drug discovery and high throughput screening are discussed. This chapter was originally published as Dean, S. F., Whalen, K. L., and Spies, M. A., (2015) Biosynthesis of a Novel Glutamate Racemase Containing a Site-Specific 7-Hydroxycoumarin Amino Acid: Enzyme-Ligand Promiscuity Revealed at the Atomistic Level. *ACS Central Science* 1, 364-373 [http://pubs.acs.org/doi/full/10.1021/acscentsci.5b00211] under an open access license from the American Chemical Society. The supplementary material of the original article has been incorporated here into the body of the described work. Katie Whalen performed the experimental portion of the described work while I performed the computational studies.

2.1 Introduction

Many protein and enzyme drug targets are intractable to the most sophisticated methods in structure-based drug design and discovery (SBDDD) in large part because a high degree of flexibility negates the use of one, or even a few, crystal structures in virtual screening or *de novo* design. Recent years have seen the development of a myriad of tools for obtaining ensembles, which can be used in SBDDD approaches, with mixed results. There are many mechanistic options to consider when attempting to model a flexible enzyme drug target. One of the most critical parameters is whether a "selection" or an "induced-fit" model will be followed. Although it is exceedingly rare for any protein−ligand binding study to contain a sufficient level of physical detail to elucidate selection vs. induced-fit, a number of exhaustive studies (both experimental and

computational) have recently shown them both to take place in particular systems.[28-30] Indeed, it is possible for a variety of complicated phenomena to be operational in a single enzyme−ligand association process.[31] Nevertheless, from an SBDDD point of view, a related question is can we employ experimental and computational methods that give us insight into the ligand-associated enzyme changes, independent of exactly how they are manifested?

In the current study, experimental and computational insights into protein−ligand complexation are obtained for a highly flexible enzyme, glutamate racemase (GR), which exhibits very puzzling and idiosyncratic ligand-associated changes. These ligand-associated changes have previously been described using computational selection models, in which a very flexible apo enzyme produces a diverse ensemble of conformations possessing ligand-binding pockets of variable solvent accessible surface areas and protein solvation energies, which has significant implications on the quality of ligand binding to each of these distinct enzyme conformations.[16] The current study employs an approach that allows for both selection and induced-fit changes in the computational workflow, concomitant with parallel experimental studies.

GR is a bacterial enzyme responsible for the essential task of the stereoisomerization of L-glutamate to D-glutamate, a member of the cross-linking peptide side chain component of all bacterial peptidoglycan cell walls.[2] D-Glutamate is not readily available in the environment, and no other viable biosynthetic pathways exists for its generation intracellularly, making GR a vital enzyme to bacterial growth and survival.[32] The lack of any human isozyme makes GR a valuable target for potential antibacterial therapeutics.

Several isozymes of GR have been the subject of X-ray crystallography campaigns, providing a suite of cocrystal structures for use in SBDDD.[11,15,33] One of the first structures solved belonged to the nonpathogenic, mesophilic bacteria *Bacillus subtilis*.[11] Thus, this isozyme has served as a model system for an assortment of functional studies and drug discovery campaigns.[16-18,34] In general, GR has proven to be a flexible enzyme capable of binding a variety of small molecules both in the active site and at allosteric binding pockets,[15,19] some of which have included cyclic substrate-product analogues.[35] Further complications of GR−ligand complexation are derived from the important role that interstitial waters play in recognizing various classes of competitive inhibitors.[34]

Inherent flexibility is notoriously a bane to traditional SBDDD, largely due to the unknown nature of the relationship between ligand binding and enzyme conformation. This limits the representative nature of single cocrystal structures and requires the acquisition of additional structural information concurrent with hit discovery and lead optimization. The current study addresses these difficulties in the case of GR, by employing an integrated experimental and computational approach to gain insight into the nature of ligand-dependent changes in complexation. Site-specific incorporation of a genetically encoded fluorescent non-natural amino acid is used to generate a GR which substitutes Tyr53, located in a dynamic region remote from the active site, with L-(7-hydroxycoumarin-4-yl) ethylglycine (7HC) (using the approach developed by Wang et al.,[20] as described below), to make an active enzyme ($GR^{Y53/7HC}$). $GR^{Y53/7HC}$ is a sensitive reporter for ligand-dependent changes and is studied with two differing competitive inhibitors (e.g., croconate and glucuronate), which have been previously identified as

selecting distinct forms of GR.[16] A parallel computational study of ligand-GR$^{Y53/7HC}$ selection and dynamics is employed to elucidate the source of the ligand-associated allosteric changes. The use of GR$^{Y53/7HC}$ offers insight into ligand-associated changes taking place within the active site, which is reflected by global changes occurring remote from the active site and is reported on by the fluorescent probe. The implications of this study are highly relevant to improving approaches to fragment-based drug discovery, virtual screening, and high throughput screening against flexible enzymes and proteins (*vide infra*).

2.2 Experimental Methods

2.2.1 Site-directed Mutagenesis

The mutant racE_g819a gene and subsequent racE_g819a_Y53tag gene was prepared using a QuikChange II Site-Directed Mutagenesis Kit (Stratagene, Santa Clara, CA) and primers obtained from Eurofins MWG Operon (Huntsville, AL). Previously prepared and recently isolated pET15b (Novagen, San Diego, CA) containing the gene of interest was used as the template DNA. Template DNA was constructed previously, as described in Spies et al.[17] A BioRad MJ Mini Personal Thermal Cycler (BioRad, Hercules, CA) was used for all PCR reactions. Mutagenesis was confirmed via in-house DNA sequencing using an ABI 3730XL capillary sequencer. Primer sequences are detailed in Table 1. *E. coli* XL1 Blue cells were transformed with PCR products.

Table 1. Primers for site-directed mutagenesis.

Gene	Desired Mutation	Primer	Primer Sequence (5'→3')
racE	g819a	g819aTermFor	5'- ctgcaagaaccgattaaaagataaggatccggctgc - 3'
		g819aTermRev	5'- gcagccggatccttatcttttaatcggttcttgcag - 3'
racE-g819a	Tyr53tag	Y53AmberFor	5'- cctgaagaagaggtgcttcaatagacgtgggagctg - 3'
		Y53AmberRev	5'- cagctcccacgtctattgaagcacctcttcttcagg - 3'

2.2.2 Protein Expression and Purification

Proteins were expressed using the BL21-DE3 strain (Novagen, San Diego, CA) of *E. coli* and pET15b expression vector (Novagen, San Diego, CA). Cells were first transformed via heat shock with the pEB-JYRS(couRS) plasmid acquired from the Peter Schultz lab (Univ. of California at San Diego). The resulting transformants were made chemically competent and transformed with the pET15b plasmid containing the genes of interest, described above. Overnight cultures were incubated at 37°C in the presence of 15 µg/mL tetracycline and 50 µg/mL ampicillin. Cultures were back-diluted into 50 mL Luria Broth with antibiotics and grown at 37°C with shaking until reaching an optical density at 600 nm of 0.5. L-(7-hydroxycoumarin-4- yl) ethylglycine was added at a final concentration of 1 mM and cultures were incubated an additional 15 min. IPTG was added at a final concentration of 0.2 mM to induce expression and the temperature was reduced to 30°C. Expression was carried out for 18-20 h. Cells were harvested and lysed via sonication. His-tagged proteins were purified via cobalt-affinity chromatography (His-Select, Sigma-Aldrich) and concentrated to a concentration of 1-2 mg/mL.

2.2.3 Bulk Fluorescence Measurements

Fluorescent scans were measured with the following excitation/emission wavelengths using a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA) at 25°C: 340 nm/400-500 nm; 295 nm/300-375 nm; and 280 nm/300-375 nm. GR-WT (2.5 µM) and $GR^{Y53/7HC}$ (10 µM) were diluted with protein storage buffer (50 mM Tris, 100 mM NaCl, pH 8.0) to the indicated final concentrations and measured in a quartz cuvette (pathlength = 1 cm). The PMT voltage was set to 800

V. Titration experiments were conducted at the protein concentrations indicated above inside the quartz cuvette. Stock solutions of croconic acid and glucuronic acid in protein storage buffer were prepared at 1 mM or 50 mM, respectively. Neither ligand has significant intrinsic fluorescence at this particular excitation/emission (Figure 1). The excitation wavelength was set to 340 nm and the emission wavelength was fixed at 455 nm. Fluorescence intensity was measured continuously over the course of the titration, with 1-1.5 min intervals between injections of ligand. Fluorescence was averaged over the 1-1.5 min interval to produce each data point. Fluorescence was corrected for dilution caused by the addition of ligand and then plotted as a function of ligand concentration. A one-site binding equation was fit to the data using QtiPlot. Titration data fitting of fluorescent titrations of croconate and glucuronate to $GR^{Y53/7HC}$ yield a Hill coefficient of unity (Figure 2).

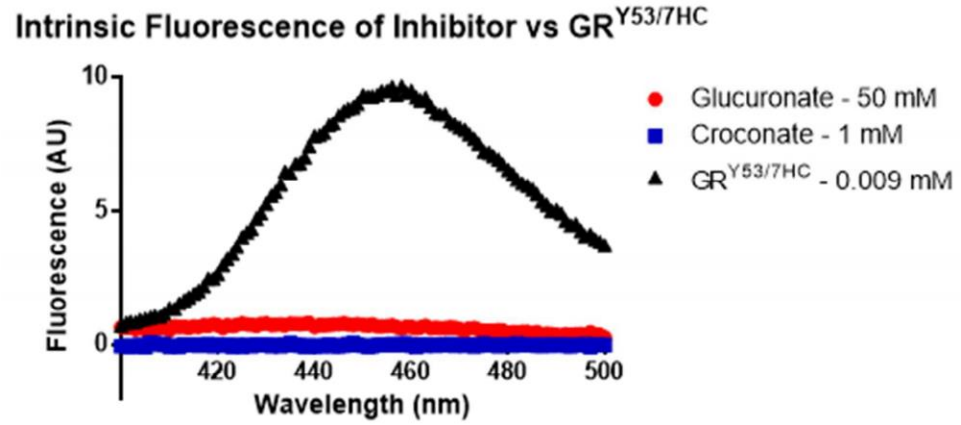

Figure 1. Croconate and glucuronate were scanned for emission at the excitation wavelength (340 nm) and no significant fluorescent signal was detected by either compound relative to $GR^{Y53/7HC}$.

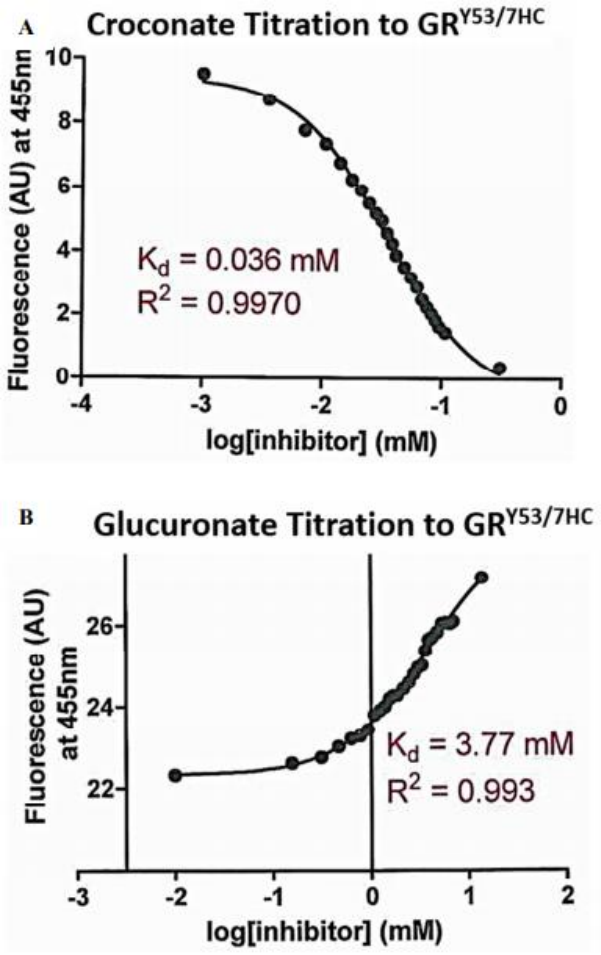

Figure 2. Data fitting of fluorescent titrations of (A) croconate and (B) glucuronate to $GR^{Y53/7HC}$ demonstrate Hill coefficients of unity indicating 1:1 binding stoichiometry of ligand to enzyme. Titration of croconate to $GR^{Y53/7HC}$ yielded a Hill coefficient of -1.158 ± 0.060 and titration of glucuronate to $GR^{Y53/7HC}$ yielded a Hill coefficient of 0.9649 ± 0.099.

2.2.4 Materials

L-(7-hydroxycoumarin-4-yl) ethylglycine was synthesized under contract from AsisChem Inc., using the procedure described in Wang et al.[20]

2.3 Computational Methods

2.3.1 Atomistic Molecular Dynamics Simulations of Unliganded GR and GR$^{Y53/7HC}$-ligand Complexes

The protocol used for the atomistic MD simulations was as described in Whalen and Spies,[34] using the starting structure from PDB 1ZUW (*B. subtilis* GR),[11] except for a few notable exceptions, which are summarized below. The YASARA Structure package version 13.4.21[36] was used to perform all simulations, employing an TIP3P explicit solvent model with a periodic simulation cell with boundaries extending 10 Å from the surface of the complex, and the cell was neutralized with NaCl (0.9% by mass), as described in Whalen and Spies,[34] but employing the YAMBER3 knowledge based force field,[37] which was used with long-range electrostatic potentials calculated with the Particle Mesh Ewald (PME) method,[38] with a cutoff of 7.86 Å. For GR$^{Y53/7HC}$, the Build utility of MOE v2013.08 (Chemical Computing Group)[39] was used to transform Tyr53 to the 7HC residue. The ligand and 7HC force field parameters were generated with the AutoSMILES utility,[40] which employs semiempirical AM1 geometry optimization and assignment of charges, followed by assignment of the AM1BCC atom and bond types with refinement using the RESP charges, and finally the assignments of general AMBER force field atom types.[41,42] Prior to initiating the MD simulation, an optimization of the hydrogen bond network of the various apo GR or GR$^{Y53/7HC}$–ligand complexes was obtained using the method established by Hooft et al.[43] in order to address ambiguities arising from multiple side chain conformations and protonation states that are not well resolved in the electron density. Following neutralization, a final density of 0.997 g/mL was employed. A previously described simulation annealing protocol was followed before initiation of simulations using the NVT ensemble at 298 K, and integration time

steps of 1.25 and 2.5 fs for intra- and intermolecular forces, respectively. Figure 3 shows the RMSD changes for this species over the course of the simulation. An equilibrated ensemble was reached after 1 ns of initial MD simulation.

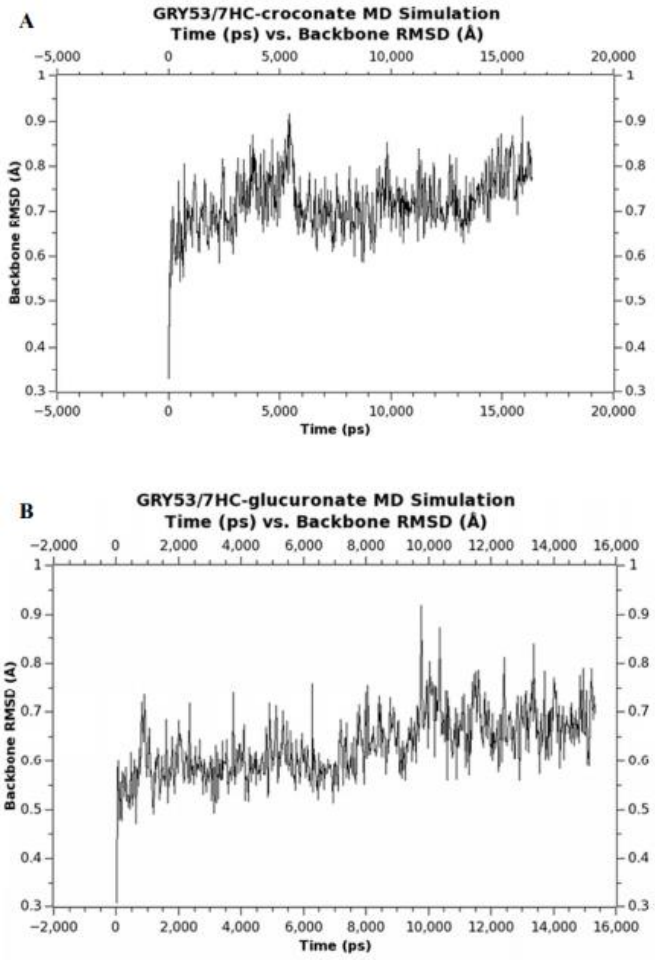

Figure 3. RMSD plots for MD simulations of (A) $GR^{Y53/7HC}$-croconate and (B) $GR^{Y53/7HC}$ - glucuronate complexes.

2.3.2 Structural Clustering of MD Snapshots

Clustering of the MD snapshots was conducted utilizing the Ensemble Cluster tool of the UCSF Chimera package.[44] This tool applies the methodology of Kelley et al.[45] A structural ensemble is partitioned into clusters of structures of similar conformations and a single structure is chosen from each of these clusters which best typifies the

conformations of structures represented within that cluster. The first step prior to the clustering process is pairwise superposition of the ensemble of structures to generate a matrix of r.m.s. values describing the similarity between all structures. Clustering of the ensemble of structures is then carried out by applying the average linkage algorithm for hierarchical cluster analysis to the described matrix. The spread of each cluster, describing the similarity of the conformations within the cluster, is assessed throughout the clustering process. After the clustering process, the spread values of all clusters at each stage is averaged and normalized. A penalty value is also determined for every step in the cluster analysis process based on the number of clusters and the average spread at that stage. The stage of the cluster analysis from which the clusters are obtained is selected based on a defined minimum penalty value. Eigen analysis is conducted on the clusters to select a representative structure from each cluster. For our purposes, we selected to use only the major representative clustered forms which represented roughly 70% of all MD snapshots.

2.3.4 Random Structural Sampling of Unliganded GR using CONCOORD (CONstraints to COORDinates)

The CONCOORD approach is a method of generating random structures, while employing empirically determined upper and lower atomic distance constraints, yielding non-correlated protein ensembles that have compared favorably to experimental NMR structures. The Python based CONCOORD plugin for the Linux distribution of YASARA Structure 13.4.21[36] was used to calculate favorable conformational isomers of the GR enzyme, based on the method of de Groot et al.,[46] using the following specifications: van der Waals parameters = OPLS-X, maximum number of iterations/structure generated = 2500, Damping factor = 2.0. Eight CONCOORD

structures were generated, as listed in Table 2. The more structures generated by CONCOORD, the greater the structural similarity between the forms generated. Therefore, this number allowed us to sample sufficient conformational space while making efficient use of our computational time and energy. An unliganded GR snapshot that represented the lowest energy structure from the production phase of the 20 ns atomistic MD simulations (described above) was used as the starting structure for calculations of the CONCOORD ensemble.

2.3.5 Ensemble Docking of Croconate and Glucuronate

The ensemble of GR structures obtained by CONCOORD sampling (described above) was prepared for virtual docking by introducing a simulation cell centered on the catalytic cysteine residues, Cys74 and Cys185, with dimensions adjusted to encompass the entirety of the active site resulting in the following cell dimensions (x−y−z) 24 X 24 X 24 Å. The ligands were constructed and minimized in MOE v2011.10 (Chemical Computing Group)[47]and imported into YASARA Structure 12.4.1[48] for ensemble virtual docking. YASARA v12.4.1[48] employs AutoDock VINA in its docking functionality.[49] Further information regarding the details of ligand pose generation and scoring can be found in the work of Whalen et al.[19] The top ranking complexes (i.e., those complexes with the strongest predicted binding using the VINA scoring function) were then used as starting structures for molecular dynamics simulations.

2.3.6 Solvent Accessible Surface Area Calculations of $GR^{Y53/7HC}$-ligand Complexes

Solvent accessible surface areas for the 7HC ring of $GR^{Y53/7HC}$ were constructed with a solvent probe radius of 1.4 Å, and the following radii for the solute elements:, non-polar hydrogens 1.0717 Å, polar hydrogens 0.32 Å, oxygen 1.344 Å, carbon 1.8 Å,

nitrogen 1.14 Å, sulfur 2.0 Å. All surface area calculations were performed with the YASARA Structure version 13.4.21.[36]

2.3.7 Calculation of Dynamic Cross-correlation Matrix (DCCM)

The DCCM between residues i and j was calculated by dividing the dot product of any two residue displacements relative to an average structure, as described by:

$$DCCM_{i,j} = \frac{\langle \vec{d}_i \cdot \vec{d}_j \rangle}{\sqrt{\langle d_i^2 \rangle \langle d_j^2 \rangle}}$$

The value of d is the displacement of an atomic position from the ensemble average position, and the brackets represent averaging over the snapshots from the MD simulation of the $GR^{Y53/7HC}$-ligand complexes.

2.3.8 Superimposition of GR Structures and Application of the Global Distance Test (GDT)

YASARA Structure version 13.4.21[36] was used to superpose the unliganded GR crystal structure, CONCOORD structures, and representative cluster structures and apply the Global Distance Test (GDT).[50] This assessment tool evaluates the percentage of atoms of two structures that may be superposed within a defined distance. Here, we defined a cutoff value of 1 Å.

2.4 Results

2.4.1 Selection of Probe and Locus for Site-Specific Incorporation into GR

Molecular dynamics (MD) simulations on GR from previous studies have suggested particular regions of GR that undergo relatively large changes, both in terms of substrate unbinding (via steered MD studies), as well as equilibrated GR−ligand complexes that show movement relative to one another (i.e., GR complexes with

different types of active site small molecules equilibrated to distinct conformers).[16] A salient residue associated with these altered regions, which could likely be a locus of incorporation for a non-natural fluorescent amino acid, is Tyr53. Incorporation of a site-specific genetically encoded non-natural amino acid 7HC by an orthogonal tRNA/aminoacyl-tRNA synthetase pair, pioneered by Wang et al.,[20] is an ideal solution in the case of GR. This system allows targeted inclusion of a single coumarin based fluorophore with high quantum yield and high environmental sensitivity to pH and solvent polarity.[20,21] This is particularly attractive in the case of GR in that a number of MD studies have suggested that Tyr53, although not directly involved in ligand binding and not part of the active site, nevertheless has a dynamic pattern highly associated with distinctive ligand conformers. Previous MD studies have revealed that binding of a variety of ligands causes obvious conformational changes in two regions: a helix and loop/helix that form the entrance to the ligand cleft.[16] The loop/helix itself contains Tyr42, which according to *in silico* docking is seen interacting with specific ligands. The loop/helix is immediately followed by a turn and helix which contains a tyrosine residue, Tyr53. Tyr53 is distinctly outside of the active site binding cleft and does not interact directly with ligands bound to the active site. Thus, the Tyr53 mutated to the 7HC functional group should provide ligand-associated fluorescent sensitivity to changes in the local environment (and thus serve as an allosteric reporter), without sacrificing particular contacts with ligands. The placement of the 7HC moiety at the surface of the enzyme, remote from any ligand pockets, places it in a microenvironment where dielectric values are significantly larger,[51] which should restrict fluorescence changes to largely water polarization effects.

2.4.2 Site-Directed Mutagenesis and Expression Conditions

Site-directed mutagenesis was carried out to make two mutations to *B. subtilis* GR (also known as RacE), the first being the alteration of the endogenous stop codon from amber (TAG) to ochre (TAA), in order to prevent probe placement at this location and a read-through mutation. Using this altered gene, mutagenesis was again utilized to mutate the Tyr53 codon to the amber stop codon. This mutation was confirmed by sequencing. The RacE gene (*B. subtilis*) was placed in a pET15b plasmid, which codes for placement of a cleavable hexahistidine-tag at the N-terminal of the recombinant protein. The pEB-JYRS-courRS plasmid containing the non-natural tRNA and corresponding tRNA synthetase was transformed first into BL21(DE3) *E. coli* cells. Cells were made chemically competent and transformed with the pET15b plasmid. The resulting cells are resistant to both tetracycline and ampicillin.

The $GR^{Y53/7HC}$ mutant protein was expressed and purified (Figure 4). It should be noted that cell growth was attenuated upon induction with IPTG, implicating protein expression of these mutants as being toxic. The expression of both the GR pET15b and pEB-JYRS plasmids within *E. coli* BL21(DE3) appears to reduce the overall fitness of the strain, as demonstrated by a reduced growth rate in liquid culture.

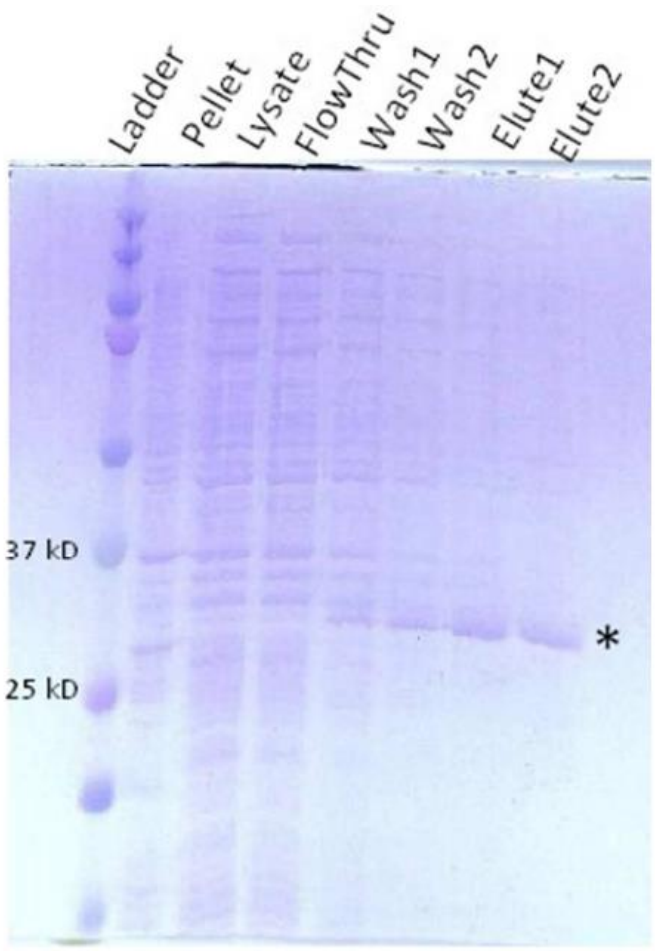

Figure 4. A representative SDS-PAGE gel is shown for $GR^{Y53/7HC}$, confirming purity and approximate molecular weight.

A circular dichroism spectrum of the wild-type and mutant protein was performed to assess proper folding (Figure 5). The two CD spectra are nearly identical, yielding the characteristic profile seen in previous studies.[17,19]

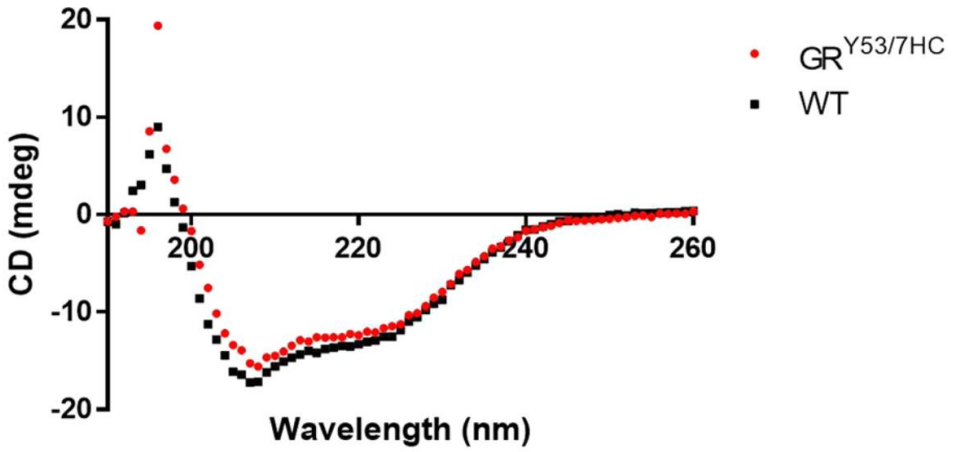

Figure 5. Circular dichroism spectra confirms retained secondary structure of equal concentrations of the $GR^{Y53/7HC}$ mutant (red) relative to WT (black).

2.4.3 Fluorescence Wavelength Profile for $GR^{Y53/7HC}$

The intrinsic fluorescence of the $GR^{Y53/7HC}$ mutant protein was compared to purified wild-type GR to confirm incorporation of the non-natural amino acid (Figure 6). Wild-type GR has no fluorescence signal when excited at 340 nm, even at substantially higher concentrations than the $GR^{Y53/7HC}$ mutant. On the contrary, the mutant has substantial fluorescence signal in this range, which increases with protein concentration. Fluorescence at 455 nm was monitored for 60 min without any notable fluctuations in signal intensity (Figure 7).

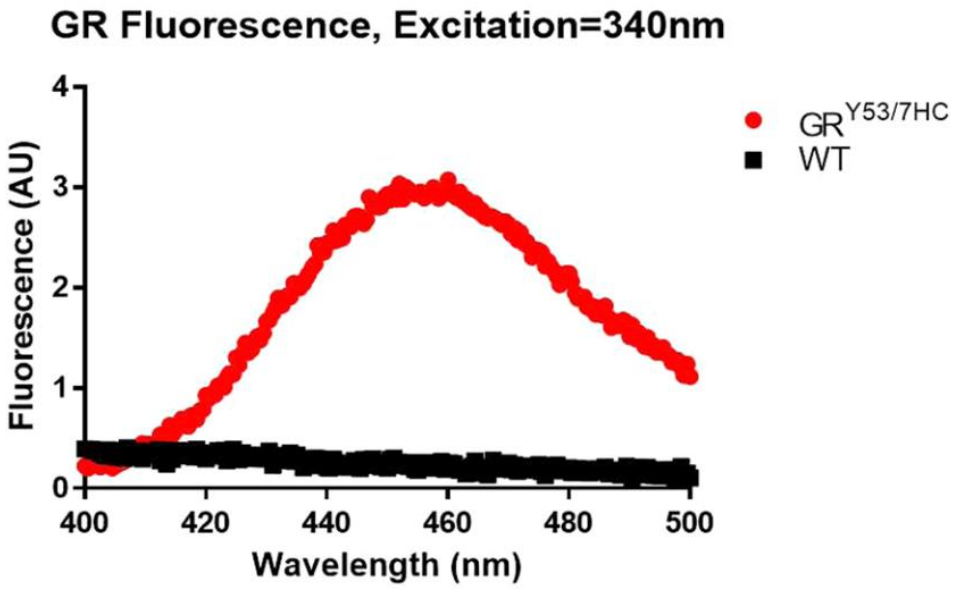

Figure 6. Fluorescence scans using an excitation wavelength of 340 nm confirms the presence of 7HC (λmax = 455 nm) in the $GR^{Y53/7HC}$ mutant (red), but not in a similar purification of wild-type GR (black). The concentration of wild-type GR is 4-fold more than that of $GR^{Y53/7HC}$.

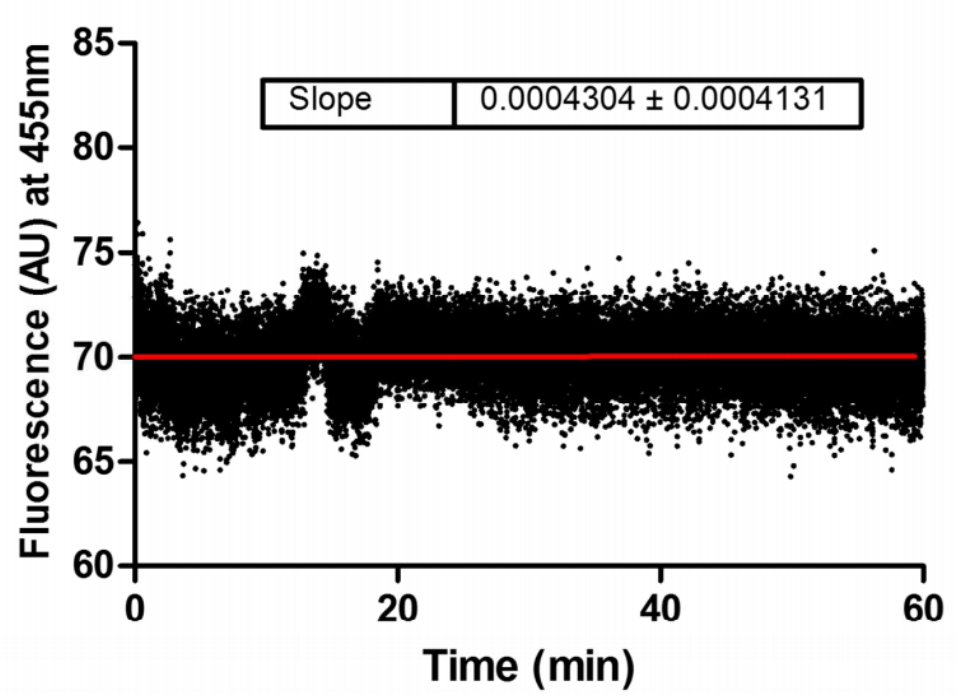

Figure 7. No fluorescent signal degradation of $GR^{Y53/7HC}$ over time.

2.4.4 $GR^{Y53/7HC}$ is an Active Racemase

To ensure the fluorescent mutant could in fact be used as a representative model of wild-type GR, activity of the $GR^{Y53/7HC}$ mutant was assessed *in vitro* using a coupled-enzyme assay measuring the turnover of L- to D-glutamate (Figure 8). The observed K_M value of the mutant was 0.27 ± 0.03 mM, within error of the published value of 0.25 mM.[17] The observed k_{cat} value was 0.031 ± 0.001 s^{-1}, a ~40-fold reduction relative to the published value of 1.3 s^{-1}.[17]

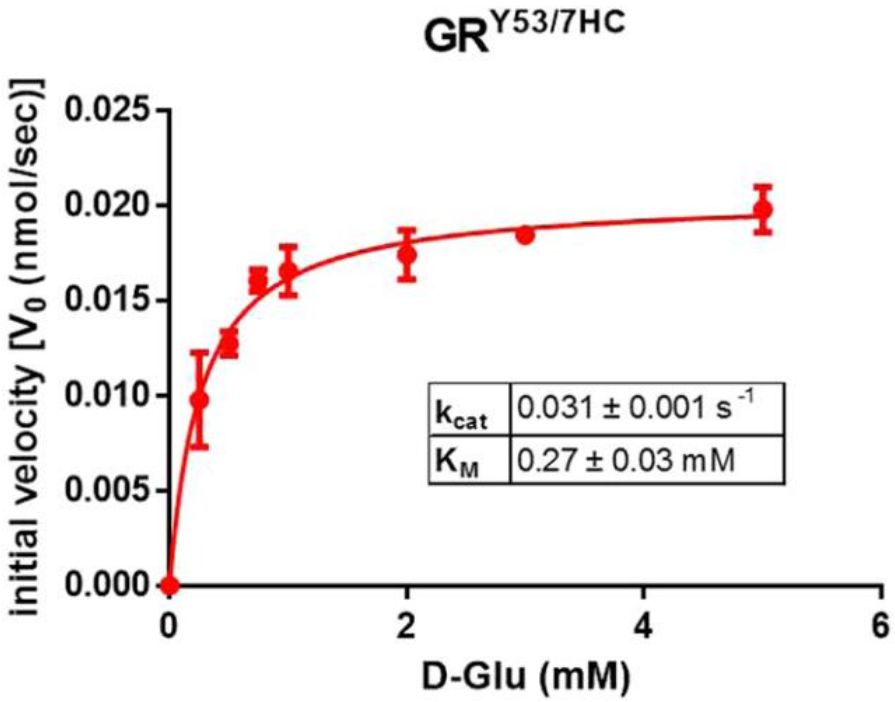

Figure 8. Racemase activity of $GR^{Y53/7HC}$ mutant has a K_M value within error of the published values (0.25 mM)[17] of wild-type GR and a k_{cat} reduced by ~40-fold relative to the published wild-type GR value (1.3 s^{-1}).[17]

2.4.5 Titration of Ligands into $GR^{Y53/7HC}$

The fluorescence wavelength profiles of both croconate and glucuronate were obtained at saturating concentrations to determine their specific contribution to fluorescence intensity at the measured wavelength. Croconate had little to no

fluorescence at a saturating concentration, while glucuronate did show minimal signal contribution at a concentration of 50 mM. Titration results are presented in Figure 9. Fitting to a one-site binding model results in K_D values that are within experimental error of K_I values determined previously via fitting of the Michaelis−Menten expression for competitive inhibition to the steady state kinetic data. This reaffirms that the fluorescent event being measured is directly related to inhibitor binding to $GR^{Y53/7HC}$. Most striking is the dramatic difference in the titration profiles between the two inhibitors. Croconate results in a quenching of the fluorescence intensity, while glucuronate binding results in fluorescence intensity enhancement. It is also encouraging that the incorporation of the probe has not altered the GR structure significantly enough to alter inhibitor binding affinities. These results provide strong experimental support for hypotheses formed from the computational studies by Whalen et al.[16] that ascribed dramatically different receptor conformations to the croconate and glucuronate-bound GR. Importantly, it is also known that both of these inhibitors are competitive with the glutamate substrate, implicating distinct allosteric linkages between the Y53 position and the small buried active site of GR.

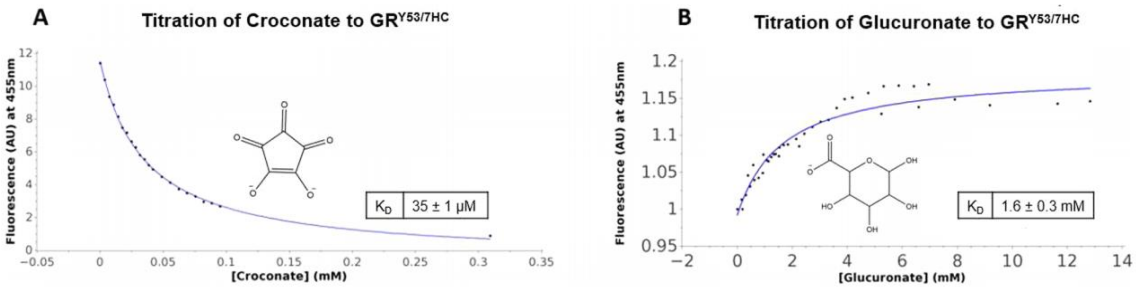

Figure 9. Binding of (A) croconate and (B) glucuronate to $GR^{Y53/7HC}$ elicits distinct changes in the fluorescent signature of $GR^{Y53/7HC}$. The fluorescent titrations were fit to a one-site binding equation giving a K_D of 35 ± 1 μM and 1.6 ± 0.3 mM for croconate and glucuronate, respectively. In both cases, the K_D is within error of the K_I, 42 ± 10 μM for croconate[18] and 1.5 ± 0.3 mM for glucuronate. The similarity of the K_I and K_D values suggests that the fluorescent titration is reporting on actual binding events.

Given the properties of the fluorescent probe, changes to fluorescence intensity can be explained in several ways: conformational changes associated with ligand binding to $GR^{Y53/7HC}$ may be altering solvent exposure of the 7HC ring; significant changes in the microenvironment of the 7HC ring may be perturbing the pK_a of the 7HC ring hydroxyl; and/or there may be changes in a simple stacking interaction with several local Tyr residues. In order to examine which of these scenarios is most likely, we employed a computational study on the ligand-$GR^{Y53/7HC}$ systems, as described below.

2.4.6 Study of $GR^{Y53/7HC}$-Ligand Complexes in silico

In order to gain insight into the types of ligand-associated conformational changes that are responsible for the observed *in vitro* differential fluorescence pattern outlined above, an MD and ensemble docking based *in silico* study of croconate- and glucuronate-bound $GR^{Y53/7HC}$ was performed. Figure 10 provides a broad overview of the computational workflow.

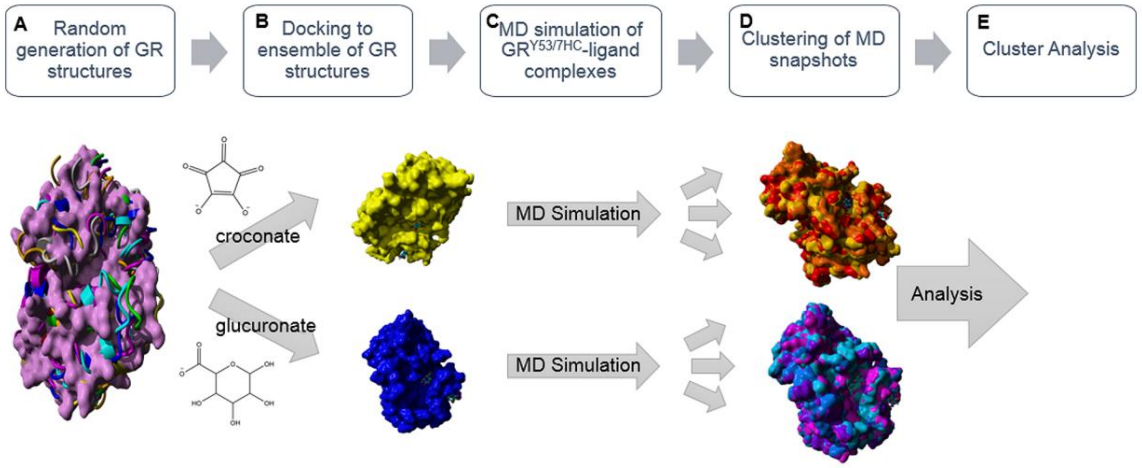

Figure 10. Computational workflow used to rationalize the differential fluorescence pattern of croconate and glucuronate binding to GR. (A) A ~20 ns MD simulation using the YAMBER3 force field,[37] a derivative of AMBER,[41] was performed on the unliganded form of GR to relax the structure. The low energy structure from the MD simulation was then subjected to the CONCOORD algorithm, as described by de Groot et al.[46] to sample conformational space of the GR structure. (B) Croconate and glucuronate were docked to an ensemble of the original unliganded GR crystal structure and eight CONCOORD generated structures using AutoDock VINA.[49] These two ligands selected distinctly different conformations for the top-docked form. (C) The 7HC ring was built into the croconate and glucuronate associated GR structures to generate the GR$^{Y53/7HC}$ mutant *in silico*. A ~15 ns MD simulation using the YAMBER3 force field[37] was then performed on each of the GR$^{Y53/7HC}$–ligand complexes. (D) Clustering of the MD snapshots was performed using the method of Pettersen et al,[44] which generated representative forms of the enzyme−ligand complexes from the MD simulation. (E) The low energy structure from the MD simulation, the time averaged structure from the MD simulation, and top representative clustered forms of the MD snapshots were used in a variety of surface area analyses. Molecular graphics created with YASARA (www.yasara.org) and POVRay (www.povray.org).

The unliganded GR crystal structure was relaxed in a ~20 ns atomistic MD simulation using the YAMBER3 knowledge based force field,[37] employing explicit solvent and physiological conditions. The low energy structure from the production portion of the simulation was used as a seed structure for the CONCOORD (from CONstraints to COORDinates) method of de Groot et al.,[46] in which an empirical restraint system is used to attempt to capture the few dominant motions that govern a protein's flexibility (a simplified form of essential dynamics, ED, which is used to

calculate eigenvectors of the covariance matrix of atomic fluctuations that occur in long atomistic MD simulations). The CONCOORD approach is extremely computationally efficient and has yielded results comparable to long MD simulations and experimental NMR data on a number of proteins.[46] One of the advantages of this approach, especially for ensemble docking, is that a covariance analysis of randomly generated structures, which satisfy a set of distance constraints, is employed. We elected to generate eight mutually distinct CONCOORD structures using this utility (henceforth referred to as 'CONCOORD 1−8'). We employed the Global Distance Test (GDT) as a metric for assessing structural uniqueness, which is summarized in the matrix in Table 2.

Table 2. Matrix of global distance test (GDT) results showing % similarity of the unliganded GR crystal structure and CONCOORD generated structures using a 1 Å cutoff. [a]The crystal structure (PDB 1ZUW)[11] of GR bound to D-glutamate was used as the initial structure. D-glutamate was deleted from the structure, and the unliganded GR form was subjected to a 20 ns MD simulation. The low energy structure from the production phase of the MD simulation was used as the seed structure for the CONCOORD method of structural sampling. The aggregate of the structures in Table 2 represent a crude model for the apo GR ensemble. YASARA's CONCOORD utility was used to generate structures which sample the conformational space of the GR structure.[46] A GDT using a 1 Å cutoff was then performed on a structural superpose of the ensemble of the unliganded GR crystal structure and the CONCOORD generated structures. The GDT assesses the quality of superpose of two structures within a defined distance, providing the percentage of superposed atoms.

structure	unliganded GR[a]	CONCOORD 1	CONCOORD 2	CONCOORD 3	CONCOORD 4	CONCOORD 5	CONCOORD 6	CONCOORD 7	CONCOORD 8
unliganded GR[a]	100	28.15	29.577	20.079	17.667	26.427	19.193	16.683	23.228
CONCOORD 1	28.15	100	20.472	16.093	32.677	14.124	14.665	17.224	20.276
CONCOORD 2	29.577	20.471	100	23.278	21.555	28.593	11.86	10.433	22.195
CONCOORD 3	20.079	16.093	23.278	100	12.894	10.531	17.372	13.829	13.484
CONCOORD 4	17.667	32.677	21.555	12.894	100	17.52	8.563	14.075	16.978
CONCOORD 5	26.427	14.124	28.593	10.531	17.52	100	11.811	7.136	33.612
CONCOORD 6	19.193	14.665	11.86	17.372	8.563	11.811	100	14.764	10.974
CONCOORD 7	16.683	17.224	10.433	13.829	14.075	7.136	14.764	100	11.614
CONCOORD 8	23.228	20.276	22.195	13.484	16.978	33.612	10.974	11.614	100

Next, an ensemble docking screen of croconate and glucuronate against the unliganded GR low energy structure of the 20 ns MD simulation (the seed structure used in the CONCOORD calculations) and the eight CONCOORD generated structures was performed using AutoDock VINA.[49] As previously discussed, croconate and glucuronate

were found to select unique forms of GR for their top-docked pose, suggesting these small molecule scaffolds may be selecting distinct forms of GR in solution.[16] Glucuronate prefers a conformer with a more solvent-exposed and opened active site cleft, which is unsurprising given the previous studies of Whalen et al. using the FERM-SMD approach.[16] On the other hand, croconate is predicted to have the most binding affinity to a conformer with a more closed active site cleft, with less solvent exposure, as well as a smaller active site volume (*vide infra*). Again this is consistent with previous computational studies.[16,18]

The GR−ligand complexes were then used as starting points for 15 ns of production atomistic MD simulations of $GR^{Y53/7HC}$–ligand complexes in physiological conditions, using the YAMBER3 force field.[37] Note that the 7HC moiety was built into the respective structures and subjected to a refinement before the production phase of the respective simulations. A variety of structures from the respective trajectories, including time-averaged, low energy, and major forms from a clustering protocol (henceforth referred to as 'croconate 1−3' and 'glucuronate 1−4' for representative clustered forms of croconate- and glucoronate-bound GR, respectively), were all subjected to further analysis, which is summarized below. The clustering protocol that was employed, the method of Pettersen et al.,[44] is based on the approach of Kelley et al.[45] The resulting $GR^{Y53/7HC}$–ligand representative clustered structures showed sufficient structural variation among themselves, based on the matrix of their respective GDT values, that all were subjected to further analysis (Table 3). Through visual inspection, surface area analyses, and analysis of correlated motions of the 7HC ring structure with residues

throughout the protein, we studied ligand-associated binding events to gain insight into the varied fluorescence pattern upon croconate and glucuronate binding, respectively.

Table 3. Matrix of global distance test (GDT) results showing % similarity of representative clustered forms of enzyme−ligand complexes from MD simulation using a 1 Å cutoff. [a]Representative clustered forms of $GR^{Y53/7HC}$–ligand complexes. Clustering of the MD snapshots was performed based on the method of Pettersen et al.,[44] which generated representative forms of the enzyme−ligand complexes from the MD simulations. A GDT using a 1 Å cutoff was then performed from a structural superpose of the top representative enzyme-croconate and enzyme-glucuronate clustered forms.

representative structure	glucuronate 1[a]	glucuronate 2[a]	glucuronate 3[a]	glucuronate 4[a]
croconate 1[a]	69.721	69.206	69.598	70.823
croconate 2[a]	67.32	64.919	66.438	69.255
croconate 3[a]	66.291	62.788	69.892	66.879

2.4.7 Ligand-Binding Associated Structural Changes

Analysis of the clustered, time-averaged, and low energy structures of the respective $GR^{Y53/7HC}$–ligand complexes all show the general trend of a greater volume active site in the glucuronate complex and a smaller volume active site in the croconate complex (Table 4). These results demonstrate consistency between this work and previous studies in which glucuronate was shown to prefer a more open, solvent-exposed active site, while croconate was shown to prefer a more closed, less solvent-exposed active site.[16,18]

Table 4. Volume of active site (Å^3) of key GR$^{Y53/7HC}$– ligand structures. [a]Low energy structure from the MD simulation. [b]Time averaged structure from the MD simulation. [c]Representative clustered forms of GR$^{Y53/7HC}$–ligand complexes. The smaller volume of the GR$^{Y53/7HC}$– croconate complexes suggests a more closed active site, while the greater volume of the GR$^{Y53/7HC}$–glucuronate complexes suggests a more open active site.

GR$^{Y53/7HC}$–croconate complexes		GR$^{Y53/7HC}$–glucuronate complexes	
structure	volume of active site (Å^3)	structure	volume of active site (Å^3)
LowE[a]	124	LowE[a]	142
time avg[b]	104	time avg[b]	241
croconate 1[c]	99	glucuronate 1[c]	154
croconate 2[c]	96	glucuronate 2[c]	124
croconate 3[c]	135	glucuronate 3[c]	198
		glucuronate 4[c]	101

There was no substantial difference in interaction of the 7HC ring with any other neighboring Tyr residues between the GR$^{Y53/7HC}$–croconate and GR$^{Y53/7HC}$–glucuronate complexes. Therefore, the differential fluorescence pattern exhibited in the *in vitro* studies cannot be explained by a simple stacking interaction with nearby Tyr residues. Additionally, no significant changes within the local environment of the 7HC moiety seemed to be occurring that may have altered the pK_a of the 7HC hydroxyl group. There have been several reported cases where pK_a shifts in a non-natural 7HC moiety are ascribed as partial contributors to the fluorescence changes upon binding of a ligand.[22,23] However, in both of these cases, the 7HC moiety was placed in buried regions proximal to the binding of highly negatively charged ligands, which is not analogous to the environment of the probe in the GR$^{Y53/7HC}$ system. However, a difference in solvent exposure of the 7HC ring between the croconate and glucuronate complexes was observed. Therefore, it is most probable that this is the cause of the differential fluorescence pattern exhibited in the *in vitro* studies. Surface area analysis of the 7HC

ring in the $GR^{Y53/7HC}$–croconate and $GR^{Y53/7HC}$–glucuronate complexes showed that the 7HC ring is more solvent exposed in the $GR^{Y53/7HC}$–croconate complex than in the $GR^{Y53/7HC}$– glucuronate complex (Table 5 and Figure 11). That this trend was evident looking at the low energy and time averaged structures from the MD as well as the representative cluster structures strongly suggests that this represents a real difference between $GR^{Y53/7HC}$–croconate and $GR^{Y53/7HC}$-glucuronate. The increased solvent exposure of the 7HC ring in the $GR^{Y53/7HC}$– croconate complex and relative decreased solvent exposure of the 7HC ring in the $GR^{Y53/7HC}$–glucuronate complex is consistent with the experimental results, which show fluorescence quenching due to complexation with croconate and an increase in fluorescence intensity due to complexation with glucuronate. It is not possible to predict the magnitude of the fluorescence changes that would result from the current ensemble using molecular mechanics methods, as performed in the current study. Nevertheless, there is a consensus among the representative forms of the individual trajectories, low energy, time averaged, and structurally clustered snapshots, that the $GR^{Y53/7HC}$–croconate complex has a 7HC moiety that is more solvent exposed than the $GR^{Y53/7HC}$–glucuronate 7HC moiety. The magnitude of change in the solvent accessible surface area of the 7HC ring between the two complexes (ranging from 70 to 107 $Å^2$) represents a change that is ~13% of the total potential solvent accessible surface of the 7HC ring. Although it is challenging to extrapolate the changes observed in the $GR^{Y53/7HC}$ system to early model studies that were performed on free 7HC (due to the uncertainty of knowing the dielectric of the protein environment), the work of Zinsli,[21] in which the effect of fractional solutions of ethanol and water on 7HC fluorescence was studied, showed very large changes in the

fluorescence when going from no water to 14% mole fraction of water (nearly half of the fluorescence signal was lost).

Table 5. Solvent-accessible surface area (Å^2) of 7HC residue of key GR$^{Y53/7HC}$–ligand structures. [a]Low energy structure from the MD simulation. [b]Time-averaged structure from the MD simulation. [c]Representative clustered forms of GR$^{Y53/7HC}$–ligand complexes. Greater solvent exposure of the 7HC moiety is observed in the GR$^{Y53/7HC}$–croconate complexes than in the GR$^{Y53/7HC}$–glucuronate complexes. The magnitude of the change in solvent accessible surface area, ranging from 70 to 107 Å^2 ,represents ~13% of the total potential solvent accessible surface of the 7HC ring system.

GR$^{Y53/7HC}$–croconate complexes		GR$^{Y53/7HC}$–glucuronate complexes	
structure	7HC solvent accessible surface area (Å^2)	structure	7HC solvent accessible surface area (Å^2)
LowE[a]	91.68	LowE[a]	76.08
time avg[b]	97.43	time avg[b]	70.34
croconate 1[c]	104.5	glucuronate 1[c]	76.42
croconate 2[c]	107.08	glucuronate 2[c]	76.36
croconate 3[c]	107.27	glucuronate 3[c]	80.06
		glucuronate 4[c]	92.62

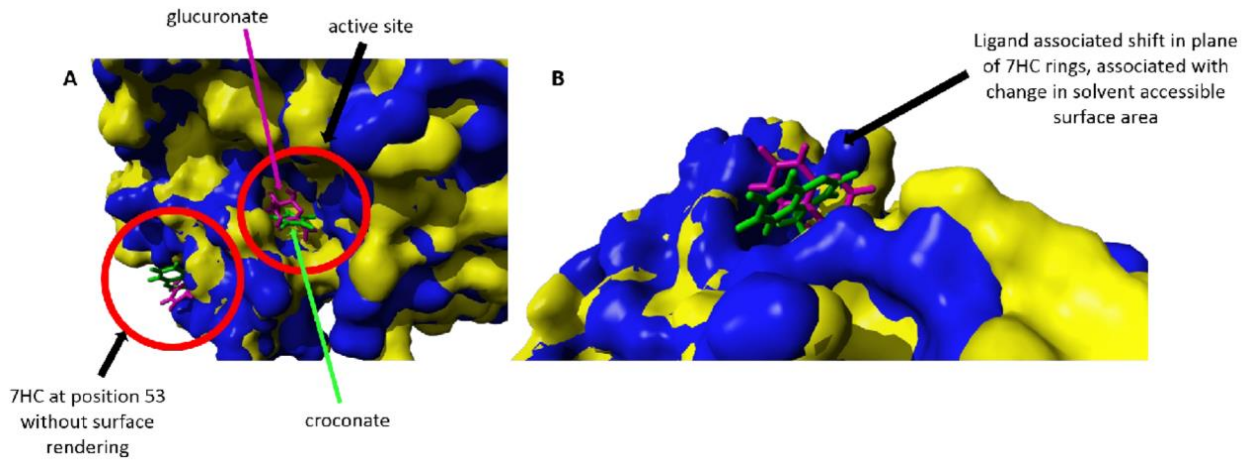

Figure 11. Superpose of most representative clustered $GR^{Y53/7HC}$ structures selected by croconate (solvent accessible surface area of this complex is indicated by yellow surface; croconate and 7HC moiety of this complex is indicated by green stick structures) and glucuronate (solvent accessible surface area of this complex is indicated by blue surface; glucuronate and 7HC moiety of this complex is indicated by magenta stick structures). (A) Conformational changes at the active site correspond to changes at the 7HC position. The closing of the active site in the $GR^{Y53/7HC}$–croconate complex corresponds to greater solvent accessible surface area of the 7HC ring, while opening of the active site in the $GR^{Y53/7HC}$–glucuronate complex corresponds to reduced solvent accessible surface area of the 7HC ring. (B) Ligand associated shift in plane of 7HC rings, concomitant with changes in solvent associated surface areas. Molecular graphics created with YASARA (www.yasara.org) and POVRay (www.povray.org).

2.4.8 Correlated Movements of the 7HC Ring with Helix System (res 74-87) Behind the Active Site

In order to investigate the dynamic relationship between the Y53/7HC position and the rest of the enzyme, especially the region of the active site, a dynamic cross-correlation matrix was calculated (DCCM) for the $GR^{Y53/7HC}$–croconate and $GR^{Y53/7HC}$–glucuronate complexes, which is represented graphically and numerically in Figure 12 and Table 6, respectively. This matrix indicates how movements of all residue pairs correlate. These values are normalized and range from −1 (perfectly anticorrelated, graphically represented by shades of blue) to +1 (perfectly correlated, graphically represented by shades of yellow). The values on the diagonal must be +1 (because the

motion of any atom is perfectly autocorrelated). Figure 12 shows regions of strong cross-correlation of residues highlighted within the white box, which links residues 50−60 (which comprise the helix on which the 7HC ring is located) to residues 74−87 (which comprise two helices connected by a short loop region located behind the croconate/glucuronate binding site). Table 6 lists the normalized cross-correlation values between the position with the 7HC ring and this key region. Behind the active site, there is a helix (residues 74−79) with which both croconate and glucuronate interact. This helix is connected to a second helix (residues 82−87) by a short loop region. The environment surrounding the 7HC ring is formed in part by this second helix (residues 82−87). Visual inspection suggested that the interaction of croconate and glucuronate with the first helix causes significant movement of the second helix, altering the 7HC environment and the solvent exposure of the 7HC ring between the $GR^{Y53/7HC}$–croconate and $GR^{Y53/7HC}$–glucuronate complexes, potentially contributing to the fluorescence changes observed experimentally. Figure 13 depicts the proposed interaction profile. These data provide a structural and dynamic link to the observed changes seen in the experimental fluorescence titration data described above.

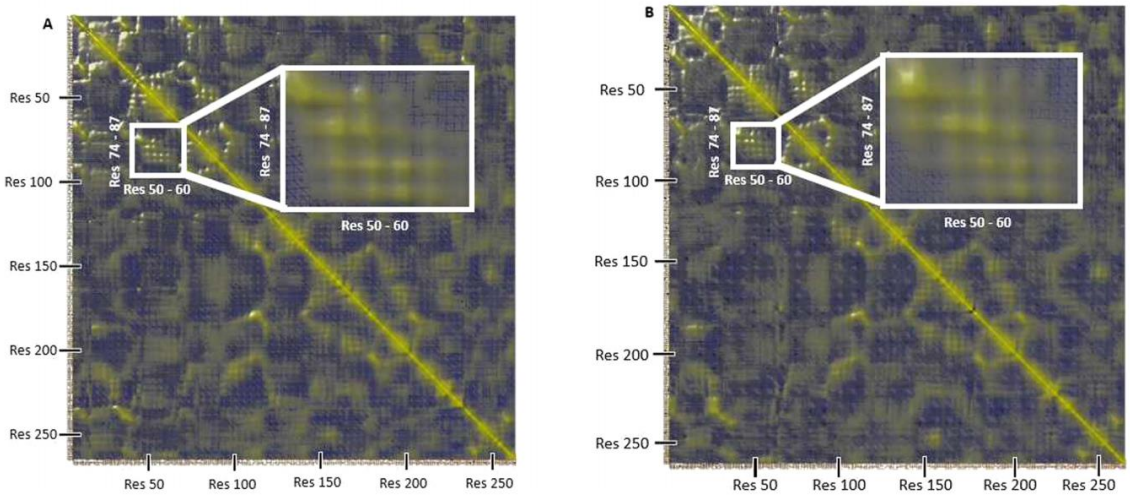

Figure 12. DCCM of (A) $GR^{Y53/7HC}$–croconate and (B) $GR^{Y53/7HC}$–glucuronate. Yellow regions are indicative of positively correlated residue movements, while blue regions are indicative of anticorrelated residue movements. Correlated movements between residues 50−60 (which comprise the helix on which the 7HC is located) to residues 74−87 (which comprise two helices connected by a short loop region located behind the ligand binding site) is shown by the raised yellow regions highlighted by the white boxes. Table 6 provides the normalized values for the DCCM. Molecular graphics created with YASARA (www.yasara.org) and POVRay (www.povray.org).

Table 6. Normalized Cross-Correlation Values between the 7HC Ring and Selected Regions of $GR^{Y53/7HC}$–Ligand Complexes. The normalized values of the DCCM are given where a value of −1 indicates residue movements are perfectly anticorrelated, and a value of +1 indicates residue movements are perfectly correlated. A strong correlation is seen in the movement of the 7HC ring with helix 1 (residues 74−79) and helix 2 (residues 82−87) in both the $GR^{Y53/7HC}$–croconate and $GR^{Y53/7HC}$–glucuronate complexes. There is no statistically significant correlation in the movement of the 7HC ring with all residues, as expected.

$GR^{Y53/7HC}$–croconate complex		$GR^{Y53/7HC}$–glucuronate complex	
region	averaged correlated movement with 7HC	region	averaged correlated movement with 7HC
helix 1 (Res 74–79)	0.10 ± 0.06	helix 1 (Res 74–79)	0.13 ± 0.04
helix 2 (Res 82–87)	0.2 ± 0.1	helix 2 (Res 82–87)	0.2 ± 0.1
entire protein (Res 1–262)	0.0 ± 0.2	entire protein (Res 1–262)	0.0 ± 0.2

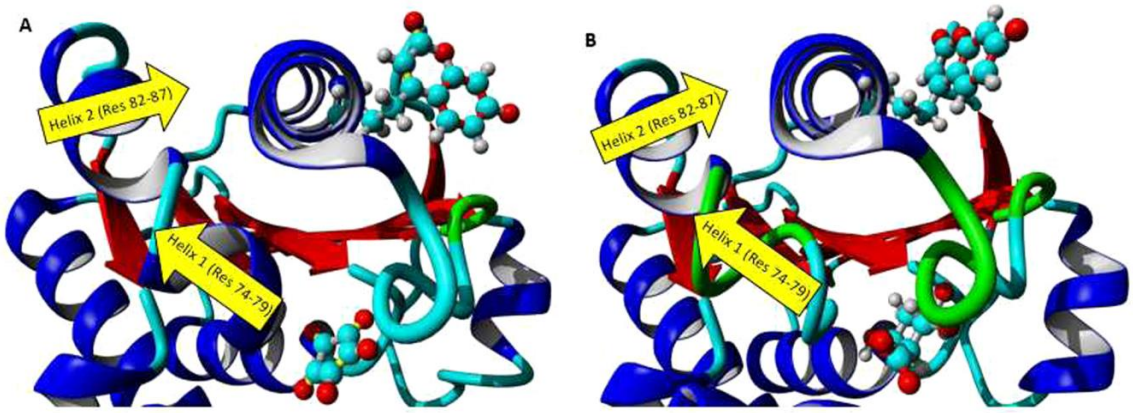

Figure 13. Results of the DCCM suggest that interaction of croconate/glucuronate with helix 1 (residues 74−79) alters the movement of helix 2 (residues 82−87), which in turn causes significant movement of the helix possessing the 7HC ring and alters its environment in the (A) $GR^{Y53/7HC}$ croconate and (B) $GR^{Y53/7HC}$–glucuronate complexes. Molecular graphics created with YASARA (www.yasara.org) and POVRay (www.povray.org).

2.5 Conclusion

Previous MD simulation studies on GR have implicated the region in the helix comprised of residues 50−60 as undergoing large movements, depending on the type of ligand bound to the active site of GR, and thus represented a reasonable position for incorporation of the fluorescent reporter moiety L-(7- hydroxycoumarin-4-yl) ethylglycine by use of a mutant orthogonal tRNA/aminoacyl-tRNA synthetase approach developed by Wang et al.[20] $GR^{Y53/7HC}$, an active glutamate racemase with the 7HC moiety inserted into the Y53 position, was subsequently biosynthesized and utilized as an experimental probe to study conformational changes occurring upon ligand binding. Although the $GR^{Y53/7HC}$ mutant has maintained the same K_M value as the wild-type GR, it exhibits a ~40 fold decrease in k_{cat}. This represents a ~2.2 kcal/mol increase in the kinetic barrier for racemization in the D → L direction. It is not possible to determine the source of this kinetic difference from the current MD studies. The small magnitude of

this change may be due to any number of factors, such as an additional hydrogen bond in the ground state of the $GR^{Y53/7HC}$–substrate complex, subtle repositioning of the catalytic base, or perhaps dynamic events in the domain opening that leads to product release. A complicating factor in addressing such subtle kinetic changes is that it is reasonable that product release is at least partially rate determining, based on the primary hydrogen substrate kinetic isotope effects performed on the related *L. bacillus* GR.[1]

A significant body of structural, biochemical, and computational work on GR has characterized it as a very plastic enzyme, which has been necessary to rationalize both its structure, function, and nature of inhibition by small molecules.[11,15,17-19] Nevertheless, direct experimental feedback about the dynamics and selectivity of allostery employed by this enzyme has been difficult to obtain. A major application of $GR^{Y53/7HC}$ is to examine how different competitive inhibitor chemotypes, which remarkably share the same active site pocket, seem to be associated with different conformations of the enzyme. The results from this study provide strong support for the hypotheses from earlier steered MD simulations that point toward very large changes in enzyme structure upon binding and unbinding of ligands to the active site.[16] From a ligand- or drug-design viewpoint, one is not dealing with a small variability in the "druggability" of a single pocket, but rather an ensemble-docking of radically different active sites, with a range of druggabilities. Indeed, accurately predicting the binding affinities for these two compounds, croconate and glucuronate, involved assessing not just a range of structures from steered MD simulations of GR, but also examining relative changes in the protein solvation energy of the target.[16] The current studies with $GR^{Y53/7HC}$ complexation with these two competitive inhibitors showed that croconate binding results in a 57% reduction in the volume of the,

already very buried, active site cavity; this may be an important contributor to its significantly higher ligand efficiency (−0.6 kcal/mol for croconate vs −0.3 kcal/mol for glucuronate). Indeed, recent free energy of binding calculations, using the extended linear response (ELR) method, performed on a series of GR competitive inhibitors showed that the collection of polar ligands, which are preferred by GR, necessitate either a very narrowly tailored van der Waals contact surface or the ability to form a network of interstitial waters deep within the active site.[34] In addition to glucuronate, GR binds weakly to a number of other carboxylate-containing ligands, which tend to remain partially solvated within the active site, and lack the enhanced van der Waals contribution observed for croconate binding.[34] The two ligands examined in detail in the current study are archetypal scaffolds selected based on a previous extensive steered MD and binding free energy study,[16] which encompassed 17 different GR inhibitors. Indeed, these two scaffolds were selected because they typified radically different ligand-associated complexation events with the GR receptor (one with an extraordinarily good and one with a very poor ligand efficiency value). Indeed, many problematic scaffolds described in Whalen et al., 2011,[16] including glucuronate, were hypothesized to associate with conformations of GR that have much more solvated subpockets and lead to much poorer than expected binding free energies. These previous findings are supported by the differences in the ligand binding cavities of the respective $GR^{Y53/7HC}$ complexes, as described in Table 4. It will be an important next step to determine if the nature of the $GR^{Y53/7HC}$ ligand associated changes in fluorescence can be used to identify poor ligand efficiencies in future experimental fragment screening campaigns.

The extraordinary sensitivity of the 7HC moiety within $GR^{Y53/7HC}$, and its incorporation into the dynamic region, provides a valuable experimental probe, quickly identifying a more "open" and solvated form, which should be avoided for high quality complexation; indeed, use of $GR^{Y53/7HC}$ could be used in high throughput screening to quickly identify attractive candidates by examining ligand-associated changes in fluorescence. The dynamic cross-correlation matrix, calculated from the MD studies of the two respective GR-complexes, shows strong coupling between regions behind the active site (a two helix system comprised of res 74−87) and the position that incorporates the 7HC moiety. The results from our computational studies indicate that this coupled motion provides a reporter for these active site forms, which provides a nexus between ensemble structures and ligand binding preferences.

Fragment-based drug discovery has been a useful and productive new approach for exploiting novel chemical space in challenging protein and enzyme targets.[52,53] The most successful campaigns in optimizing fragment hits usually employ numerous biophysical approaches (X-ray crystallography, NMR, SPR, and ITC) to predict how a fragment's ligand efficiency (as well as other metrics, such as LogP) will be maintained or improved, while potency (i.e., molecular weight) is increased. Structural knowledge about the complexation is absolutely essential to optimization, yet the very nature of many plastic enzymes means that this comes with a host of potential difficulties, particularly related to different fragments associating with different forms of the target protein, which may or may not be amenable to crystallization or even prediction about these novel structural forms. Indeed, the results of the current study underscore an even more serious challenge: small molecules which bind to the same active site, but to

substantially different forms of the target enzyme! Furthermore, it is not at all apparent that attempting to optimize a fragment hit in the usual rational synthetic "growth" strategy would result in this same mode of selection (i.e., fragment modification may have nontrivial effects on selection of the target conformer). One may envisage the use of MD-informed 7HC-labeling strategies, as outlined in the current study, as a means of improving fragment optimization protocols in many other flexible protein and enzyme drug targets. Future studies will be focused on employing $GR^{Y53/7HC}$ in fragment-based screens, in order to determine a broader profile of compounds and their respective allosteric signatures with respect to fluorescence changes. It would be particularly interesting to make $GR^{Y53/7HC}$ analogues of *H. pylori* and *B. anthracis* GRs, as these have well characterized allosteric pockets remote from the active site. It is quite possible that the unifying theme in the allosteric inhibition of these enzymes may be accurately described by the type of extended selection model employed in these studies (i.e., the nature of noncompetitive, small molecule regulation of GR enzymes resides in an extended selection model, as presented in this study). The generation of $GR^{Y53/7HC}$ analogues for these species would yield potentially powerful insight into how small molecules associate with particular conformers of a flexible protein target. Indeed, as our understanding about the linkage between *in vitro* and *in silico* studies grows, such knowledge may allow future campaigns to presage how large libraries of fragments will dynamically interact with flexible protein targets.

3. CONCLUSION AND FUTURE WORK

In the described study, we have created a fluorescently tagged GR that allows us to overcome one of the most difficult problems associated with flexible enzyme systems – elucidating ligand-associated structural changes. Using this system, we were able to acquire vital structural information of the enzyme-ligand complexes studied, GR-croconate and GR-glucuronate. This method reports on active site changes, indicating whether a ligand has associated with a conformation with a favorable "closed" active site or an unfavorable "open" active site. Such information is vital to structure-based drug design and discovery efforts of flexible enzymes; in particular, optimization attempts using these methods.

Methods that have been applied to flexible proteins in order to acquire structural information have frequently included other fluorescence spectroscopy techniques,[54,55] NMR,[27,56] x-ray crystallography,[27,57,58] and small angle X-ray scattering (SAXS).[59] The advantages of the use of 7HC versus other spectroscopic techniques include that it may be incorporated into nearly any location on a protein, possesses a large fluorescence quantum yield, has an improved Stoke's shift, and is sensitive to solvent pH and polarity.[20,21] Compared to NMR and crystallization methods, structural information from $GR^{Y53/7HC}$ can be acquired with relative ease. It does not require time consuming residue assignment or refinement efforts. Furthermore, it can be applied to proteins of diverse size, unlike NMR, which is largely limited to proteins of 20 kDa or less. Additionally, SAXS alone only provides parameters describing general global protein flexibility,[60] whereas the $GR^{Y53/7HC}$ system specifically reports on active site changes. To acquire more detailed structural information from SAXS, it is necessary to validate the SAXS data against crystallization or NMR data.[60] This is not the case for the $GR^{Y53/7HC}$ system.

The prior work that went into building this system negates the need for validation against other models. Computational methods previously identified croconate and glucoronate as associating with unique GR conformations as well as a dynamic region of the protein that's movement correlated with active site changes.[16] We then created the $GR^{Y53/7HC}$ system and experimentally assessed the fluorescence pattern upon titration of croconate and glucoronate to $GR^{Y53/7HC}$. A differential fluorescence pattern was observed. A retrospective computational study confirmed that ligand-associated changes of the $GR^{Y53/7HC}$-croconate and $GR^{Y53/7HC}$-glucuronate complexes produced changes in the local environment of the 7HC residue responsible for the observed differential fluorescence pattern.

The $GR^{Y53/7HC}$ system may be applied to HTS studies. Titration of ligands of interest to $GR^{Y53/7HC}$ in an HTS format should readily identify whether they associate with GR conformations with more open or closed active sites. As was the case with croconate, fluorescence quenching should be observed upon binding of ligands that preferentially associate with GR complexes with smaller, less solvated active sites. Conversely, as with glucoronate, fluorescence enhancement should be observed upon binding of ligands that preferentially associate with GR complexes with larger, more solvated active sites. Such information will help to guide selection of ligands for optimization. Certainly, ligands that bind to less solvated active sites would be more desirable. Additionally, the observed fluorescence pattern upon titration of a ligand of interest would give insight into a starting structure to which to dock the ligand in a computational study and begin to optimize the hit and pick up additional protein contacts.

The $GR^{Y53/7HC}$ system is particularly useful for fragment-based drug discovery and design. This is because fragments are especially difficult to apply to structure-based drug design and discovery campaigns against flexible enzymes since they do not stabilize protein conformations well.[61] This means that growing out or cross-linking fragment hits may result in the optimized compounds associating with quite different conformations than the original hits. Consequently, it is necessary to monitor conformational changes throughout the optimization process to ensure that the conformational changes do not produce unfavorable active site changes such as increased solvation and to ensure that an accurate structure is guiding optimization efforts. This would be quite difficult to accomplish with more rigorous methods such as crystallization or NMR, but is fairly simple with our system.

The GR isoform utilized in the described study was that of *Bacillus subtilis*. However, it is likely that this methodology could successfully be applied to many GR isoforms. This is due to the fact that numerous GR isoforms exhibit considerable plasticity. For example, the MurI enzymes from *H. pylori*, *E. coli*, *E. facecalis*, *E. faecium*, and *S. aureus* all undergo a significant hinge movements in order to bind substrate to the active site[15] – similar to RacE.[11] Additionally, that allostery has shown to be significant in a number of GR isoforms also seems to indicate that the described methodology could be applied to other GR isoforms. For example, dipicolinic acid (DPA) was found to inhibit both GR isozymes from *Bacillus anthracis* (RacE1 and RacE2) through a noncompetitive model[19] and a pyrazolopyrimidinedione compound was found to inhibit Helicobacter pylori GR (MurI) through an uncompetitive model.[15] That these enzymes are susceptible to allosteric inhibition indicates that changes in the active site

are linked to regions outside the active site, the underlying principle that permitted the success of our $GR^{Y53/7HC}$ system.

Future work will be focused on attempting to apply the $GR^{Y53/7HC}$ system to GR isozymes from other bacterial species. In particular, work will be focused on applying this system to GR isozymes from pathogenic bacterial species. *H. pylori* GR seems to hold the greatest promise given that the *H. pylori* bacteria poses significant health risks (it is known to be associated with cancers of the stomach[62]) and allosteric regulation has already been confirmed in this enzyme.[15]

REFERENCES

1 Tanner, M. E., Gallo, K. A. & Knowles, J. R. Isotope effects and the identification of catalytic residues in the reaction catalyzed by glutamate racemase. *Biochemistry* **32**, 3998-4006 (1993).

2 Walsh, C. T. Enzymes in the D-alanine branch of bacterial cell wall peptidoglycan assembly. *J Biol Chem* **264**, 2393-2396 (1989).

3 Fisher, S. L. Glutamate racemase as a target for drug discovery. *Microb Biotechnol* **1**, 345-360 (2008).

4 Doublet, P., van Heijenoort, J., Bohin, J. P. & Mengin-Lecreulx, D. The murI gene of Escherichia coli is an essential gene that encodes a glutamate racemase activity. *J Bacteriol* **175**, 2970-2979 (1993).

5 Bae, T. *et al.* Staphylococcus aureus virulence genes identified by bursa aurealis mutagenesis and nematode killing. *Proc Natl Acad Sci U S A* **101**, 12312-12317 (2004).

6 Kada, S., Nanamiya, H., Kawamura, F. & Horinouchi, S. Glr, a glutamate racemase, supplies D-glutamate to both peptidoglycan synthesis and poly-gamma-glutamate production in gamma-PGA-producing Bacillus subtilis. *FEMS Microbiol Lett* **236**, 13-20 (2004).

7 Kimura, K., Tran, L. S. & Itoh, Y. Roles and regulation of the glutamate racemase isogenes, racE and yrpC, in Bacillus subtilis. *Microbiology* **150**, 2911-2920 (2004).

8 Song, J. H. *et al.* Identification of essential genes in Streptococcus pneumoniae by allelic replacement mutagenesis. *Mol Cells* **19**, 365-374 (2005).

9 Morayya, S., Awasthy, D., Yadav, R., Ambady, A. & Sharma, U. Revisiting the essentiality of glutamate racemase in Mycobacterium tuberculosis. *Gene* **555**, 269-276 (2015).

10 Li, Y. *et al.* Investigation of the essentiality of glutamate racemase in Mycobacterium smegmatis. *J Bacteriol* **196**, 4239-4244 (2014).

11 Ruzheinikov, S. N., Taal, M. A., Sedelnikova, S. E., Baker, P. J. & Rice, D. W. Substrate-induced conformational changes in Bacillus subtilis glutamate racemase and their implications for drug discovery. *Structure* **13**, 1707-1713 (2005).

12 Bugg, T. D. & Walsh, C. T. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance. *Nat Prod Rep* **9**, 199-215 (1992).

13 de Dios, A. *et al.* 4-Substituted D-glutamic acid analogues: the first potent inhibitors of glutamate racemase (MurI) enzyme with antibacterial activity. *J Med Chem* **45**, 4559-4570 (2002).

14 Whalen, K. L., Chau, A. C. & Spies, M. A. In silico optimization of a fragment-based hit yields biologically active, high-efficiency inhibitors for glutamate racemase. *ChemMedChem* **8**, 1681-1689 (2013).

15 Lundqvist, T. *et al.* Exploitation of structural and regulatory diversity in glutamate racemases. *Nature* **447**, 817-822 (2007).

16 Whalen, K. L., Chang, K. M. & Spies, M. A. Hybrid Steered Molecular Dynamics-Docking: An Efficient Solution to the Problem of Ranking Inhibitor Affinities Against a Flexible Drug Target. *Mol Inform* **30**, 459-471 (2011).

17 Spies, M. A. *et al.* Determinants of catalytic power and ligand binding in glutamate racemase. *J Am Chem Soc* **131**, 5274-5284 (2009).
18 Whalen, K. L., Pankow, K. L., Blanke, S. R. & Spies, M. A. Exploiting Enzyme Plasticity in Virtual Screening: High Efficiency Inhibitors of Glutamate Racemase. *ACS Med Chem Lett* **1**, 9-13 (2010).
19 Whalen, K. L., Tussey, K. B., Blanke, S. R. & Spies, M. A. Nature of allosteric inhibition in glutamate racemase: discovery and characterization of a cryptic inhibitory pocket using atomistic MD simulations and pKa calculations. *J Phys Chem B* **115**, 3416-3424 (2011).
20 Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid. *J Am Chem Soc* **128**, 8738-8739 (2006).
21 Zinsli, P. E. Investigation of rate parameters in chemical reactions of excited hydroxycoumarins in different solvents. *J Photochem* **3**, 55-69 (1974).
22 Mendes, K. R., Martinez, J. A. & Kantrowitz, E. R. Asymmetric allosteric signaling in aspartate transcarbamoylase. *ACS Chem Biol* **5**, 499-506 (2010).
23 Lacey, V. K. *et al.* A fluorescent reporter of the phosphorylation status of the substrate protein STAT3. *Angew Chem Int Ed Engl* **50**, 8692-8696 (2011).
24 Talukder, P. *et al.* Cyanotryptophans as Novel Fluorescent Probes for Studying Protein Conformational Changes and DNA-Protein Interaction. *Biochemistry* **54**, 7457-7469 (2015).
25 Chen, S. *et al.* Detection of dihydrofolate reductase conformational change by FRET using two fluorescent amino acids. *J Am Chem Soc* **135**, 12924-12927 (2013).
26 Kuhn, S. M., Rubini, M., Muller, M. A. & Skerra, A. Biosynthesis of a fluorescent protein with extreme pseudo-Stokes shift by introducing a genetically encoded non-natural amino acid outside the fluorophore. *J Am Chem Soc* **133**, 3708-3711 (2011).
27 Teague, S. J. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* **2**, 527-541 (2003).
28 Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471-1475 (2008).
29 Gsponer, J. *et al.* A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure* **16**, 736-746 (2008).
30 Grant, B. J., McCammon, J. A. & Gorfe, A. A. Conformational selection in G-proteins: lessons from Ras and Rho. *Biophys J* **99**, L87-89 (2010).
31 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 (2010).
32 Kobayashi, K. *et al.* Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A* **100**, 4678-4683 (2003).
33 May, M. *et al.* Structural and functional analysis of two glutamate racemase isozymes from Bacillus anthracis and implications for inhibitor design. *J Mol Biol* **371**, 1219-1237 (2007).
34 Whalen, K. L. & Spies, M. A. Flooding enzymes: quantifying the contributions of interstitial water and cavity shape to ligand binding using extended linear response free energy calculations. *J Chem Inf Model* **53**, 2349-2359 (2013).

35 Pal, M. & Bearne, S. L. Inhibition of glutamate racemase by substrate-product analogues. *Bioorg Med Chem Lett* **24**, 1432-1436 (2014).
36 YASARA (13.4.21) (YASARA Biosciences Gmbh, Vienna, Austria, 2013).
37 Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* **57**, 678-683 (2004).
38 Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys B* **103**, 8577-8593 (1995).
39 Molecular Operating Environment (2013.08) (Chemical Computing Group, Inc., Montreal, Quebec, Canada, 2013).
40 Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **23**, 1623-1641 (2002).
41 Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **117**, 5179-5197 (1995).
42 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-1174 (2004).
43 Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature* **381**, 272 (1996).
44 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
45 Kelley, L. A., Gardner, S. P. & Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng* **9**, 1063-1065 (1996).
46 de Groot, B. L. *et al.* Prediction of protein conformational freedom from distance constraints. *Proteins* **29**, 240-251 (1997).
47 Molecular Operating Environment (2011.10) (Chemical Computing Group, Montreal, Quebec, Canada, 2011).
48 YASARA (12.4.1) (YASARA Biosciences GmbH, Vienna, Austria, 2012).
49 Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-461 (2010).
50 Theobald, D. L. & Wuttke, D. S. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22**, 2171-2172 (2006).
51 Li, L., Li, C., Zhang, Z. & Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J Chem Theory Comput* **9**, 2126-2136 (2013).
52 Murray, C. W. & Blundell, T. L. Structural biology in fragment-based drug design. *Curr Opin Struct Biol* **20**, 497-507 (2010).
53 Murray, C. W., Verdonk, M. L. & Rees, D. C. Experiences in fragment-based drug discovery. *Trends Pharmacol Sci* **33**, 224-232 (2012).

54 Somogyi, B., Lakos, Z., Szarka, A. & Nyitrai, M. Protein flexibility as revealed by fluorescence resonance energy transfer: an extension of the method for systems with multiple labels. *J Photochem Photobiol B* **59**, 26-32 (2000).
55 Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838-844 (2007).
56 Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520-1523 (2002).
57 Lee, A. Y., Gulnik, S. V. & Erickson, J. W. Conformational switching in an aspartic proteinase. *Nat Struct Biol* **5**, 866-871 (1998).
58 Rahuel, J., Priestle, J. P. & Grutter, M. G. The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J Struct Biol* **107**, 227-236 (1991).
59 Lipfert, J. & Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu Rev Biophys Biomol Struct* **36**, 307-327 (2007).
60 Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett* **589**, 2570-2577 (2015).
61 Surade, S. & Blundell, T. L. Structural biology and drug discovery of difficult targets: the limits of ligandability. *Chem Biol* **19**, 42-50 (2012).
62 Cover, T. L. Helicobacter pylori Diversity and Gastric Cancer Risk. *MBio* **7**, e01869-01815 (2016).