Supporting Information

Schwans et al. 10.1073/pnas.1206710110

SI Text

Analysis of the Conservation of Aromatic Residues at Positions Identified Having a Phe Residue Neighboring an Asp or Glu General Base. The UniProt Knowledgebase (1) was used to identify 500 similar sequences for each enzyme listed in Table S4 with an angle of interaction between 0–10°. Sequences were aligned using Clustal Omega (2, 3), and the identity of the residues at the position of the Phe residue were counted manually. The same procedure was followed using starting sequences that did not contain a Phe residue to control for bias from the starting sequences used for the database search.

Analysis of the Identity of Residues Nearby Asp and Glu General Bases Without Neighboring Hydrogen-Bonding Groups. We analyzed the structure database to determine if aromatic residues commonly residue near Asp and Glu general bases that do not have nearby hydrogen-bonding groups to help position the carboxylate group. We first used the Catalytic Site Atlas (4–6), Protein Data Bank (PDB) (http://rcsb.org/pdb), MACiE (Mechanism, Annotation, and Classification in Enzymes) (7, 8), BRaunschweig ENzyme DAtabse (BRENDA) (9, 10), and the primary literature to identify enzymes with Asp or Glu residues implicated as general bases that do not have nearby hydrogen-bonding groups (oxygen, nitrogen, and sulfur atoms not observed within 3.2 Å of the carboxylate oxygen atoms). Multiple structures of the same enzyme were identified manually and removed. We identified 72 nonredundant enzymes with Asp or Glu residues implicated as general bases that do not have nearby hydrogen-bonding groups (Table S6*A*). Structures with aromatic residues near the carboxylate general bases were determined by manually inspecting the PDB files and identifying all residues within 4 Å of the carboxylate oxygen atoms.

Our analysis revealed that 46 of the 72 structures identified have aromatic residue near the carboxylate general base. The remaining 26 structures had hydrogen bonds to crystallographically observed water molecules (20 structures), were clearly mispositioned (5 structures), or had low electron density (1 structure; Table S6*B*).

The observation of aromatic residues commonly found near carboxylate general bases without nearby hydrogen-bonding groups to position the base is consistent with the model that aromatic residues are important for helping to position carboxylate general bases in enzyme active sites. The structural analysis also revealed a preponderance of Phe residues in the subset of structures with aromatic residues near Asp and Glu general bases; Phe was the nearby aromatic residue in 40 or the 46 structures with an aromatic residue near the carboxylate general base. Future studies will be needed to investigate the possible roles of Phe versus other aromatic amino acids near carboxylate general bases.

1. Consortium TU; UniProt Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40(Database issue): D71–D75.
2. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
3. Goujon M, et al. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38(Web Server issue):W695–W699.
4. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. *J Mol Biol* 324(1):105–121.
5. Porter CT, Bartlett GJ, Thornton JM (2004) The Catalytic Site Atlas: A resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res* 32(Database issue):D129–D133.
6. Torrance JW, Bartlett GJ, Porter CT, Thornton JM (2005) Using a library of structural templates to recognise catalytic sites and explore their evolution in homologous families. *J Mol Biol* 347(3):565–581.
7. Holliday GL, et al. (2007) MACiE (Mechanism, Annotation and Classification in Enzymes): Novel tools for searching catalytic mechanisms. *Nucleic Acids Res* 35(Database issue):D515–D520.
8. Holliday GL, et al. (2005) MACiE: A database of enzyme reaction mechanisms. *Bioinformatics* 21(23):4315–4316.
9. Scheer M, et al. (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* 39(Database issue):D670–D676.
10. Schomburg I, et al. (2002) BRENDA: A resource for enzyme data and metabolic information. *Trends Biochem Sci* 27(1):54–56.

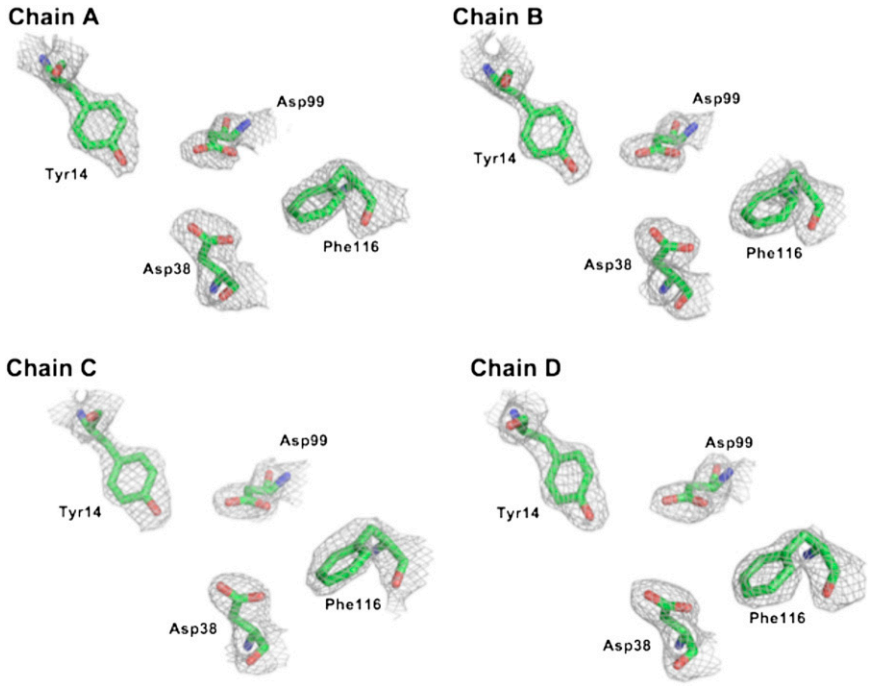

Fig. S1. $2F_o–F_c$ electron density map (contoured at 2.0σ) from the 2.5-Å resolution structure of Phe54Gly for the four independently refined chains contained in the asymmetric unit. Chain A has Asp38 indistinguishable from its position in wild-type ketosteroid isomerase (KSI), whereas the other three chains have Asp38 displaced ∼1.0 Å from the wild-type KSI position.

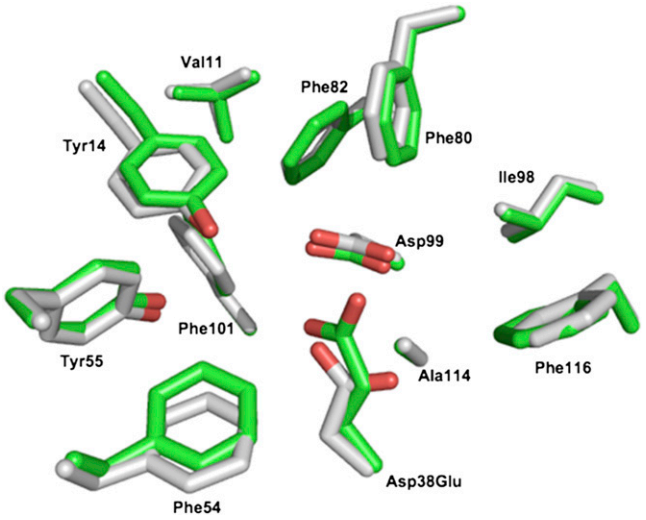

Fig. S2. Crystal structure of tKSI Asp38Glu shows that the base is mispositioned relative to wild-type tKSI. Superposition of the 2.1-Å Asp38Glu structure determined herein (PDB ID code 4L7K; carbon atoms are colored green) and the previously determined 2.3 Å wild-type KSI structure (PDB 8CHO; carbon atoms are color white). The overall root-mean-square deviation between the two structures for backbone atoms is 0.4 Å. X-ray data and refinement statistic are listed in Table S2.

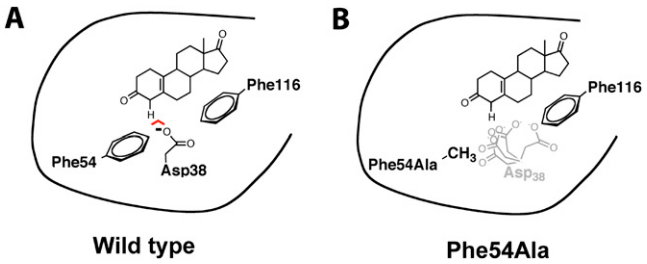

Fig. S3. Model for an increase in substrate binding for Phe54 mutants. (*A*) The anionic Asp general base is juxtaposed near to the hydrophobic steroid substrate in wild-type KSI, leading to an unfavorable interaction (red line). (*B*) Mutation of Phe54 to Ala, Gly, Val, and Ile allows the general base to access alternative conformations that reduce unfavorable interactions between the general base and steroid substrate leading to an increase in substrate binding (Fig. 3*B*).

Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)
Table S4 (DOC)
Table S5 (DOC)
Table S6 (DOC)