SUPPORTING INFORMATION

Detecting the Native Ligand Orientation by Interfacial Rigidity: SiteInterlock

Sebastian Raschka[1], Joseph Bemister-Buffington[1], and Leslie A. Kuhn[1,2,3]

[1]Department of Biochemistry & Molecular Biology, Michigan State University, Biochemistry Building, 603 Wilson Road, East Lansing, MI 48824

[2]Department of Computer Science & Engineering, Michigan State University

[3]Corresponding author: Leslie A. Kuhn

E-mail: kuhnl@msu.edu

Website: http://www.kuhnlab.bmb.msu.edu

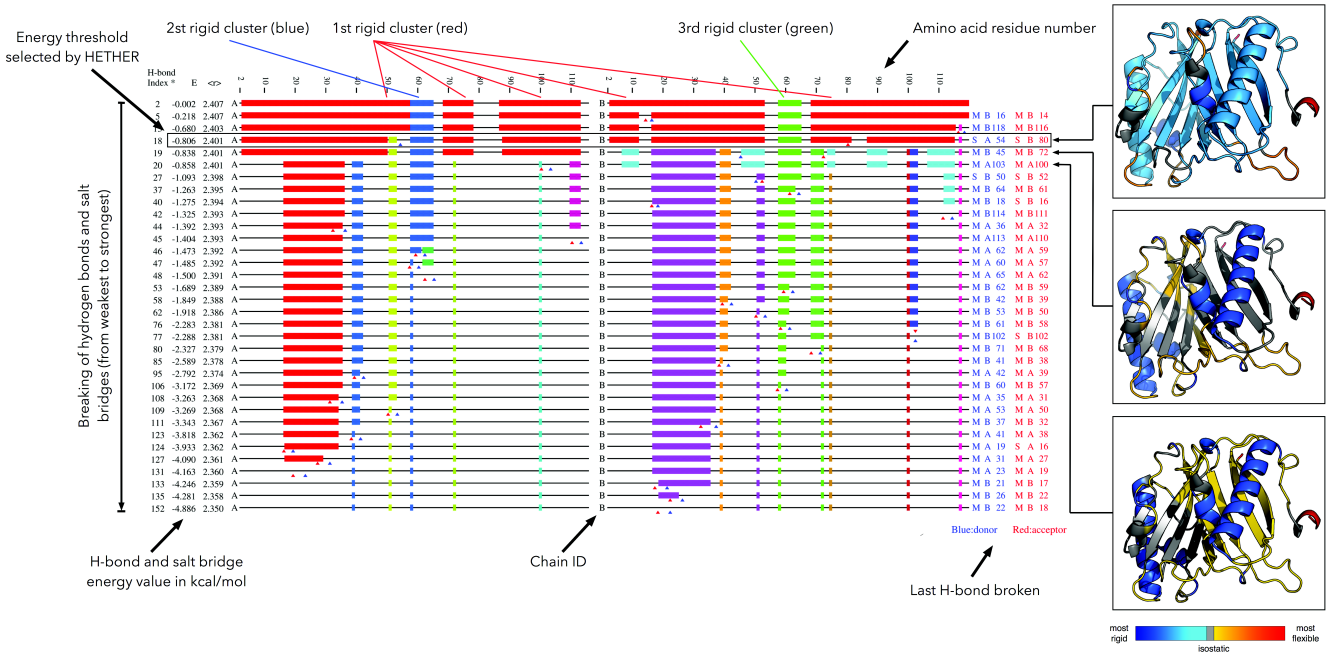
Telephone: 517-353-8745

Fax: 517-353-9334

Supporting Information Figure 1

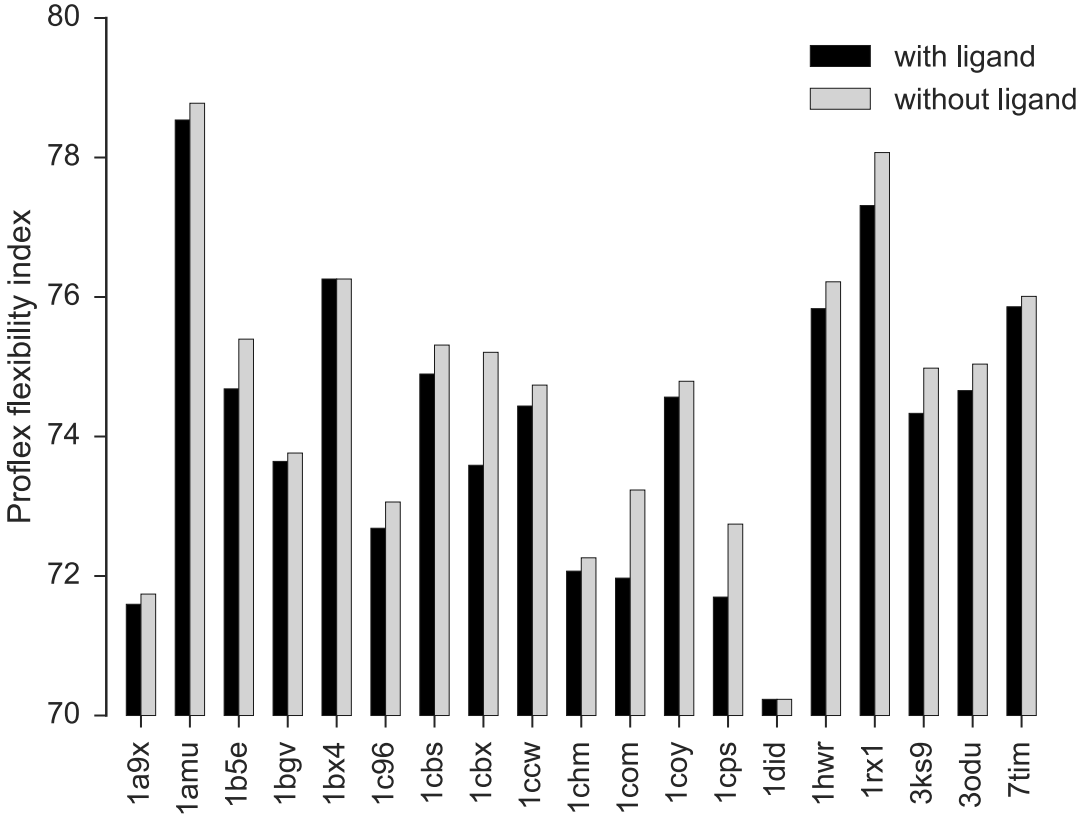
ProFlex hydrogen-bond dilution plot of the de-ligated protein structure of a monofunctional chorismate mutase from Bacillus subtilis (PDB code: 1com), showing the transition from mostly rigid to mostly flexible as hydrogen bonds and salt bridges are broken with increasing energy. The HETHER module of SiteInterlock is designed to identify the energy just before the protein becomes substantially flexible, as described below. The distinct lines in this plot show the rigid and flexible regions of the protein at different energy values, with successive lines representing increasingly flexible states of the protein as the energy level (temperature) increases. Residues of the protein chain are numbered from left to right at the top of the plot. At a given energy value, the thick, colored blocks in each row indicate the rigid clusters of the protein main chain, with a different color used for each independently rigid cluster of atoms. The thin, black lines correspond to intervening flexible regions observed in the protein bond network at that energy. A rigid region may be comprised of residues that are not contiguous in sequence; thus, blocks of residues with the same color indicate residues belonging to the same mutually-rigid region. The energy value for each row is listed in the second column from the left. The first row shows the predicted state of the protein when all hydrogen bonds and salt bridges are included in the bond network. The number of salt bridges and hydrogen bonds is listed in the leftmost column. The third column shows the average number of bonds connecting to each atom (averaged over all atoms in the protein) at that energy level, including covalent single and double bonds, bond-coordination constraints (constraining sp3 and sp2 centers in the correct geometry), hydrophobic tethers, hydrogen bonds, and salt bridges. For instance, the second row, at an energy value of -0.218 kcal/mol, shows the rigid and flexible regions in the protein when all hydrogen bonds and salt bridges with an energy of -0.218 kcal/mol or stronger are included in the bond network. Moving down the rows of the plot, the energy values increase and hydrogen bonds and salt bridges are incrementally broken (from weakest to strongest), resulting in an overall increase of flexible regions in the protein structures indicated by the intervening, black lines and fragmentation of rigid regions.

The energy value selected by HETHER is highlighted by the black frame shown at -0.806 kcal/mol, in which the main chain is mostly rigid (comprised by the large rigid region shown in red, plus two ~10-residue independent rigid regions colored in blue and green, and a very short rigid region in lime green appearing at residue 50). This state shows some residual flexibility that is sensitive to native-like ligand interactions, as described in the results in the main text. The rigid and flexible regions mapped onto the corresponding, ligand-free protein structure at different energy levels are shown at the far right, now colored by flexibility index (with colors defined in the spectrum bar shown beneath the structures). At the next energy step (-0.838 kcal/mol) above that chosen by HETHER, the protein structure decomposes into eight rigid clusters (red, yellow, blue, green, cyan, orange, lime green, and dark blue), which results in a structure with about one-third of the main chain being flexible. Thus, HETHER selected the last substantially stable state of the protein structure, as intended.



Supporting Information Figure 2

Rigidity of interfacial protein atoms (within 9 Å of ligand heavy atoms) in the presence (black bars) and absence (gray bars) of the crystallographic ligand pose for the 19 holo structures. Lower ProFlex values indicate greater rigidity. For 17 cases, the protein interface is more rigid in the presence of the ligand, and for 2 cases (PDB entries 1bx4 and 1did), it is equally rigid.



Supporting Information Figure 3

Relationship between the SiteInterlock score and ligand RMSD relative to the crystallographic pose for (A) dockings spanning the RMSD range of 0-5 Å for prephenic acid in complex with chorismate mutase (PDB entry 1com; also see Figure 4 for SiteInterlock results on two of these poses) and (B) 331 dockings from all 30 protein-ligand complexes. A funnel-like tendency is seen that discriminates more native-like dockings (closer to 0 Å RMSD) based on these dockings having more negative (rigid) SiteInterlock scores, particularly for dockings with RMSD values of ≤ 3 Å.