Testing of the Additivity-Based Protein Sequence to Reactivity Algorithm†

M. A. Qasim,‡ Wuyuan Lu,‡,§ Stephen M. Lu,‡,|| Michael Ranjbar,‡ ZhengPing Yi,‡ Yi-Wen Chiang,⊥ Kevin Ryan,⊥ Stephen Anderson,⊥ Wenlei Zhang,‡,@ Sabiha Qasim,‡ and Michael Laskowski, Jr.*,‡

Department of Chemistry, Purdue University, 560 Oval Drive, Brown Building, West Lafayette, Indiana 47907-2038, and Center for Advanced Biotechnology and Medicine, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854-5638

Received November 18, 2002; Revised Manuscript Received March 20, 2003

ABSTRACT: The standard free energies of association (or equilibrium constants) are predicted for 11 multiple variants of the turkey ovomucoid third domain, a member of the Kazal family of protein inhibitors, each interacting with six selected enzymes. The equilibrium constants for 38 of 66 possible interactions are strong enough to measure, and for these, the predicted and measured free energies are compared, thus providing an additional test of the additivity-based sequence to reactivity algorithm. The test appears to be unbiased as the 11 variants were designed a decade ago to study furin inhibition and the specificity of furin differs greatly from the specificities of our six target enzymes. As the contact regions of these inhibitors are highly positive, nonadditivity was expected. Of the 11 variants, one does not satisfy the restriction that either P_2 Thr or P_1' Glu should be present and all three measurable results on it, as expected, are nonadditive. For the remaining 35 measurements, 22 are additive, 12 are partially additive, and only one is (slightly) nonadditive. These results are comparable to those obtained for a set of 398 equilibrium constants for natural variants of ovomucoid third domains. The expectation that clustering of charges would be nonadditive is modified to the expectation that major nonadditivity will be observed only if the combining sites in both associating proteins involve large charge clusters of the opposite sign. It is also shown here that an analysis of a small variant set can be accomplished with a smaller subset, in this case 13 variants, rather than the whole set of 191 members used for the complete algorithm.

The sequence to reactivity algorithm for proteins (*1*, *2*) predicts the reactivity of all (subject to restrictions) possible members of a protein family from known reactivities of several appropriate target members of the family. The needed data consist of two parts. One is the identification of a subset of residues that can be varied without changing the structure yet determine the reactivity. The second is the production of all 19 single variants at each of the above positions and measurement of the reactivity of the wild type and of the 19*n* variants. Assuming that the free energy contributions of each of the positions are independently additive, we can calculate the reactivity of any sequence. It is highly likely that some residue pairs will be nonadditive.

In our sequence to reactivity algorithm (*1*), the standard free energy of association (equilibrium constant) between two different proteins has been studied. One member of the pair is one of the six serine proteinases: CHYM,[1] PPE, CARL, SGPA, SGPB, and HLE. The other member of the pair is any possible member (subject to restrictions) of the Kazal family of protein inhibitors of serine proteinases (for a recent list of such families, see ref *3*). For the wild type, the turkey ovomucoid third domain, OMTKY3, was selected (Figure 1). X-ray structures of its complexes with several of our six serine proteinases have been obtained (*4*−*9*), and a consensus set of 12 contact residues has been formed. These are colored in Figure 1. The two green residues are structurally very important and therefore were not varied. The 191 needed variants, including the wild type, were expressed, and the 6×191 (=1146) needed equilibrium constants were measured.

Now consider the association of just one of the 19*n* variants with a specified target protein. Its standard free energy of association, $\Delta G°(X_i)$, will in general be quite different from that of the wild type, $\Delta G°_{wt}$, for association with the same target protein. We call the difference between these two standard free energies the increment due to the replacement

$$\Delta\Delta G°(X_{wt}\ i\ X) = \Delta G°(X_i) - \Delta G°_{wt} \qquad (1)$$

where *i* is the position where the replacement was made, X_{wt} is the wild-type residue, and X is the variant residue at the *i*th position. The additivity assumption asserts that for members of a protein family the value of the free energy

† This research was supported at Purdue in part by funds from National Institutes of Health (NIH) Grants GM10831 and GM63539-01 and at Rutgers by NIH Grants AG11525 and AG10462.

* To whom correspondence should be addressed. Telephone: (765) 494-5291. Fax: (765) 494-0239. E-mail: michael.laskowski.1@purdue.edu.

‡ Purdue University.

§ Current address: Institute of Human Virology, University of Maryland, Baltimore, MD 21201.

|| Current address: University of Colorado Health Sciences Center, Denver, CO 80262.

⊥ Rutgers University.

@ Current address: Spectra Gases, Inc., 6480 Dobbin Rd., Columbia, MD 21045.

[1] Abbreviations: CHYM, bovine chymotrypsin Aα; PPE, porcine pancreatic elastase; CARL, subtilisin Carlsberg; SGPA, *Streptomyces griseus* protease A; SGPB, *S. griseus* protease B; HLE, human leukocyte elastase; OMTKY3, turkey ovomucoid third domain; CNBr, cyanogen bromide.

10.1021/bi027186u CCC: $25.00
Published on Web 05/06/2003

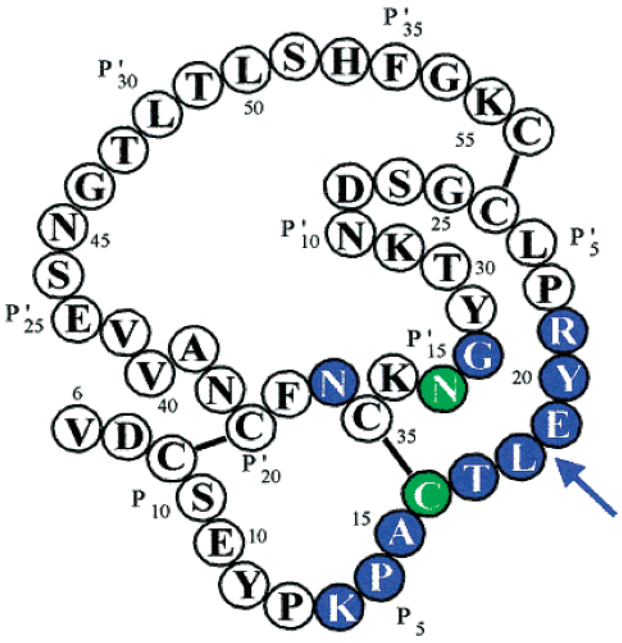

FIGURE 1: Covalent structure of OMTKY3. The bars indicate disulfide bridges. The arrow points to the reactive site peptide bond between P_1 and P_1' residues. The 12 colored residues comprise the consensus contact residue set. Of these 12, the two in green are structural and accept very few mutations in evolution. In contrast, the remaining 10 are hypervariable.

increment $\Delta\Delta G°(X_{wt}\ i\ X)$ is independent of the sequence context. In other words, changing other residues does not affect the value of the free energy increment.

Assuming additivity, the data allow for the prediction of association constants of multiple variants of the blue residues. There are 20^{10} or approximately 10^{13} such multiple variants. However, three-dimensional structure determinations show that in free inhibitors the side chain of Thr^{17} (P_2 residue) is hydrogen bonded to the side chain of Glu^{19} (P_1' residue) (*10*). Importantly, the strength (length) of this hydrogen bond appears to change greatly upon complex formation with some, but not all, enzymes (*4*−*8*). This is the hallmark of nonadditivity. Therefore, we decided to restrict the algorithm to those P_2−P_1' pairs that were measured directly. An easy way to state the restriction is that either Thr at P_2 or Glu at P_1' must be present in the variant that can be predicted. This restriction decreases the scope of the algorithm by a factor of 10.

The algorithm was tested with a set of 398 equilibrium constants of the six target enzymes interacting with numerous natural variants of ovomucoid third domains (*1*). The results were highly satisfactory as were the results with five very distantly related Kazal inhibitors. However, it was clear that more tests are needed. Here we ran into major difficulty. It turned out that most of our colleagues whom we asked for samples of inhibitors they had isolated, sequenced, and characterized no longer had any material left. Even in our own laboratory, some samples we needed were gone. Clearly, expression of test variants for more extensive tests is the ultimate solution, but these must be either rather complicated multiple variants or a very large number of simpler but still multiple ones.

Fortunately, we found in our laboratory a set of 11 constructs designed a decade ago to test the inhibition of furin (*11*). We expressed and characterized all of them. The set appeared to be interesting as it was designed for the inhibition of furin. As the specificity of furin is strikingly different from those of the six serine proteinases we have studied, this test set was clearly not biased by its design. Furin prefers Arg and Lys residues in many contact positions. Therefore, the contact regions of the inhibitors in the set are very positive. There was ample reason to think that highly positive or highly negative contact regions would be nonadditive as repulsive interactions between the residues would change upon binding to the enzyme (see also the Discussion). The "furin" set therefore provided a possible test of this expected nonadditivity.

EXPERIMENTAL PROCEDURES

Inhibitor Variants. The variants used in this study were constructed and expressed as described in earlier publications from the Purdue group (*11*, *12*). The fusion proteins containing two identical engineered B domains (designated as Z domains) of protein A, and the inhibitor variant, were expressed in the periplasmic space of *Escherichia coli*. The cells were grown in 9 L of 2× YT medium for 24 h at 30 °C. The fusion proteins were obtained in soluble and folded form by osmotic shock (*12*, *13*) and were purified by affinity chromatography on an IgG−Sepharose column. The two Z domains are connected to the inhibitor by an oligopeptide linker which has engineered sites for cleavage by Glu-specific *Staphylococcus aureus* protease V8 and CNBr. In the study presented here, the fusion protein was treated with 6% (w/v) CNBr in 5% (v/v) trifluoroacetic acid for 12 h. The inhibitor variants were separated from Z domains by size exclusion column chromatography on P-10 and further purified by ion exchange chromatography on S- and Q-Sepharose columns. The variants were routinely characterized by amino acid analysis and mass spectroscopy.

K_a Measurements. Association equilibrium constants for the association of inhibitors with serine proteinases were measured in 0.1 M Tris-HCl buffer containing 0.005% Triton X-100 (0.025% for measurements with HLE) and 0.02 M $CaCl_2$ at 21 ± 1 °C. The procedure used for K_a measurements is in principle based on the procedure of Green and Work (*14*), but it has been extensively modified in our laboratory over the past two decades (*13*, *15*−*17*). Ten separate mixtures of enzyme and inhibitor in various ratios are incubated for an appropriate period of time to reach equilibrium. An appropriate *p*-nitroanilide peptide substrate is added, and the free enzyme concentration is determined by spectrophotometrically monitoring the release of *p*-nitrophenol. Determination of correct inhibitor and enzyme concentrations is vital for accurate K_a measurements. We routinely determined the inhibitor concentration by amino acid analysis and the enzyme concentration by titrating against an appropriate inhibitor variant whose K_a is $\geq 5 \times 10^8$. Our procedure produces reliable results over a wide range from 10^3 to 10^{13} M^{-1}. From the fits obtained and the reproducibility of ~3000 $\Delta G°$ values determined in our laboratory, we conclude that the average standard deviation, σ, is ±100 cal/mol. This finding corresponds to log K_a ± 0.075 and K_a ± 20%. For further details, readers are referred to publications or Ph.D. theses from Purdue (*12*, *13*, *16*, *17*).

Table 1: Free Energies of Association in Kilocalories per Mole at pH 8.3 and 21 ± 1 °C

	CHYM	PPE	CARL	SGPA	SGPB	HLE
P_2K/P_1R						
measured	−7.86		−9.03	−7.33	−9.27	
predicted	−7.21	−1.31	−9.39	−7.00	−9.16	−2.75
$\Delta G°_I$	0.65		−0.36	0.33	0.11	
P_2R/P_1R						
measured	−8.10		−7.80	−6.54	−9.03	
predicted	−7.41	−1.36	−8.00	−6.71	−9.23	−2.68
$\Delta G°_I$	0.69		−0.20	−0.17	−0.20	
P_4R/P_1R						
measured	−10.32		−6.57	−10.05	−9.42	−5.98
predicted	−10.79	−2.05	−5.83	−9.42	−9.16	−6.36
$\Delta G°_I$	−0.47		0.74	0.63	0.26	−0.38
$P_4R/P_2K/P_1R$						
measured	−7.35		−5.03	−6.60	−7.82	
predicted	−6.74	1.61	−3.91	−5.74	−7.14	−3.01
$\Delta G°_I$	0.61		1.12	0.86	0.68	
$P_4R/P_2P/P_1R$						
measured	−6.43			−5.96	−6.04	
predicted	−6.60	0.77	−0.93	−5.70	−6.01	−3.32
$\Delta G°_I$	−0.17			0.26	0.03	
$P_5S/P_4R/P_2P/P_1R$						
measured	−6.86			−6.41	−6.39	
predicted	−7.03	1.05	−0.79	−5.68	−5.94	−3.89
$\Delta G°_I$	−0.17			0.73	0.45	
$P_6R/P_4R/P_2K/P_1R$						
measured	−7.31			−7.29	−8.99	
predicted	−7.11	1.62	−3.77	−6.27	−7.82	−3.64
$\Delta G°_I$	0.20			1.02	1.17	
$P_6R/P_5F/P_4R/P_2K/P_1R$						
measured	−7.24			−6.95	−8.85	
predicted	−7.41	1.86	−2.40	−6.94	−8.16	−5.48
$\Delta G°_I$	−0.17			0.01	0.69	
$P_4R/P_2K/P_1R/P_3'L$						
measured				−6.04	−7.62	
predicted	−3.94	1.99	−3.68	−5.21	−6.91	−3.86
$\Delta G°_I$				0.83	0.71	
$P_5S/P_4L/P_2K/P_1R$						
measured	−8.03		−9.39	−7.71	−9.60	
predicted	−7.50	−1.74	−10.99	−7.42	−8.92	−3.30
$\Delta G°_I$	0.53		−1.60	0.29	0.68	
$P_4R/P_2K/P_1R/P_1'S$						
measured	−6.95			−6.12	−7.41	
predicted	−3.97	3.91	−0.77	−3.17	−5.21	−1.18
$\Delta G°_I$	2.98			2.95	2.20	

RESULTS

Predicted ΔG° Values. Table 1 lists all available $\Delta G°$ values for the association of 11 multiple variants with our set of serine proteinases. Of these, three are double variants, two are triple variants, five are quadruple variants, and one is a pentuple variant. The additivity assumption allows us to predict their standard free energies of association from the standard free energies of association of all needed single variants and the wild type. Table 2A is a listing of the free energy of association of the wild type and of the 12 relevant single variants with six serine proteinases. The values for the wild type and the P_1 Arg variant have already been published (*12*). The remaining ones are excerpted from the database of 191 single variants recently described (*1*) but not as yet published.

The $\Delta G°_{predicted}$ values can now be computed by summing the observed $\Delta G°_{wt}$ value shown in Table 2A and the appropriate standard free energy increments (eq 1, Table 2B)

$$\Delta G°_{predicted} = \Delta G°_{TKY} + \sum_{i=1}^{i=10} \Delta\Delta G°(X_{TKY}\ i\ X) \quad (2)$$

Table 2: Experimental Free Energies of Association $\Delta G°(X_i)$ in Kilocalories per Mole at pH 8.3 and 21 ± 1 °C

single variant	CHYM	PPE	CARL	SGPA	SGPB	HLE
			(A) OMTKY3			
P_6R	−15.60	−14.33	−14.08	−16.02	−15.19	−13.84
P_6F	−14.14	−14.33	−15.00	−16.26	−15.23	−12.26
P_5S	−15.66	−14.06	−14.08	−15.47	−14.44	−13.78
P_5F	−15.53	−14.10	−12.85	−16.16	−14.85	−15.05
P_4L	−15.09	−15.05	−15.96	−15.93	−14.34	−13.19
P_4R	−14.76	−11.42	−8.74	−14.23	−12.49	−13.47
P_2P	−11.04	−11.52	−9.32	−11.77	−11.36	−10.17
P_2K	−11.18	−10.68	−12.30	−11.81	−12.49	−9.86
P_2R	−11.38	−10.73	−10.91	−11.52	−12.56	−9.79
P_1L, WT	−15.23	−14.34	−14.22	−15.49	−14.51	−13.21
P_1R	−11.26	−4.97	−11.31	−10.68	−11.18	−6.10
$P_1'S$	−12.46	−12.04	−11.08	−12.92	−12.58	−11.38
$P_3'L$	−12.43	−13.96	−13.99	−14.96	−14.28	−14.06
			(B) $\Delta\Delta G°(X_{TKY}\ i\ X) = \Delta G°(X_i) - \Delta G°_{TKY}$			
P_6R	−0.37	0.01	0.14	−0.53	−0.68	−0.63
P_6F	1.09	0.01	−0.78	−0.77	−0.72	0.95
P_5S	−0.43	0.28	0.14	0.02	0.07	−0.57
P_5F	−0.30	0.24	1.37	−0.67	−0.34	−1.84
P_4L	0.14	−0.71	−1.74	−0.44	0.17	0.02
P_4R	0.47	2.92	5.48	1.26	2.02	−0.26
P_2P	4.19	2.82	4.90	3.72	3.15	3.04
P_2K	4.05	3.66	1.92	3.68	2.02	3.35
P_2R	3.85	3.61	3.31	3.97	1.95	3.42
P_1L, WT	0	0	0	0	0	0
P_1R	3.97	9.37	2.91	4.81	3.33	7.11
$P_1'S$	2.77	2.30	3.14	2.57	1.93	1.83
$P_3'L$	2.80	0.38	0.23	0.53	0.23	−0.85

The readers are invited to check the predicted entries in Table 1 with a hand calculator to appreciate how simple the algorithm is.

The range of K_a measurements by direct techniques employed in our laboratory is from 10^3 to 10^{13} M^{-1}, which is equivalent to $\Delta G°$ values ranging from −4.05 to −17.55 kcal/mol. It is immediately seen that only somewhat more than half of the $\Delta G°$ values are measurable. This is also quite common with some natural inhibitors. The specific explanation for most of the difficulty here is straightforward. All 11 variants have Arg at P_1. This is extremely poor for HLE and even poorer for PPE. Both elastases have very small pockets, and the Arg side chain is very large. In addition, Arg is positive and elastases prefer hydrophobic, neutral side chains. Weak association of CARL is found for some of the variants with Arg at P_4, which is highly deleterious for that enzyme (*18*).

Measurement of ΔG° and Scoring. Measurements of $\Delta G°$ were obtained for 38 of the 66 possible combinations and are listed in Table 1. When both the measured and predicted standard free energy of association were available, then the $\Delta G°_I$, the free energy of interaction, or the difference between these two values is also given. Wells (*19*) defined the difference between the predicted and measured free energy of association as $\Delta G°_I$

$$\Delta G°_I = \Delta G°_{predicted} - \Delta G°_{measured} \quad (3)$$

where the subscript I denotes interaction, the hallmark of nonadditivity. It should be noted that both $\Delta G°_{predicted}$ and $\Delta G°_{measured}$ contain experimental errors as $\Delta G°_{predicted}$ is based on combining (eq 2) several of the free energy increments (Table 2B of the single variants). Thus, several experimental

Table 3: Free Energies of Interaction, $\Delta G°_I$, in Kilocalories per Mole (Predicted − Measured Standard Free Energy Changes)

	net charge[a]	$\Delta G°_I$ (kcal/mol) CHYM	CARL	SGPA	SGPB	HLE	additive	partially additive	nonadditive
wild-type OMTKY3	+1								
P_2K/P_1R	+3	0.65	−0.36	0.33	0.11		3	1	0
P_2R/P_1R	+3	0.69	−0.20	−0.17	−0.20		3	1	0
P_4R/P_1R	+3	−0.47	0.74	0.63	0.26	−0.38	2	3	0
$P_4R/P_2K/P_1R$	+4	0.61	1.12	0.86	0.68		0	4	0
$P_4R/P_2P/P_1R$	+3	−0.17		0.26	0.03		3	0	0
$P_5S/P_4R/P_2P/P_1R$	+3	−0.17		0.73	0.45		3	0	0
$P_6R/P_4R/P_2K/P_1R$	+4	0.20		1.02	1.17		1	2	0
$P_6R/P_5F/P_4R/P_2K/P_1R$	+4	−0.17		0.01	0.69		3	0	0
$P_4R/P_2K/P_1R/P_3L$	+3			0.83	0.71		1	1	0
$P_5S/P_4L/P_2K/P_1R$	+3	0.53	−1.60	0.29	0.68		3	0	1
					total		22	12	1
$P_4R/P_2K/P_1R/P_1'S$	+5	2.98		2.95	2.20		0	0	3

no. of changes	$\lvert\Delta G°_I\rvert$ (kcal/mol)[b] additive	partially additive	nonadditive
0	0−0.28	0.28−0.56	>0.56
1	0−0.28	0.28−0.56	>0.56
2	0−0.40	0.40−0.80	>0.80
3	0−0.57	0.57−1.14	>1.14
4	0−0.75	0.75−1.50	>1.50
5	0−0.94	0.94−1.88	>1.88

[a] Net charge in the contact region of inhibitor variants. [b] Error analysis deals with propagated error in multiple substitutions. It is discussed in detail in ref *1*.

errors, rather than one, are involved. Also, each $\Delta\Delta G°$ value involves not one but two measurements. The propagated error in $\Delta G°_I$ was analyzed in ref *1*. The relevant cases are given at the bottom of Table 3. The second column in Table 3 gives the estimated net charge in the combining region of the inhibitor (colored residues in Figure 1) at the pH (pH 8.3) of K_a measurements. The net charge in this region in wild-type OMTKY3 is +1 due to Lys at P_6, Glu at P_1', and Arg at P_3'. The changes in charge are easily inferred from the listed substitutions in the variants. It is clear that all the variants have highly positive combining regions and an unusually large concentration of like charges.

The remainder of the table evaluates the $\Delta G°_I$ values of Table 1. $\Delta G°_I$ can deviate from zero for three reasons (*1*): (1) nonadditivity, (2) effect of changing the white residues in Figure 1, and (3) experimental errors in measured $\Delta G°$ and especially in predicted $\Delta G°$ as these values are based on experimental data in Table 2. A 1σ error in $\Delta G°$ was estimated to be 100 cal/mol. This is used to divide the results into additive, partially additive, and nonadditive groups (*1*). We arbitrarily call the cases where $|\Delta G°_I| \leq 2\sigma$ of the expected propagated experimental error additive. We call cases where $2\sigma < |\Delta G°_I| \leq 4\sigma$ partially additive and cases where $|\Delta G°_I| > 4\sigma$ nonadditive. The exact criteria are listed at the bottom of Table 3. For the first 10 variants in Table 3, the results are strikingly good: of 35 cases, 22 (63%) are additive, 12 cases (34%) are partially additive, and only one (3%) is nonadditive. Note that this last nonadditive entry is borderline. These results are comparable to those obtained with a much larger analysis of 398 free energies of association of natural ovomucoid third domains (*1*). Those were 63% additive, 27% partially additive, and 10% nonadditive. The results are highly gratifying since we anticipated they would be nonadditive due to the large positive charge on the combining region. In sharp contrast are the three nonadditive results out of three measured for the last variant in Table 3. This is analyzed in more detail in the Discussion.

DISCUSSION

We were pleased by the significant additivity of ovomucoid third domain natural variants (*1*). We are also pleased by comparable results that are seen here because we had strong reason to expect that these results would be worse.

One of the additive results listed in Tables 1 and 2 is shown in the double mutant cycle in Figure 2. Note that in this figure the two $\Delta\Delta G°$(ThrP_2Lys) values are within experimental error of each other as are the two $\Delta\Delta G°$(LeuP_1-Arg) values. An easy overall assessment of additivity is that the $\Delta G°_I$ term (eq 3) is 0.11 kcal/mol, while the allowed interval for an additive double mutant cycle is 0.4 to −0.4 kcal/mol as given at the bottom of Table 3.

Nonadditivity of the Last Variant. This variant has Lys at P_2, i.e., ThrP_2Lys change from the wild type, and Ser at P_1', i.e., GluP_1'Ser change from the wild type. As stated in the introductory section, when, as seen frequently (*20*), Thr at P_2 and Glu at P_1' are present in the same molecule, their side chains are hydrogen bonded. The strength of this hydrogen bond is greatly affected by formation of complexes with some enzymes. To study this further, we designed a cycle where the P_2Thr−P_1'Glu pair is gradually replaced by a hydrophobic Val−Val pair (Figure 3). This is a highly nonadditive cycle, as the presence of Val at either P_2 or P_1' breaks the hydrogen bond and therefore weakens the binding of some enzymes. This suggests that the P_2Val/P_1'Val double variant should be very weak. It is not as weak as predicted, and the cycle is nonadditive. Figure 3 shows the cycle only for CHYM. However, Table 4 lists the $\Delta G°_I$ values for CHYM, SGPA, and SGPB and compares them to $\Delta G°_I$ values for the $P_4R/P_2K/P_1R/P_1'S$ variant given in Tables 1

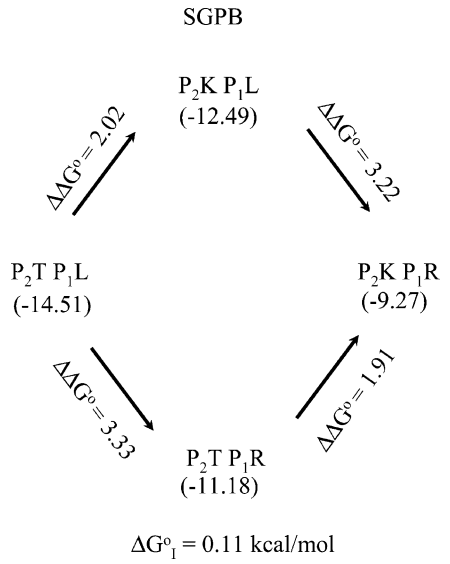

FIGURE 2: Double mutant cycle (*27*) for an additive interaction. The numbers under each entry are the measured standard free energies of interaction of SGPB with the wild-type OMTKY3 (middle left), the two single variants (center at the top and bottom), and the double variant (middle right). The first three entries are taken from Table 2, and the last entry is from Table 1 (top line). There are two pairs of $\Delta\Delta G°$ values: $\Delta\Delta G°$(ThrP$_2$Lys) and $\Delta\Delta G°$-(LeuP$_1$Arg). Within each pair, the same change occurs in two sequence contexts. Note that here the $\Delta\Delta G°$ values are independent of the sequence context well within the allowed (0.40 kcal/mol; see Table 3) experimental error. We can conclude that interactions between P_2 and P_1 residues that are studied here do not change on complex formation. In this case, the standard free energy of association can be predicted from the first three values by using additivity.

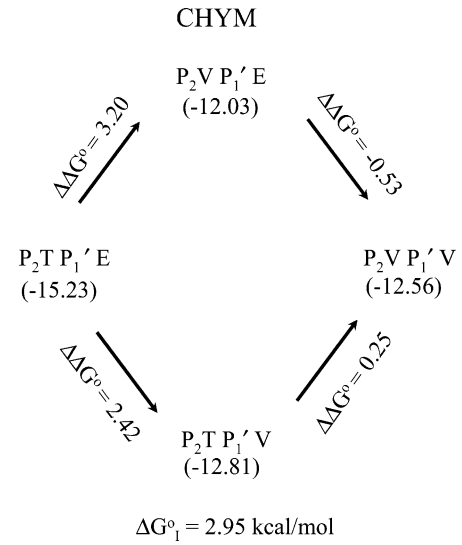

FIGURE 3: Double mutant cycle illustrating a strongly nonadditive situation. The data are from Qasim et al. (*28*). Note that the two values for the $\Delta\Delta G°$(ThrP$_2$Val) change (top left and bottom right, 2.95 kcal/mol) differ far more than the allowed experimental error of 0.4 kcal/mol (see Table 2). The same large difference exists between the two $\Delta\Delta G°$(GluP$_1$′Val) values (bottom left and top right, 2.95 kcal/mol). The results are consistent with expectations based on X-ray crystallography (see the text). In free OMTKY3, the P_2T side chain forms a hydrogen bond with the P$_1$′E side chain. Upon formation of a complex with CHYM, the bond becomes much shorter and therefore stronger. This strengthening is reflected in the very negative $\Delta G°$ for the OMTKY3−CHYM association. Making either the P_2V or P$_1$′V substitution disrupts this hydrogen bond both in the free inhibitor and in the complex and produces a weakening of binding of 2.95 kcal/mol, the $\Delta G°_I$ value. Therefore, the values of $\Delta\Delta G°$(ThrP$_2$Val) and $\Delta\Delta G°$(GluP$_1$′Val) are large and positive as each carries the additional 2.95 kcal/mol term as well as a smaller term associated with the substitution in a non-hydrogen-bonded system. As the hydrogen bond is disrupted in both middle forms, going from the middle to the right gives only the $\Delta\Delta G°$ for the substitution in the non-hydrogen-bonded system. $\Delta G°$ of the double mutant cannot be predicted by simple additivity because two 2.95 kcal/mol terms for rupture of the hydrogen bond would be added. Only one is needed. Obviously, a correction for this can be made and is planned.

and 3. The match is excellent. It appears that the nonadditivity of the last variant arises from the same cause.

Highly Charged Combining Region. The highly charged combining region might lead to nonadditivity. This is because we might expect a change in the repulsive interaction between the several positive side chains upon complex formation. Such a change might arise from a change in either the dielectric environment, the intercharge distances, or both. The observed additivity is therefore a surprise.

We also, somewhat hastily, expected nonadditivity from the excellent double variant studies on barnase−barstar association (*21*). Barnase is a nuclease and, like most nucleases that hydrolyze highly negatively charged nucleic acids, carries a large positive charge in its combining sites. As many protein enzyme inhibitors mimic the enzyme substrates, the barnase inhibitor, barstar, has a highly negative combining site (for a recent review of such interactions of nucleases and their inhibitors, see ref *22*). Many of the double variant cycles on the interaction of barnase and barstar were highly nonadditive, confirming our nonadditivity expectation. However, we and many other protein chemists to whom we talked misread these results. In the previous studies with furin (*11*) as well as in the current paper, all multiple mutations are made in the inhibitor. The enzyme is left unchanged. In the barnase−barstar system (*21*), one change is made in the enzyme (barnase) and the other in the inhibitor (barstar). This is an excellent strategy for evaluation of the interaction energies. However, (1) the information that is obtained does not lend itself well to the construction of a predictive algorithm, and (2) the likelihood of nonadditivity is much greater than in our approach. Therefore, the otherwise excellent paper on barnase−barstar interaction (*21*) does not serve us well as a predictive analogue.

Table 4: Nonadditivity Involving P_2T/P$_1$′E Changes

mutational changes	$\Delta G°_I$ (kcal/mol) CHYM	SGPA	SGPB
P_2T/P$_1$′E → P_2V/P$_1$′V	2.95	2.98	2.07
P_4A/P_2T/P$_1$L/P$_1$′E → P_4R/P_2K/P$_1$R/P$_1$′S	2.98	2.95	2.20

However, such an analogue was almost at hand. As stated previously, this set of variants was designed as inhibitors of furin and furin has a negative combining site as clearly demonstrated by the conversion of subtilisin BPN′ to an enzyme with furin-like specificity by introduction of three Asp residues into the combining region of subtilisin (*23*). The interaction of these variants with furin was studied in our laboratory, and we found the P_4R/P$_1$R OMTKY3 variant to be strongly nonadditive. Also, the association equilibrium constants for wild-type OMTKY3, P_4R OMTKY3, and P$_1$R OMTKY3 were weaker than the lower bound of the measurement range. Only the double variant P_4R/P$_1$R OMTKY3 was sufficiently strong as an inhibitor to be well

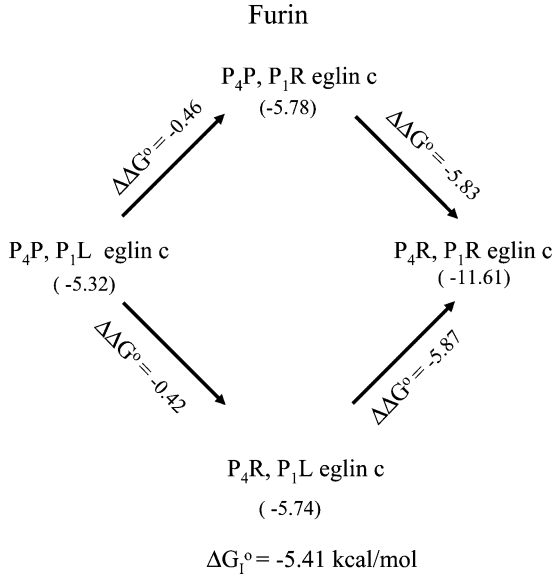

FIGURE 4: Double mutant cycle showing strong nonadditivity. The data are from Komiyama and Fuller (*24*). The enzyme is furin, and the inhibitors are eglin c and its variants. It is clear that the two values of $\Delta\Delta G°$(ProP_4Arg) differ dramatically. So do the two values of $\Delta G°$(LeuP_1Arg). Further inspection shows that introducing only one Arg, either at P_4 or at P_1, increases the level of biding only slightly. Adding the second Arg increases the level of binding dramatically. The nonadditivity observed here results from two Arg residues interacting with furin, which uses Asp residues to recognize its substrates (for a discussion, see the text). This effect was mistakenly expected by us for the two-Arg OMTKY3 variant interacting with CHYM, CARL, SGPA, SGPB, and HLE. As can be seen in the third line of Table 3, two results are additive and three partially additive. We believe that the difference between furin and these enzymes is that furin's site is strongly negatively charged while the others are nearly neutral.

inside the measurement range. As the measurements of furin inhibition are rather difficult, we decided the results were not publishable after presenting them at a meeting. However, such results could be improved. One possibility would be to change some residues in positions different from P_4 and P_1 to make OMTKY3 a better furin inhibitor. Such an approach (*1*, *2*) requires a wealth of data on furin inhibition that we do not have. The other approach is to do the experiment in a different inhibitor family, which is inherently a better furin inhibitor. The second approach was quite independently adopted by using a set of variants closely similar to our OMTKY3 set but using eglin c as the scaffold (*24*). An additivity cycle using these data for the inhibition of furin is shown in Figure 4. The nonadditivity observed in this case is in good quantitative agreement with our unpublished OMTKY3 results.

Previous quantitative studies from our laboratory (*25*, *26*) show that (1) changes in the P_1 position in OMTKY3 and in eglin c are strikingly interscaffolding additive and (2) eglin c is frequently a better inhibitor than OMTKY3 even though both share a Leu residue at P_1. The enzymes that are better inhibited by eglin c are α-lytic proteinase, *Streptomyces griseus* proteinases A and B, bovine chymotrypsins A and B, mosquito chymotrypsin, human cathepsin G, human leukocyte elastase, and cercarial elastase. Only porcine pancreatic elastase exhibits strikingly opposite behavior. Thus, the stronger inhibition of furin by analogous constructs in the eglin c frame is not unexpected, nor is the interscaffolding additivity of the results.

It appears that if only one of the interacting partners has a strongly charged regions, multiple variants made entirely in that partner will be largely additive.

Discrimination of SGPA from SGPB. OMTKY3 inhibits SGPA 1.0 kcal/mol more strongly than it inhibits SGPB. These two enzymes are very similar and generally exhibit similar $\Delta G°$ values for inhibition by natural avian ovomucoid third domains and by their variants, but in most cases, SGPA is ∼1 kcal/mol stronger. The variants used in this study mostly inhibit SGPB more strongly, a reversal of the common observation. The reason for this reversal is in part due to the fact that all these variants have Arg at P_1, and a single variant of this type [see Table 2, the R18 (P_1R) line] inhibits SGPB 0.5 kcal/mol more strongly than SGPA. It seemed worthwhile to check whether the predictions are sufficiently accurate to pick up the subtle difference in SGPA and SGPB discrimination. They are correct in 10 cases out of 11. They are even correct for the last entry in Table 1 where there is a very large difference between the predicted and measured values.

Additivity and Predictability Studies Can Be Carried Out on Relatively Small Databases. We received numerous well-intentioned criticisms of our additivity-based predictive algorithm (*1*). They can be divided into two classes. The first is that proteins are (frequently) cooperative. As cooperativity is an antithesis of additivity, one should not expect additivity. The results of both the previous (*1*) and the current paper show that while nonadditivity is found in some protein systems, it occurs only in a few residue pairs. We believe that it is sufficiently rare to initially start with a fully additive model in which then we can identify and deal with relatively rare nonadditivities. It is also worth noting again that for the association of inhibitors with highly positively charged combining regions interacting with neutral combining region enzymes, nonadditivity was expected but not found.

The other criticism is somewhat the obverse of the first. Our approach is too much work. The complete algorithm (*1*) required the expression of the wild type and of 190 specified variants (19 variants at each of the 10 contact positions). It then required measurement of 6×191 (=1146) equilibrium constants. Not only was the effort involved large, we were conspicuously lucky to express all 191 variants and to measure all the 1146 equilibrium constants. The current paper provides many examples of associations falling outside the current measurement range. However, in this paper, we demonstrate that all 191 variants and 1146 measurements were not needed for the additivity analysis carried out here. The wild type and 12 variants produced all 78 entries given in Table 2. This was sufficient to predict all needed $\Delta G°$ predicted values of Table 1. Indeed, the 191 variants and 1146 measurements were not all needed for the much larger algorithm tests (*1*). They were, however, all needed to provide the best possible inhibitor for each of the enzymes we study, or the most specific one or the distribution function of standard free energies of association for all possible inhibitors.

Will Such an Approach Succeed with Other Systems? We believe it will, but we are not sure. To the best of our knowledge, a comparable amount of work was not invested in any other protein family. A more detailed answer is provided in a recent review (*2*).

Difficulty in Obtaining Protein Test Samples. We already mentioned that it proved to be extremely difficult to obtain samples of Kazal inhibitors that were isolated, sequenced, and characterized by other laboratories. They are unavailable as they were exhausted in earlier experiments, lost, or discarded in laboratory cleanups. As this is in contrast to DNA (genes or plasmids), it may be worthwhile for granting agencies to take an interest in preserving the proteins, which were described in the literature with their support.

ACKNOWLEDGMENT

Dr. Etti Harms of our Biological Sciences Department provided the P_2Val/P_1'Val OMTKY3. Data obtained by us for this variant are given in Figure 3 and Table 4.

REFERENCES

1. Lu, S. M., Lu, W., Qasim, M. A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y.-W., Cook, J., James, M. N. G., Kato, I., Kelly, C., Kohr, W., Komiyama, T., Lin, T.-Y., Ogawa, M., Otlewski, J., Park, S.-J., Qasim, S., Ranjbar, M., Tashiro, M., Warne, N., Whatley, H., Wieczorek, A., Wieczorek, M., Wilusz, T., Wynn, R., Zhang, W., and Laskowski, M., Jr. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 1410−1415.
2. Laskowski, M., Jr., Qasim, M. A., and Yi, Z.-P. (2003) *Curr. Opin. Struct. Biol. 13*, 130−139.
3. Laskowski, M., Jr., and Qasim, M. A. (2000) *Biochim. Biophys. Acta 1422*, 324−337.
4. Fujinaga, M., Read, R. J., Sielecki, A., Ardelt, W., Laskowski, M., Jr., and James, M. N. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4868−4872.
5. Bode, W., Wei, A.-Z., Huber, R., Meyer, E., Travis, J., and Neumann, S. (1986) *EMBO J. 5*, 2453−2458.
6. Fujinaga, M., Sielecki, A. R., Reed, R. J., Ardelt, W., Laskowski, M., Jr., and James, M. N. G. (1987) *J. Mol. Biol. 195*, 397−418.
7. Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr., and James, M. N. G. (1995) *Protein Sci. 4*, 1985−1997.
8. Bateman, K. S., Anderson, S., Lu, W., Qasim, M. A., Laskowski, M., Jr., and James, M. N. G. (2000) *Protein Sci. 9*, 83−94.
9. Bateman, J. S., Huang, K., Anderson, S., Lu, W., Qasim, M. A., and Laskowski, M., Jr. (2001) *J. Mol. Biol. 305*, 839−849.
10. Bode, W., Epp, O., Huber, R., Laskowski, M., Jr., and Ardelt, W. (1985) *Eur. J. Biochem. 147*, 387−395.
11. Lu, W., Zhang, W., Molloy, S., Thomas, G., Ryan, K., Chiang, Y.-W., Anderson, S., and Laskowski, M., Jr. (1993) *J. Biol. Chem. 268*, 14583−14585.
12. Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W., Anderson, S., Chiang, Y.-W., Ogin, E., Rothberg, I., Ryan, K., and Laskowski, M., Jr. (1997) *J. Mol. Biol. 266*, 441−461.
13. Lu, W. (1994) Ph.D. Thesis, Purdue University, West Lafayette, IN.
14. Green, N. M., and Work, E. (1953) *Biochem. J. 54*, 347−352.
15. Empie, M. W., and Laskowski, M., Jr. (1982) *Biochemistry 21*, 2274−2284.
16. Park, S. J. (1985) Ph.D. Thesis, Purdue University, West Lafayette, IN.
17. Wynn, R. (1985) Ph.D. Thesis, Purdue University, West Lafayette, IN.
18. Lu, S. M. (2000) Ph.D. Thesis, Purdue University, West Lafayette, IN.
19. Wells, J. A. (1990) *Biochemistry 29*, 8509−8517.
20. Apostol, I., Giletto, A., Komiyama, T., Zhang, W., and Laskowski, M., Jr. (1993) *J. Protein Chem. 12*, 419−433.
21. Schreiber, G., and Fersht, A. R. (1995) *J. Mol. Biol. 248*, 478−486.
22. Kleanthous, C., and Pommer, A. J. (2000) *Protein−protein recognition: The frontiers in molecular biology series* (Kleanthous, C., Ed.) pp 280−311, Oxford University Press, New York.
23. Ballinger, M. D., Tom, J., and Wells, J. A. (1996) *Biochemistry 35*, 13579−13585.
24. Komiyama, T., and Fuller, R. S. (2000) *Biochemistry 39*, 15156−15165.
25. Qasim, M. A., Ganz, P. J., Saunders, C. W., Bateman, K. S., James, M. N. G., and Laskowski, M., Jr. (1997) *Biochemistry 36*, 1598−1607.
26. Qasim, M. A., Lu, S. M., Ding, J., Bateman, K. S., James, M. N. G., Anderson, S., Song, J., Markley, J. L., Ganz, P. J., Saunders, C. W., and Laskowski, M., Jr. (1999) *Biochemistry 38*, 7142−7150.
27. Laskowski, M., Jr., Tashiro, M., Empie, M. W., Park, S. J., Kato, I., Ardelt, W., and Wieczorek, M. (1983) in *Proteinase Inhibitors: Medical and Biological Aspects* (Katanuma, N., Umezawa, H., and Holtzer, H., Eds.) pp 55−68, Japan Scientific Societies, Tokyo, and Springer-Verlag, Berlin.
28. Qasim, M. A., Harms, E., Laskowski, M., Jr., Ranjbar, M., Yi, Z.-P., and Qasim, S. (2002) *Protein Sci. 11*, 161.

BI027186U