Amino Acid Physical Chemistry Furnishes a Two-Dimensional Basis Set for Computational Structural Biology Charles W. Carter, Jr.[1], Richard Wolfenden[1], Celia A. Schiffer[2], Ronald Swanstrom[1], and Marc Potempa[3] [1]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill [2]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts, Worcester, MA, [3]Department of Microbiology and Immunology, University of California, San Francisco, CA.

Efforts to relate the physical properties of amino acids quantitatively to protein folding have not met with notable success. Recent work on aminoacyl-tRNA synthetase recognition elements in tRNA highlighted the possibility that experimental transfer free energies from water to cyclohexane and vapor to cyclohexane—closely related to side chain hydrophobicity and size—for side chain mimics furnish an improved basis set to account for the accessible surface areas in folded proteins of 18 of the 20 amino acids [1,2]. Cysteine and proline are outliers, presumably because their roles in metal binding and in turns, respectively, produce large deviations in their expected distributions in folded proteins. We test the generality of that proposal by using these transfer free energies as a 2D basis set to predict cleavage rates for 140 variants of the cleavage sites processed by the HIV-1 protease, based on 6 natural cleavage sites processed during virus maturation. Previous work [3,4] attributed mutational perturbations of HIV-1 protease cleavage rates qualitatively to the size and polarity of the eight residues in the protease recognition site. Wild type sequences for six of the nine cleavage sites are cleaved at rates that vary over a range of nearly three orders of magnitude. Using the cleavage rates of these six natural sites as a test set, we trained regression models for cleavage rates of the mutant cleavage sites using as predictors two parameters per amino acid in the eight-residue protease recognition site. The training set could be fitted to R^2 = 0.88 with robust statistical t-test probabilities, and these models predicted the six wild-type cleavage rates omitted from the training set with comparable R^2 = 0.87-0.89 [5]. Extensive jackknife studies afford further validation. These results highlight the potential utility of applying amino acid side chain phase transfer free energies to the quantitative analysis of peptide-protein interactions, opening the way to broader applications in computational structural biology.

References

1. Carter CW, Jr., Wolfenden R (2015) tRNA Acceptor-Stem and Anticodon Bases Form Independent Codes Related to Protein Folding. Proc Nat Acad Sci USA 112 (24):7489-7494. doi:http://www.pnas.org/cgi/doi/10.1073/pnas.1507569112
2. Wolfenden R, Lewis CA, Yuan Y, Carter CW, Jr. (2015) Temperature dependence of amino acid hydrophobicities. Proc Nat Acad Sci USA 112 (24):7484-7488. doi:10.1073/pnas.1507565112
3. Rögnvaldsson T, You L, Garwicz D (2015) State of the art prediction of HIV-1 protease cleavage sites. Bioinformatics 31 (8):1204–1210. doi:10.1093/bioinformatics/btu810
4. You L, Garwicz D, Rögnvaldsson T (2005) Comprehensive Bioinformatic Analysis of the Specificity of Human Immunodeficiency Virus Type 1 Protease. Journal of Virology 79 (Oct. No. 19):12477–12486. doi:10.1128/JVI.79.19.12477–12486.2005
5. Potempa M, Carter CW, Jr, Schiffer, CA, Swanstrom R (2017) Site Specificity in the HIV-1 Protease Cleavage Rates. In Preparation