Supporting Materials:

# Biosynthesis of a Novel Glutamate Racemase Containing a Site Specific 7-Hydroxycoumarin Amino Acid: enzyme-ligand promiscuity revealed at the atomistic level

Sondra F. Dean[1**], Katie L. Whalen[2**], and M. Ashley Spies[1,3*]

[1] Division of Medicinal and Natural Products Chemistry, College of Pharmacy
University of Iowa
Iowa City, IA 52242
[2] Department of Biochemistry
University of Illinois, Urbana-Champaign
Urbana, Illinois 61801
[3] Department of Biochemistry, Carver College of Medicine
University of Iowa
Iowa City, IA 52242

[*]To whom correspondence should be addressed.
[**]These authors have an equal contribution.
Correspondence to: M. Ashley Spies
Email: m-ashley-spies@uiowa.edu

Pages: 14

Tables: 1

Figures: 5

**SUPPORTING INFORMATION**

## Experimental Methods

### S1. 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.[1] 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 S1. *E. coli* XL1 Blue cells were transformed with PCR products.

### S2. 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 600nm 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.

**S3. 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/300-375nm; 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 S3). 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 S4).

**Materials**

L-(7-hydroxycoumarin-4-yl) ethylglycine was synthesized under contract from AsisChem Inc., using the procedure described in Wang et al.[2]

**Computational Methods**

**S4. 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,[3] using the starting structure from PDB 1ZUW (*B. subtilis* GR),[4] except for a few notable exceptions, which are summarized below. The YASARA Structure package version 13.4.21[5] 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,[3] but employing the YAMBER3 knowledge based force field,[6] which was used with long-range electrostatic potentials calculated with the Particle Mesh Ewald (PME) method,[7] with a cutoff of 7.86 Å. For $GR^{Y53/7HC}$, the Build utility of MOE v2013.08 (Chemical Computing Group)[8] was used to transform Tyr53 to the 7HC residue. The ligand and 7HC force field parameters were generated with the AutoSMILES utility,[9] 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.[10, 11] 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.[12] 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 S5 of the supporting information 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.

**S5. Structural Clustering of MD Snapshots**

Clustering of the MD snapshots was conducted utilizing the Ensemble Cluster tool of the UCSF Chimera package.[13] This tool applies the methodology of Kelley et al.[14] 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 generates 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.

**S6. 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[5] was used to calculate favorable conformational isomers of the GR enzyme, based on the method of de Groot et al.,[15] 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 1. 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.

**S7. 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)[16] and imported into YASARA Structure 12.4.1[17] for ensemble virtual docking. YASARA v12.4.1[17] employs AutoDock VINA in its docking functionality.[18] 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.

**S8. 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.[5]

**S9. 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 Equation 1:

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

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.

**S10. Superimposition of GR Structures and Application of the Global Distance Test (GDT)**

YASARA Structure version 13.4.21[5] was used to superpose the unliganded GR crystal structure, CONCOORD structures, and representative cluster structures and apply the Global Distance Test (GDT).[20] 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.

## SUPPORTING FIGURES

**Table S1.** 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' |

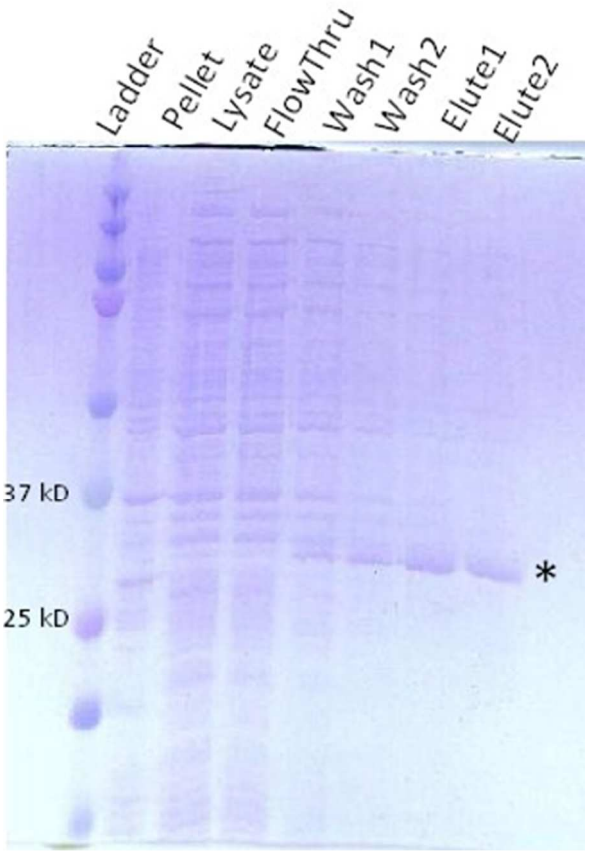


**Figure S1.** A representative SDS-PAGE gel is shown for $GR^{Y53/7HC}$, confirming purity and approximate molecular weight.

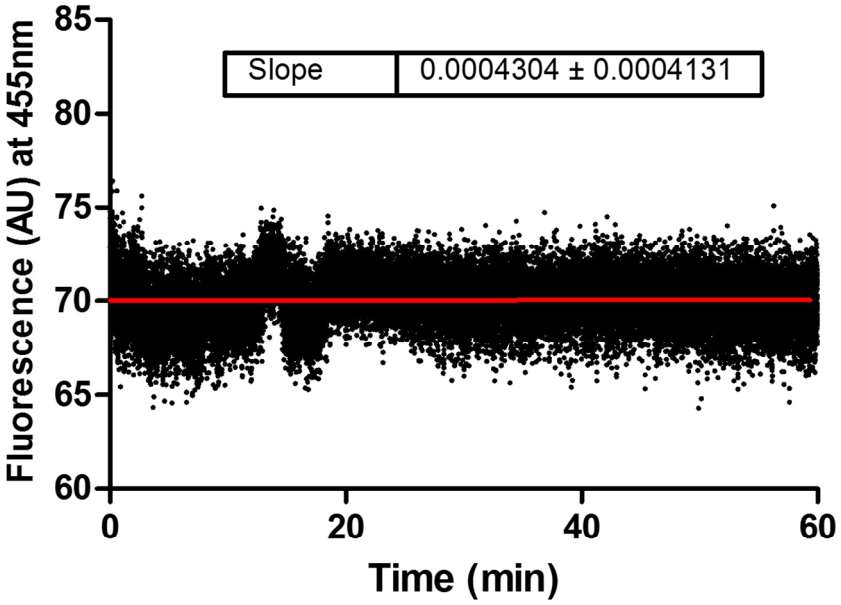


**Figure S2.** No fluorescent signal degradation of $GR^{Y53/7HC}$ over time.

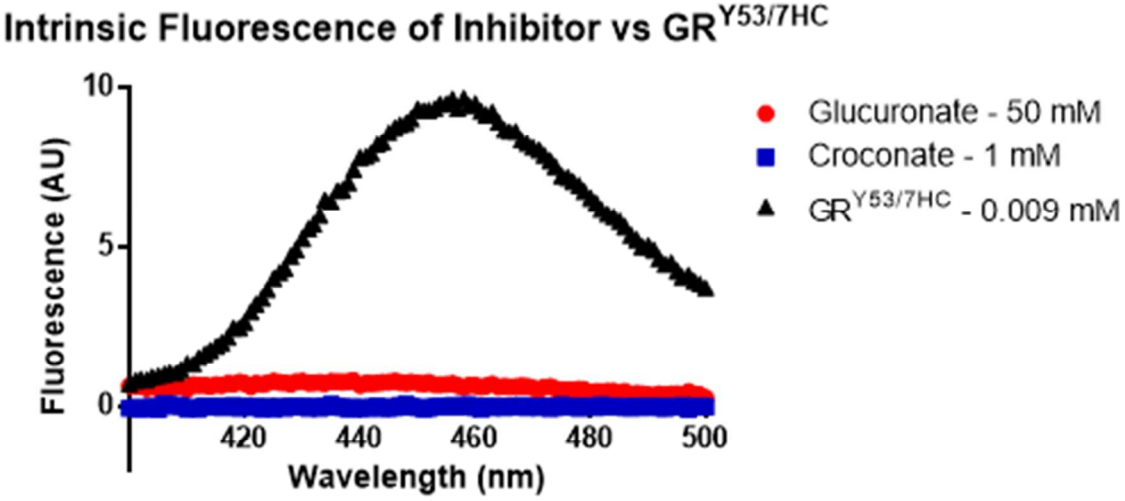


**Figure S3.** Croconate and glucuronate were scanned for emission at the excitation wavelength (340 nm) and no significant fluorescent signal was detected by either compound relative to $GR^{Y53/7HC}$.

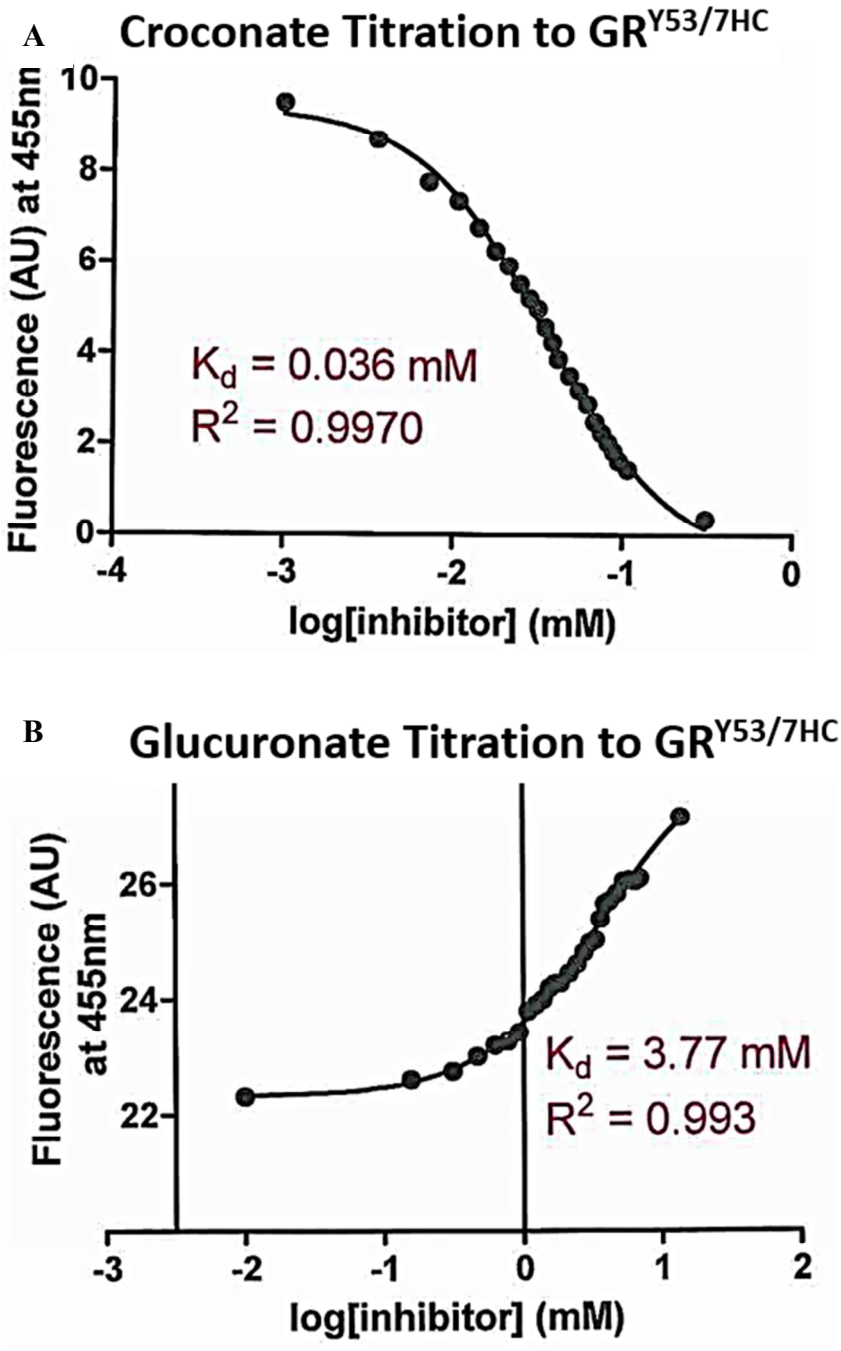


**Figure S4.** Data fitting of fluorescent titrations of (A) croconate and (B) glucuronate to $GR^{Y53/7HC}$ demonstrate Hill coefficients of unity indicating 1:1 binding stoichiometry of ligand to enzyme. Titration of croconate to $GR^{Y53/7HC}$ yielded a Hill coefficient of -1.158 ± 0.060 and titration of glucuronate to $GR^{Y53/7HC}$ yielded a Hill coefficient of 0.9649 ± 0.099.

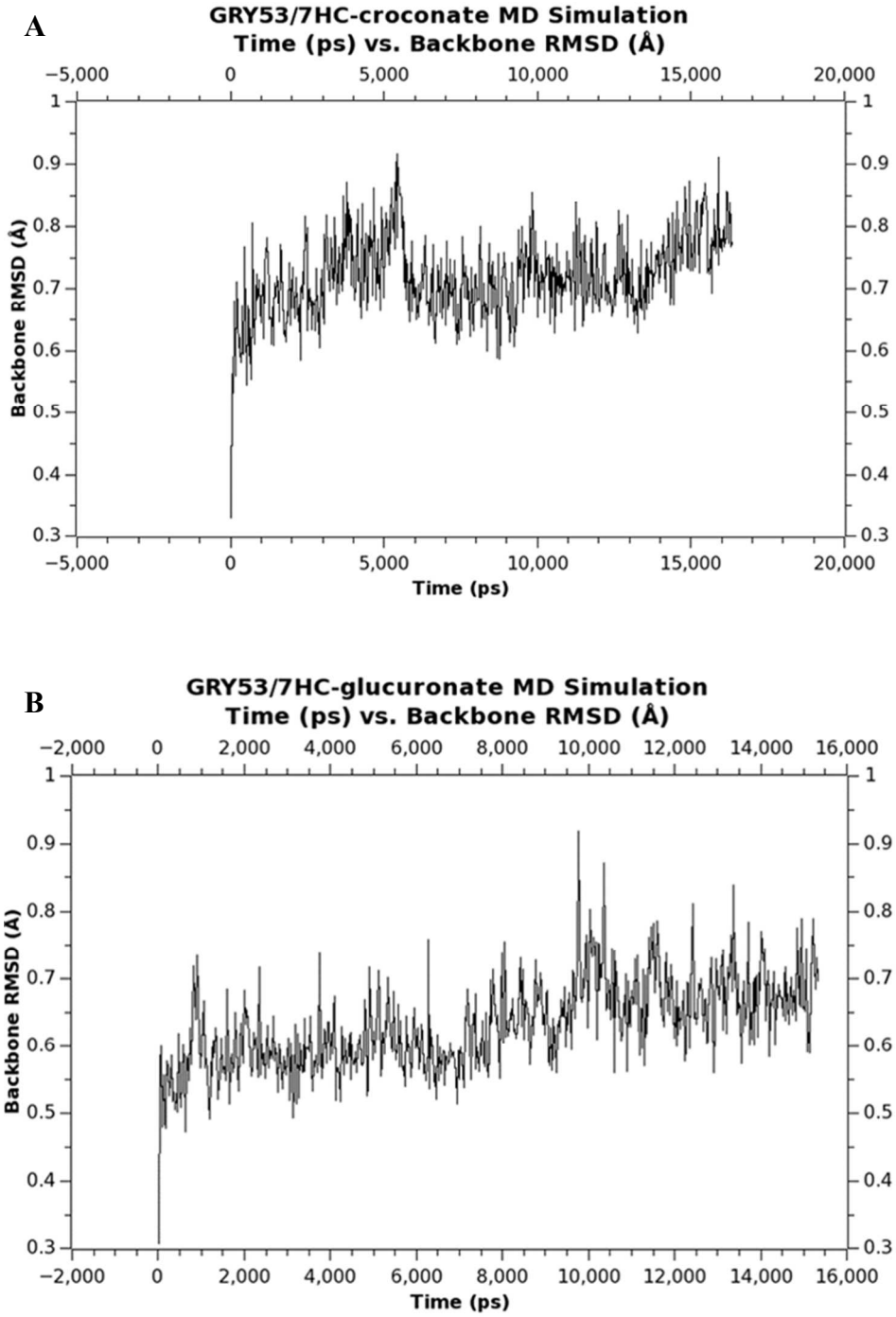


**Figure S5.** RMSD plots for MD simulations of (A) $GR^{Y53/7HC}$-croconate and (B) $GR^{Y53/7HC}$-glucuronate complexes.

**References:**


[1] Spies, M. A., Reese, J. G., Dodd, D., Pankow, K. L., Blanke, S. R., and Baudry, J. (2009) Determinants of catalytic power and ligand binding in glutamate racemase, *Journal of the American Chemical Society 131*, 5274-5284.

[2] Wang, J., Xie, J., and Schultz, P. G. (2006) A genetically encoded fluorescent amino acid, *Journal of the American Chemical Society 128*, 8738-8739.

[3] Whalen, K. L., and Spies, M. A. (2013) Flooding enzymes: quantifying the contributions of interstitial water and cavity shape to ligand binding using extended linear response free energy calculations, *Journal of chemical information and modeling 53*, 2349-2359.

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

[5] (2013) YASARA (13.4.21), YASARA Biosciences Gmbh, Vienna, Austria.

[6] Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A., and Vriend, G. (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space, *Proteins 57*, 678-683.

[7] Essman, U., Perera, L., Berkowitz, M. D., T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. B 103*, 8577-8593.

[8] (2013) Molecular Operating Environment (2013.08), Chemical Computing Group, Inc., Montreal, Quebec, Canada.

[9] Jakalian, A., Jack, D. B., and Bayly, C. I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, *J Comput Chem 23*, 1623-1641.

[10] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr., K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc. 117*, 5179-5197.

[11] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *J Comput Chem 25*, 1157-1174.

[12] Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures, *Nature 381*, 272.

[13] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—A visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.

[14] Kelley, L. A., Gardner, S. P., and Sutcliffe, M. J. (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies, *Protein engineering 9*, 1063-1065.

[15] de Groot, B. L., van Aalten, D. M., Scheek, R. M., Amadei, A., Vriend, G., and Berendsen, H. J. (1997) Prediction of protein conformational freedom from distance constraints, *Proteins 29*, 240-251.

[16] (2011) Molecular Operating Environment (2011.10), Chemical Computing Group, Montreal, Quebec, Canada.

[17] (2012) YASARA (12.4.1), YASARA Biosciences GmbH, Vienna, Austria.

[18] Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J Comput Chem 31*, 455-461.

[19] Whalen, K. L., Tussey, K. B., Blanke, S. R., and Spies, M. A. (2011) Nature of allosteric inhibition in glutamate racemase: discovery and characterization of a cryptic inhibitory pocket using atomistic MD simulations and pKa calculations, *The journal of physical chemistry. B 115*, 3416-3424.

[20] Theobald, D. L., and Wuttke, D. S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures, *Bioinformatics 22*, 2171-2172.