jmb Article

Janus: Prediction and Ranking of Mutations Required for Functional Interconversion of Enzymes

Trevor A. Addington[1]†, Robert W. Mertz[1,2]†, Justin B. Siegel[1,3,4]†, James M. Thompson[5], Andrew J. Fisher[1,6], Vladimir Filkov[7], Nicholas M. Fleischman[1], Alisa A. Suen[1], Chensong Zhang[1] and Michael D. Toney[1]

*1 - **Department of Chemistry,** University of California, Davis, CA 95616, USA*
*2 - **Biochemistry and Molecular Biology Graduate Group,** University of California, Davis, CA 95616, USA*
*3 - **Department of Biochemistry & Molecular Medicine,** University of California, Davis, CA 95616, USA*
*4 - **Genome Center,** University of California, Davis, CA 95616, USA*
*5 - **Department of Biochemistry,** University of Washington, Box 357350, 1705 NE Pacific Street, Seattle, WA 98195–7350, USA*
*6 - **Department of Molecular and Cellular Biology,** University of California, Davis, CA 95616, USA*
*7 - **Department of Computer Science,** University of California, Davis, CA 95616, USA*

***Correspondence to Michael D. Toney:** Department of Chemistry, University of California, Davis, CA 95616–3516, USA.*
mdtoney@ucdavis.edu
http://dx.doi.org/10.1016/j.jmb.2013.01.034
Edited by A. Panchenko

Abstract

Identification of residues responsible for functional specificity in enzymes is a challenging and important problem in protein chemistry. Active-site residues are generally easy to identify, but residues outside the active site are also important to catalysis and their identities and roles are more difficult to determine. We report a method based on analysis of multiple sequence alignments, embodied in our program Janus, for predicting mutations required to interconvert structurally related but functionally distinct enzymes. Conversion of aspartate aminotransferase into tyrosine aminotransferase is demonstrated and compared to previous efforts. Incorporation of 35 predicted mutations resulted in an enzyme with the desired substrate specificity but low catalytic activity. A single round of DNA back-shuffling with wild-type aspartate aminotransferase on this variant generated mutants with tyrosine aminotransferase activities better than those previously realized from rational design or directed evolution. Methods such as this, coupled with computational modeling, may prove invaluable in furthering our understanding of enzyme catalysis and engineering.

Introduction

The ability to predict which residues in a protein determine a specific function (e.g., substrate or reaction specificity, thermostability) is a long-standing and elusive problem in protein chemistry. One obstacle is that a compilation of diverse sequences, all of which have the same biological activity, includes a mixture of structural and functional information as well as evolutionary drift. For closely related enzymes of known structure, determinants of specificity can sometimes be gleaned by active-site comparisons and can be confirmed by mutagenesis. However, a growing body of evidence suggests that residues distant from the active site also contribute to functional properties.[1]

Probing distant residues for their effects on activity by site-directed mutagenesis becomes impractical when many combinations of mutations require testing. If selection or screening for activity is possible, iterative rounds of directed evolution can be employed. However, only a small fraction of the sequence space can be sampled using current directed evolution methods, and many residues that may contribute to function may well be missed.

We report a bioinformatics method (named "Janus") to identify and rank amino acid residues that specify differences in functionality between two structurally similar proteins. Janus combines analyses of structural conservation and mutation correlations and is based on physicochemical properties of amino acids rather than amino acid identity alone.

It is used here to predict the mutations required to convert one enzyme into a structurally related one of different substrate specificity, drastically reducing the exploration of sequence space required to interconvert enzyme function.

A variety of methods for identifying functional positions from sequence alignments of two or more groups of proteins have been reported (evolutionary trace,[2] SDPpred,[3] GroupSim,[4] SCA,[5] etc.). Some methods, for example, WebProAnalyst,[6] allow correlation of activity with physicochemical properties but do not compare two or more families or sequences. Analysis of correlated mutations, including correlated physicochemical properties, within protein families has also received significant attention (Chakrabarti and Panchenko,[7] CRASP,[8] SCA[5]). Most methods for comparing multiple families, with the notable exception of SCA, were primarily tested against benchmarks of existing data. To our knowledge, Janus is the only method created with transferring functional properties between proteins in mind and is validated experimentally here.

To calculate the functional importance of residues, Janus analyzes multiple sequence alignments (MSAs) of two structurally related proteins with distinct properties (referred to as "Start" and "Target" proteins) and includes structural information if available (see Supplementary Fig. 1 for an overview

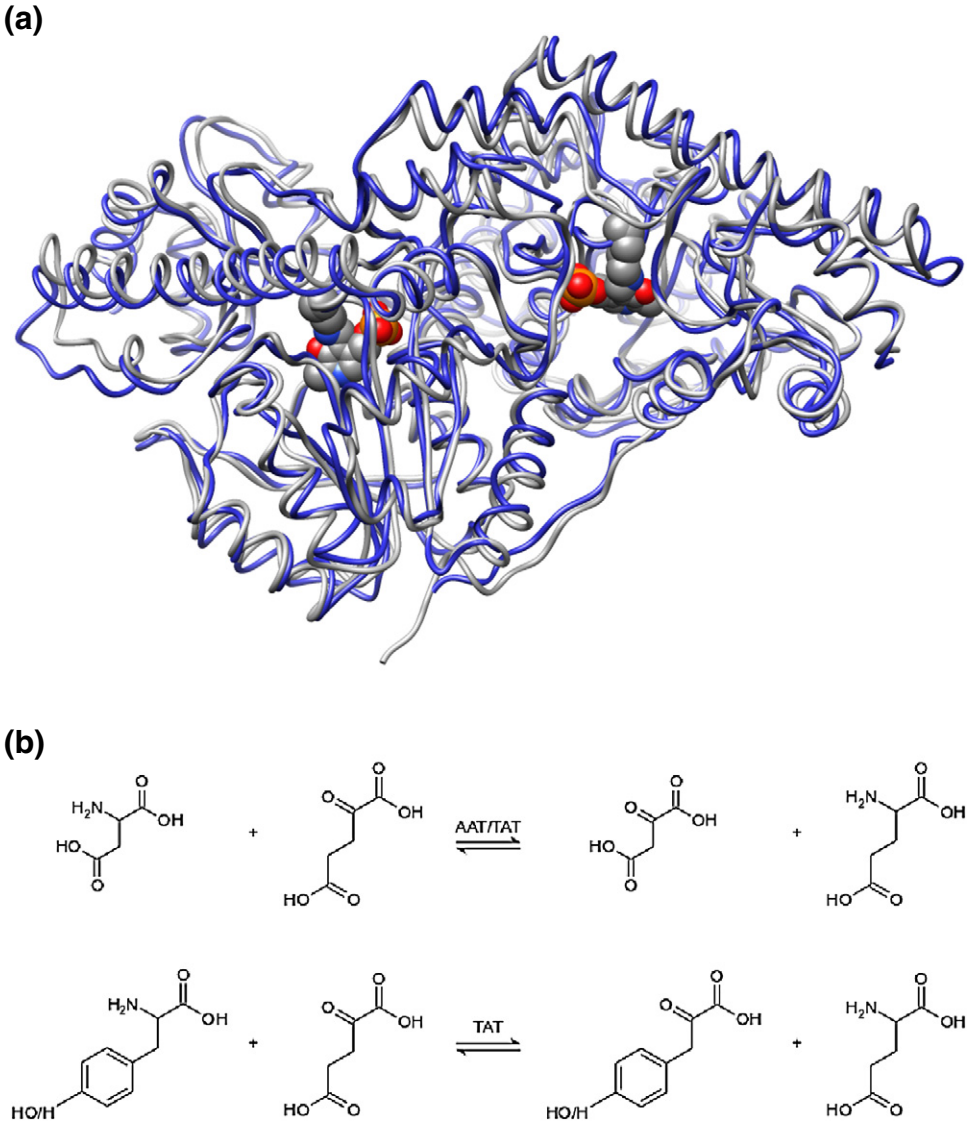

Fig. 1. (a) Structural similarity between AAT and TAT. *E. coli* AAT (gray; this work) and TAT (blue; PDB ID 3TAT) backbone traces are superposed with cofactor shown in space filling. The two proteins have 43% sequence identity and a C^{α} rmsd of 1.15 Å. Native α_2 dimers are shown for each. (b) Reactions catalyzed by AAT and TAT. AAT transfers the amino group between dicarboxylic amino acids and their corresponding α-keto acids. TAT catalyzes the same reaction but with expanded substrate specificity that includes aromatic α-amino and α-keto acids.