

# Additivity-based prediction of equilibrium constants for some protein–protein associations

Michael Laskowski Jr<sup>\*</sup>, MA Qasim<sup>†</sup> and ZhengPing Yi<sup>‡</sup>

For many protein families, such as serine proteinases or serine proteinase inhibitors, the family assignment predicts reactivity only in general terms. Both detailed specificity and quantitative reactivity are lacking. We believe that, for many such protein families, algorithms can be devised by defining the subset of  $n$  functionally important sequence positions, making the  $19^n$  possible single mutants and measuring their reactivity. Given the assumption that the contributions of the  $n$  positions are additive, the reactivities of the  $20^n$  variants can be predicted. This is illustrated by an almost complete algorithm for the Kazal family of protein inhibitors of serine proteinases.

## Addresses

Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907-2084, USA

<sup>\*</sup>e-mail: michael.laskowski.1@purdue.edu

<sup>†</sup>e-mail: abul@purdue.edu

<sup>‡</sup>e-mail: zpyi@purdue.edu

**Current Opinion in Structural Biology** 2003, **13**:130–139

This review comes from a themed issue on  
Folding and binding  
Edited by Jane Clarke and Gideon Schreiber

0959-440X/03/\$ – see front matter  
© 2003 Elsevier Science Ltd. All rights reserved.

DOI 10.1016/S0959-440X(03)00013-7

## Abbreviations

|               |  |
|---------------|--|
| <b>BPTI</b>   | bovine pancreatic trypsin inhibitor        |
| <b>CARL</b>   | subtilisin Carlsberg                       |
| <b>CHYM</b>   | bovine chymotrypsin A $\alpha$             |
| <b>HLE</b>    | human leukocyte elastase                   |
| <b>OMTKY3</b> | turkey ovomucoid third domain              |
| <b>PPE</b>    | porcine pancreatic elastase                |
| <b>SGPA/B</b> | <i>Streptomyces griseus</i> proteinase A/B |
| <b>SSI</b>    | <i>Streptomyces</i> subtilisin inhibitor   |

## Introduction

Anfinsen [1] showed that reduced and denatured bovine pancreatic ribonuclease A recovers its disulfide bridge pairing, its native structure and, most importantly, its enzymatic activity. Merrifield's [2] work on the chemical synthesis of ribonuclease eliminated any doubt that 'memory' of previous structure played a part. The work on ribonuclease was soon generalized to other proteins, even though some exceptions were found [3,4<sup>\*</sup>]. Therefore, an existence theorem — 'algorithms to predict reactivity of most proteins that are based on sequence alone exist' — is available.

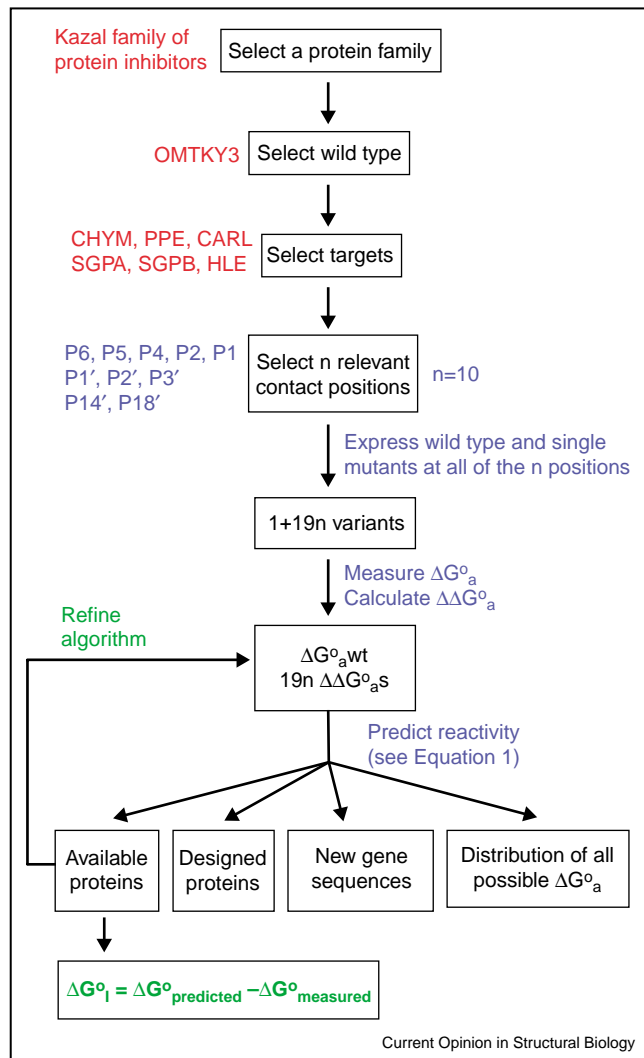
The search for such algorithms is one of the main goals of molecular biology. However, most workers, including Anfinsen [1], divide the algorithms into sequence to fold and fold to reactivity, and concentrate mostly on the first. We divide the algorithm into sequence to target protein binding and target protein binding to reactivity, if needed. However, for some of the cases on which we currently focus, target protein binding is the reactivity: protein enzyme–protein inhibitor interactions, protein antigen–antibody interactions, protein channel blockers and their target channels, and numerous protein and peptide transport proteins. Whereas the products of most algorithms are folds, the products of the algorithm described here are equilibrium constants.

An important characteristic of the algorithm described here is that it is capable of providing equilibrium constant predictions for all possible members of an entire protein family with astonishing calculation speed. This is a consequence of the simplicity of the additivity assumption and of the availability of a very large empirical database on which the calculation is based. Figure 1 provides a flow diagram for the construction of the algorithm. In this review, we summarize the only case to date in which such an approach has been taken, mostly from our own work. This is because, although many workers made decisive contributions to the analysis of protein–protein interaction by mutagenesis (and some of these were crucial to our work), no one else announced a family-wide sequence to binding algorithm. We certainly hope that many others will join us and that soon someone will be able to write a truly balanced review.

## Family-based algorithms

The assignment of new sequences to already known protein families has become more and more facile [5]. Such an assignment provides a great deal of information. Members of a family share a common global three-dimensional structure. In some cases, such as the triosephosphate isomerases, all family members share the same specificity and quantitative reactivity, such as the  $k_{cat}/K_m$  value at a diffusion-controlled limit [6]. For natural members of such families, the type of algorithm we describe here appears redundant. However, a large number of protein family members share only a common mechanism of interaction and general type of reactivity. Such is the case with the 18 known families of standard-mechanism canonical [7–9] protein inhibitors of serine proteinases. All 18 families have different global three-dimensional structures, but a common set of mainchain Ramachandran angles in the common part of the

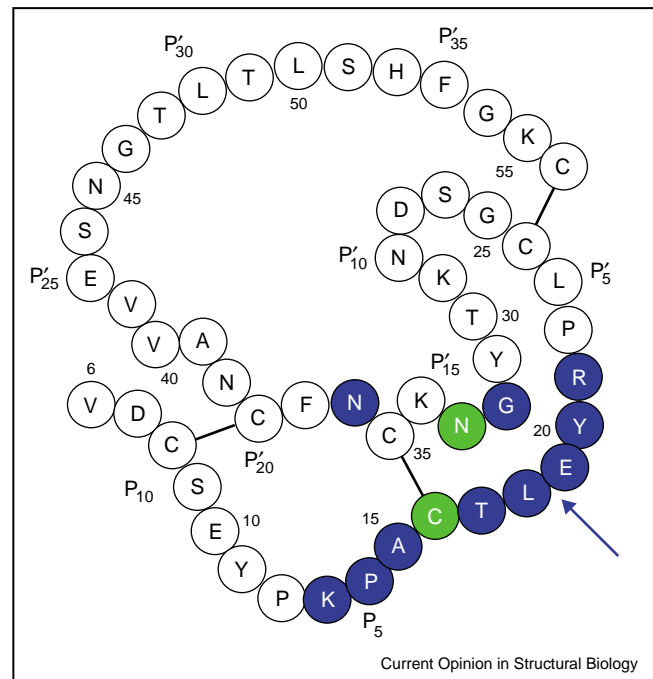
Figure 1



Flow diagram for the additivity-based prediction of the standard free energy of association between all possible members of a protein family and a selected member of another protein family. The Kazal family was selected to be analyzed because we expected that the number of contact residues contributed by these inhibitors would be smaller than 20. OMTKY3 was selected as the wild type because the natural variant was easy to obtain and because it strongly inhibits all six enzymes (CHYM, PPE, CARL, SGPA, SGPB and HLE) we chose as targets. These enzymes have been extensively studied by others. Although one enzyme would suffice, we chose six to stretch the test variant sets (we get six numbers for each variant) and to provide more evidence of the generality of the algorithm. The consensus variable contact residue set is described in Figure 2. The database for the algorithm used here is solely based on experiments. However, it seems that the  $\Delta\Delta G^\circ$  terms could be attractive for calculation. Thus, the scheme could be modified to a combination of experiments and calculation.

proteinase-binding region of the molecule, residues  $P_3$  to  $P'_3$  in Schechter and Berger [10] notation (see Figure 2, which also serves as an explanation of this notation). They all exhibit the same 'standard mechanism' of interaction

Figure 2



Covalent structure of OMTKY3, a Kazal family protein inhibitor. The three disulfide bridges are indicated by bars. Residues are numbered sequentially from Val6 to Cys56. Schechter and Berger [10] notation starts at the reactive site peptide bond (indicated by the arrow). The subscript indicates the distance from this bond. Residues preceding the reactive site are labeled  $P_1$ ,  $P_2$ – $P_n$ , those following it  $P'_1$ ,  $P'_2$ – $P'_n$ . The binding sites on the enzyme for these inhibitor residues are often referred to as  $S_1$ ,  $S_2$ – $S_n$  and  $S'_1$ ,  $S'_2$ – $S'_n$ , respectively. X-ray structures revealed that, in complexes of OMTKY3 with CHYM, SGPB and HLE, there is a clear consensus of 12 contact residues. These are indicated in color. Of these, the two marked in green are clearly structural. The structural role of  $P_3$  Cys16 is obvious and the  $P'_{15}$  Asn33 sidechain makes three hydrogen bonds to other residues. The remainder, the variable contact residue set, are marked in blue. Among them are the most variable positions in the molecule [28]. Changing these residues generally results in large changes in binding.

with their target enzymes. However, even within each family, where the global three-dimensional structures are very similar for all members, some are strong inhibitors of trypsin, some of chymotrypsin, some of elastase and some of both chymotrypsin and elastase. Other members of the same family are weak inhibitors of all of the enzymes tested thus far. It is in such situations that family-wide algorithms are needed. We chose the Kazal family [7,8] of protein inhibitors (Figure 2) as the object of our algorithm (Figure 1). However, it is quite likely that similar algorithms can be developed for several of the 18 families; several groups listed at the end of this review have gathered large amounts of data that will be helpful for this.

How many members are there in a protein family? There are two dramatically different answers to this. One is the number of natural members, that is, the number of

paralogs [11], in each species summed over all extant species. The other, much larger, number is the sum of all possible members of the family. It is the large number that we care about. The number appears intractable, but can be made tractable by two major approximations. One is that only those parts of the inhibitor that make contact with the enzyme need to be considered. The other, and more crucial one, is additivity.

### Additivity

When scientists think about an enormously greater number of objects than can be carefully and individually examined, they need a principle for estimating the properties of the untested objects. In thermodynamics, the most broad-based application of such a principle is additivity. Kistiakowsky [12] used additivity to show that the heat of hydrogenation of cyclohexadiene is almost exactly twice the heat of hydrogenation of cyclohexene. However, the heat of hydrogenation of cyclohexatriene (benzene) is much less than would be expected. The difference between the enthalpy of hydrogenation of benzene and the expected enthalpy of hydrogenation of three double bonds is now called the resonance stabilization energy of benzene. Further extensions to the additivity-based prediction of the free energy, enthalpy and entropy of formation of organic compounds are given by Benson [13].

The Kistiakowsky experiments illustrate the two applications of additivity. The first and the simplest is as a predictive tool. This is what we mostly describe in this review. The other is the measurement of interaction energy, whereby additivity is used to subtract the predicted quantity from the measured one. A classic paper by Fersht and co-workers [14] uses double-mutant cycles to measure the electrostatic stabilization of a helix in barnase into which an Asp12–Arg16 pair was introduced by mutation. The interaction between Asp12 and Arg16 is broken up by either the single mutation Asp12Ala or Arg16Thr. The double-mutant cycle is completed by making the Asp12Ala/Arg16Thr double mutant. As expected, the cycle is nonadditive at low ionic strength, where Asp12 and Arg16 interact. As the ionic strength is increased, the charges are shielded and the cycle becomes additive.

### Nonadditivity is a change in interaction

For the reaction  $A \leftrightarrow B$ , we measure  $\Delta G_{\text{wt}}^{\circ}$ , which is equivalent to measuring an equilibrium constant. Now, let us make a substitution in A to XA. The reaction now is  $XA \leftrightarrow XB$  and we measure  $\Delta G_X^{\circ}$ . A substitution at another position in A yields  $AY \leftrightarrow BY$  and  $\Delta G_Y^{\circ}$ . Now, let's make both substitutions together, yielding  $XAY \leftrightarrow XBY$  and  $\Delta G_{X,Y}^{\circ}$ . What can we say about this last term in relation to the previous three? It seems clear that, if the interaction energy between X and Y in XAY, if any, is unchanged in XBY, then the system will be additive and  $\Delta G_{X,Y}^{\circ} = \Delta G_X^{\circ} + \Delta G_Y^{\circ} - \Delta G_{\text{wt}}^{\circ}$ . If the interaction between

X and Y changes, an additional term for the change must be introduced.

The most popular application of additivity is to the thermodynamics of unfolding of individual proteins, where  $N \leftrightarrow U$  yields  $\Delta G_{\text{wt}}^{\circ}$  and  $XNY \leftrightarrow XUY$  yields  $\Delta G_{X,Y}^{\circ}$ . Virtually all investigators make the generally tacit assumption that any interactions between X and Y in XUY are energetically insignificant. The unfolding reaction is said to be additive if X and Y do not interact in XNY. It is said to be nonadditive if X and Y interact in the native form. This gives rise to the somewhat incorrect slogan 'interaction is nonadditive', because it is the change in interaction energy upon reaction that should be emphasized.

With two molecules of reactant, we can modify either only one or both. The latter is emphasized in the classic paper on the interaction of barnase with barstar [15]. For the unmodified wild types,  $A + B \leftrightarrow C$  yields  $\Delta G_{\text{wt}}^{\circ}$ , as before. In the double-mutant case, where each reactant has one mutation,  $XA + BY \leftrightarrow XCY$  yields  $\Delta G_{X,Y}^{\circ}$ , which can be predicted from the wild-type and the two single-mutant experiments if and only if there is no interaction energy between X and Y in XCY. Clearly, there is no interaction between X and Y for the unbound reactant molecules. The nonadditive term is just a measure of the interaction energy between X and Y in C. Generally, if X and Y are close to one another, there will be such a term. This conclusion is so powerful that the unknown structure of a receptor was mapped by such experiments [16\*\*].

In contrast to the above strategy for the measurement of interaction energies and the determination of structure, making both substitutions in the same reactant molecule is most useful for reactivity prediction and protein (drug) design. In such a system [17\*\*],  $\Delta G_X^{\circ}$  and  $\Delta G_Y^{\circ}$  for  $XA + B \leftrightarrow XC$  and  $AY + B \leftrightarrow CY$  are compared to  $\Delta G_{X,Y}^{\circ}$  for  $XAY + B \leftrightarrow XCY$ . It is clear that this will be nonadditive if the interaction energy between X and Y changes in XCY and XAY. Such an effect can be direct, whereby the contact between X and Y changes (e.g. the length of a hydrogen bond between them), or indirect, whereby binding of X remodels the binding pocket for Y and *vice versa*.

It seems clear that this approach will find far more additivity than the one in which a single modification is made in both reactants. Although some authors clearly realize the difference between the two approaches, others do not. Thus, many critics cite the first approach, for which nonadditivity is common, as an objection to the second approach, for which it is relatively rare.

### Setting up the algorithm

A detailed analysis of protein–protein associations [18\*] shows that they can be readily divided into rigid-body

associations and associations accompanied by conformational changes. We expected that the rigid-body associations would be more additive because they involve fewer contact residues. The typical number of contact residues is twenty for each partner. We were delighted to find that a standard-mechanism canonical [8,9] inhibitor of serine proteinases, turkey ovomucoid third domain (OMTKY3; see Figure 2), has only twelve consensus contact residues with the six serine proteinases we selected for study. Of these residues, two have a structural role and almost never varied during the evolution of the Kazal inhibitor family to which OMTKY3 belongs. The family has many members, including pancreatic secretory trypsin inhibitors (PSTIs), acrosin inhibitors from seminal plasma, two-domain submandibular inhibitors in carnivores, three-domain ovomucoids in birds and four-domain ovomucoids in alligators, seven-domain ovoinhibitors, nine-domain agrin and fifteen-domain LEKTI.

We chose to study OMTKY3 and its variants, and generated a sequence to reactivity algorithm for the entire Kazal family of inhibitors with the six serine proteinases selected (listed in Figure 1). To this end, we expressed all 19 possible single-coded variants of each of the 10 variable contact positions (in blue in Figure 2). This yielded the 190-variant set plus the wild type. We were extremely fortunate that all 191 variants could be expressed. This is in part because the contact residues in standard-mechanism canonical inhibitors are largely on the surface. The association equilibrium constants ( $K_{\text{assoc}}$ ) of all 191 OMTKY3 variants with the six enzymes listed in Figure 1 could all be measured by a direct procedure. The dynamic range of the measurements for  $K_{\text{assoc}}$  is  $10^3 \text{ M}^{-1}$  to  $10^{13} \text{ M}^{-1}$ , which is probably the widest known for any protein–protein interaction. For the OMTKY3  $P_1$  position with bovine chymotrypsin A $\alpha$  (CHYM), the range of  $K_{\text{assoc}}$  values is slightly greater than 8 orders of magnitude ( $K_{\text{assoc}}$  is  $8.1 \times 10^{12} \text{ M}^{-1}$  for  $P_1$  Tyr and  $6.7 \times 10^4 \text{ M}^{-1}$  for  $P_1$  Pro). Ranges of 5–6 orders of magnitude are not uncommon for several other enzyme–OMTKY3 position pairs, such as the subtilisin Carlsberg (CARL)– $P_4$  pair. It is thus very easy to imagine that some single-variant equilibrium constants would fall out of the measurement range. Specifically, the  $P_1$  range for bovine cationic  $\beta$ -trypsin is 9 orders of magnitude and for salmon trypsin is 11 orders of magnitude for their interaction with bovine pancreatic trypsin inhibitor (BPTI) [19]. For both enzymes, the authors could not measure the strongly interacting  $P_1$  arginine and lysine variants directly, but estimated them from measurements at pH 5.0 and extrapolation. In an interscaffolding additivity study, we failed to measure some eglin c  $P_1$  variants interacting with *Streptomyces griseus* proteinases A and B (SGPA and SGPB), because they interacted more strongly than  $10^{13} \text{ M}^{-1}$ , and some eglin c variants with porcine pancreatic elastase (PPE), because they were weaker than  $10^3 \text{ M}^{-1}$  [20]. Most recently, a set of 11 multiple variants

of OMTKY3 designed previously to inhibit furin was used in an additivity test with the six enzymes listed in Figure 1. However, 28 of 66 possible interactions were too weak for exact  $K_{\text{assoc}}$  measurement.

For each enzyme, the 191  $K_{\text{assoc}}$  values can now be converted to  $\Delta G^\circ$  values by  $-RT \ln K_{\text{assoc}}$  and the  $\Delta G_{\text{wt}}^\circ$  value is retained. The remainder are converted to  $\Delta \Delta G^\circ (X_{\text{wt}} i X)$  values by subtracting  $\Delta G_{\text{wt}}^\circ$ , thus generating one  $\Delta G_{\text{wt}}^\circ$  and 190  $\Delta \Delta G^\circ (X_{\text{wt}} i X)$  values. Here,  $i$  indicates the position in Figure 2. This set contains all the information needed to calculate the  $\Delta G^\circ$  predicted for any Kazal inhibitor. On the other hand, as it is a nonredundant set, there are relatively few self-consistency checks available.

### Testing the algorithm

We now obtain [17\*\*],

$$\Delta G_{\text{predicted}}^\circ = \Delta G_{\text{wt}}^\circ + \sum_{i=1}^{i=10} \Delta \Delta G^\circ (X_{\text{wt}} i X) \quad (1)$$

$$\Delta G_{\text{I}}^\circ = \Delta G_{\text{predicted}}^\circ - \Delta G_{\text{measured}}^\circ \quad (2)$$

where  $\Delta G_{\text{I}}^\circ$  is the free energy of interactions. It is, in fact, an error in the prediction term. This error term has three components: the difference in the free energies of interaction between the contact residue in the complex and in the unbound inhibitor — we call it the nonadditivity term for short; the contributions of changes, if any, of residues at positions other than the contact positions (the white positions in Figure 2); and the experimental errors both in the  $\Delta G_{\text{predicted}}^\circ$  and in the  $\Delta G_{\text{measured}}^\circ$ . It is only the last term that we are able to estimate by statistical analysis. As the predictions are based on an experimental database, the predictions involve experimental error and these errors increase with the number of terms in the sum of Equation 1. The propagated error rises from 0 to 10 substitutions [17\*\*]. In fact, the number of substitutions is limited to 9 by a restriction that demands that predictions are limited to those cases which differ by only one substitution from the  $P_2$  Thr17– $P_1'$  Glu19 hydrogen-bonded pair of residues present in wild-type OMTKY3. The detailed reasons for this temporary restriction are summarized in the last part of this review. We arbitrarily call all the cases where  $|\Delta G_{\text{I}}^\circ| < 2\sigma$  of the expected propagated experimental error as additive. We call cases where  $2\sigma < |\Delta G_{\text{I}}^\circ| < 4\sigma$  partially additive and cases where  $|\Delta G_{\text{I}}^\circ| > 4\sigma$  nonadditive. Our laboratory had a large (92 different sequences) set of natural avian ovomucoid third domains [21]. This should potentially give us  $92 \times 6$  (our target enzymes) = 552 equilibrium constants. However, two factors intervened. One was the restriction stated above. The other was that some equilibrium constants were too small to measure. This left us with 398 equilibrium constants (see Tables 1 and 2 of Lu *et al.* [17\*\*]). In the same paper, 17 equilibrium constants of natural Kazal inhibitors other than the ovomucoid third domain were also studied. These give better results than

the ovomucoid third domains, in spite of the fact that there are many differences between these inhibitors and OMTKY3 in the noncontact region. Combining these results with those for the furin inhibitors mentioned earlier gives a total of 450 equilibrium constants for Kazal inhibitors interacting with the six target enzymes (Figure 1). Of these, 289 (64%) are additive, 119 (26%) are partially additive and 42 (9%) are nonadditive. Other criteria, such as the average absolute error is only 500 cal/mol, suggest that the restricted algorithm works well. More tests are clearly needed and these are underway, as is an ongoing effort to remove the restriction.

## Designs

Subject to restriction, we can answer questions such as how strong is the strongest possible Kazal inhibitor. This is a particularly simple task, as selecting the best possible residue at each position suffices in an additive system. Knowledge of the awesome strength of the best possible inhibitors may be of use in drug design. For the study of the pK values of groups both on the enzyme and on the inhibitor, some of which undergo very large changes upon complex formation [22,23\*\*], it would be most useful to design inhibitors with very high  $K_{\text{assoc}}$  values that form highly stable complexes at low pH values. This can be accomplished by including the pH dependence of  $\Delta\Delta G^\circ$  [23\*\*] and incorporating such pH-dependent terms into Equation 1. The somewhat opposite goal of making the complexes very unstable at low pH but not at neutral pH might be useful for affinity chromatography separation of enzymes.

As is always the case, the most challenging design problems involve increasing the specificity. However, as the role of many inhibitors appears to be defense against bacterial or insect predators, lack of specificity, at least for families of serine proteinases, has probably been nature's design goal.

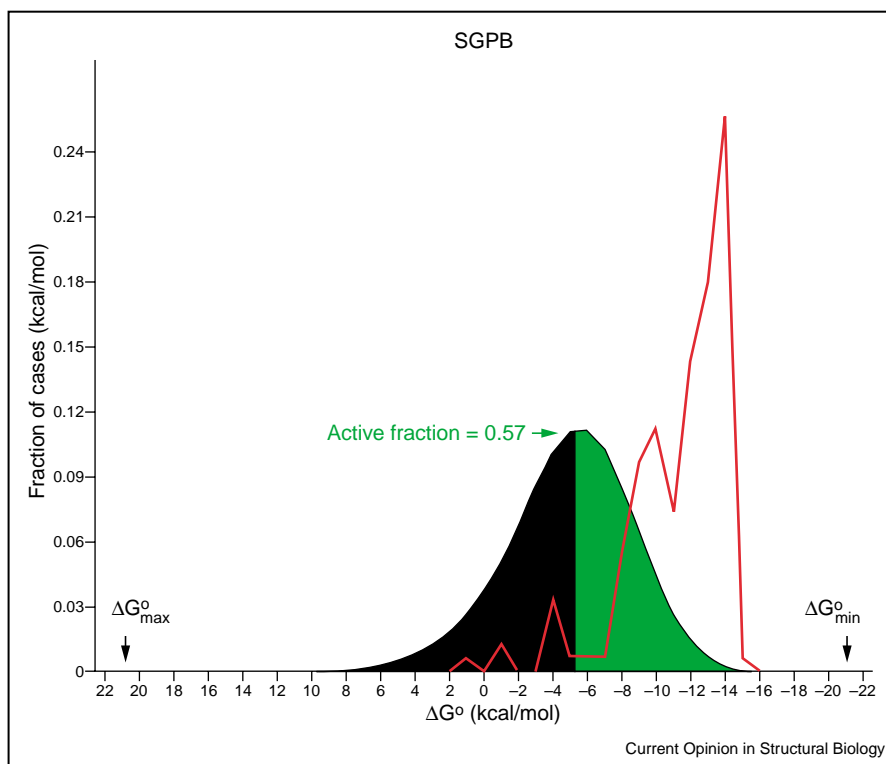
## The distribution function

Equation 1 can be applied to any of the  $20^{10}$  possible sequences of the blue residues in Figure 2. However, the restriction discussed both above and at greater length below reduces the  $20^{10}$  ( $\sim 10^{13}$ ) to  $10^{12}$ . Neither of these numbers is the total possible number of Kazal inhibitors, as we have not yet dealt in detail with the white residues in Figure 2. These consist of two types. The first are the evolutionarily unvaried residues, such as the six cysteine residues involved in disulfide bridges, and several nearly unvaried positions. These define the membership of the sequence in the Kazal family. The remaining white positions have been shown [17\*\*] to accept changes without changing the  $K_{\text{assoc}}$  values of the inhibitor. A rough estimate is that there are at least  $10^{50}$  possible members of the Kazal family to which the distribution function applies. Within this set,  $10^{13}$  differ in inhibitory strength or specificity.

Even though it is technically possible to calculate values for  $10^{12}$  or  $10^{13}$  variants, it is not necessary. Excellent additivity-based approximations have been introduced that change the calculation from cumbersome to rapid. The results of the calculation for SGPB are shown in Figure 3. The characteristics of the six enzymes we studied are given in Table 1a. Figure 3 gives us the range (40–50 kcal/mol) and informs us that the distribution function is largely symmetric. The green part of the distribution contains the variants that associate more strongly than  $\Delta G^\circ = -4$  kcal/mol, which is the lower limit of measurement ( $K_{\text{assoc}} = 10^3 \text{ M}^{-1}$ ). The upper limit of measurement is  $-17.5$  kcal/mol ( $K_{\text{assoc}} = 10^{13} \text{ M}^{-1}$ ). The fraction of  $\Delta G^\circ$  values above that limit is very small, although, because of the large total numbers, there are many such inhibitors in this set. We like to refer to the black side of the distribution as the dark side of the moon, as it deals with currently experimentally inaccessible situations. In all cases in which the interaction is measurable, the residue in the  $P_1$  position enters the  $S_1$  pocket of the enzyme, however deleterious the  $P_1$ – $S_1$  interaction is. This was amply demonstrated by X-ray crystallographic studies of SGPB–X<sup>18</sup> OMTKY3 variants [24,25], where both the  $P_1$  Leu (best) variant and  $P_1$  Pro (worst) variant bind to the  $S_1$  cavity of the enzyme. Even more impressive are the X-ray results of Helland *et al.* [26], where both the best (arginine) and worst (glycine) variants enter the  $S_1$  cavity of bovine trypsin. Thus, for these inhibitors, the docking [27\*] to the enzyme never varies. This constancy must be achieved at the price of unfavorable interactions at most of the other positions. However, when we make several positions strongly deleterious, a change in docking must occur. Thus, inhibitors that inhibit only in the gigamolar range are avoided, suggesting that at least part of the dark area of Figure 3 must be wrong. It is startling how large the fraction of the green area (the active fraction) is for all the enzymes (Table 1a). The fraction of all possible Kazal inhibitors that inhibit at least one of the six enzymes is 0.76. This is subject to the already stated restriction.

Having concluded the analysis of the distribution of all possible Kazal inhibitors, we now introduce in red in Figure 3 the distribution function calculated for avian ovomucoid third domains for 153 species (about 2% of the extant species) of birds. Each bird species has only one ovomucoid; ovomucoids are an orthologous [11] group. The detailed function of many proteinase inhibitors is not known. Additionally, many inhibitors, especially ovomucoid third domains, exhibit great hypervariability of their proteinase-binding regions [28]. Therefore, some evolutionists concluded that this rapid variation was a reflection of the drift of exposed residues during evolution and that inhibitors acquire their inhibitory activity as a consequence of this drift [29]. If this were so, the means of the  $\Delta G^\circ$  values of all possible Kazal inhibitors and of the natural set would be the same. Clearly they are not. Nature evolved avian

Figure 3



Distribution of the standard free energies of association for all possible members of the Kazal family, subject to restriction, interacting with SGPB. Note the unconventional direction of the x-axis. This was done to position strong binding on the right and weak binding on the left. The  $\Delta G_{\min}^{\circ}$  is the value for the strongest possible SGPB inhibitor in the Kazal family and  $\Delta G_{\max}^{\circ}$  is the weakest possible, although this calculated value is unreasonable (see Table 1a and text). The lower limit of measurement is  $-4$  kcal/mol. The green area comprises 0.57 of the total. The black area is where the interactions are too weak to measure. However, the calculated values there are doubtful. The red curve is for ovomucoid third domains from 153 different species of birds. Cases where several species shared the same sequence were given a statistical weight of the number of species that share the sequence. This distribution is also calculated, because the lowest values (in the black area) could not be measured. The black-green and red curves are normalized to unit area.

ovomucoid third domains to inhibit serine proteinases with a specificity similar to our six target enzymes. With the exception of PPE, where the difference is small, there is a large difference between the  $\Delta G^{\circ}$  of the best found inhibitor (Table 1b) and  $\Delta G_{\min}^{\circ}$  — the best of all possible inhibitors. The ovomucoid third domain curves greatly differ in shape from the all possible Kazal curve. For the ovomucoid third domain, the strongest found inhibitor is very near the most probable inhibitor. For the all Kazal curve, the  $\Delta G_{\min}^{\circ}$  and the most probable  $\Delta G^{\circ}$  (peak of the distribution curve) differ by about 15 kcal/mol. Why is that so? As pointed out [30<sup>\*</sup>], extremely strong inhibitors may be undesirable because the life-time of their complexes is too long. Such complexes may not have enough time to be turned over.

There are, however, many other plausible explanations. The very strong inhibitors may not be avoided, but, if somewhat weaker ones suffice, evolutionary effort need not be expended to achieve the strongest possible. The

six enzymes listed in Figure 1 are only proxy enzymes, not the real targets. Clearly, the situation may be different for the appropriate enzymes. It can be demonstrated that highly nonspecific inhibitors are much weaker than the strongest possible, for the various enzymes they inhibit. As the inhibitor's role is often defensive, low rather than high specificity may be needed. It is our favorite explanation of the less than optimal reactivity.

### The nonadditivity and the restriction

Relatively early in this very long study, we realized that, in OMTKY3 and in a large fraction of avian ovomucoid third domains, the  $P_2$  residue is Thr17 and the  $P_1'$  residue is Glu19. These two residues form a weak 3.0 Å hydrogen bond in the unbound inhibitor [31]. Formation of a complex with CHYM [32] shortens this bond to 2.5 Å. We already pointed out that nonadditivity arises from changes in the interaction between the residues in question in products (complexes) and reactants (unbound enzyme and inhibitor). As the interaction in question

**Table 1**

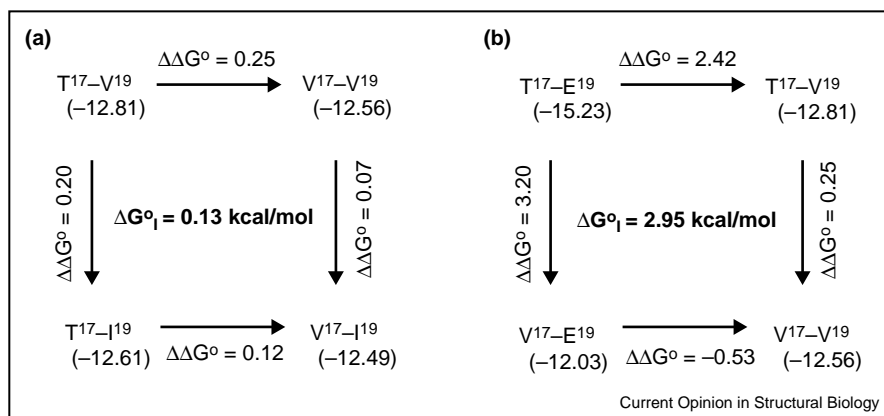
| Enzyme   | $\Delta G_{\min}$<br>(kcal/mol) | $\Delta G_{\max}$<br>(kcal/mol) | Active<br>fraction (%) <sup>†</sup> | Mean<br>(kcal/mol) | Standard<br>deviation (kcal/mol) | Skewness | Kurtosis |
|--|---------------------------------|---------------------------------|-------------------------------------|--------------------|----------------------------------|----------|----------|
| (a) Distribution for all possible sequences subject to restrictions*.        |                                 |                                 |                                     |                    |                                  |          |          |
| CHYM   | -23.69                          | 27.12                           | 28                                  | -1.38              | 4.61                             | 0.03     | -0.02    |
| PPE  | -18.44                          | 25.38                           | 21                                  | 0.23               | 4.96                             | 0.14     | -0.30    |
| CARL   | -22.83                          | 25.59                           | 45                                  | -3.24              | 4.48                             | 0.25     | 0.18     |
| SGPA   | -21.40                          | 21.32                           | 63                                  | -5.11              | 3.95                             | 0.34     | 0.28     |
| SGPB   | -20.78                          | 21.07                           | 57                                  | -4.40              | 3.75                             | 0.46     | 0.50     |
| HLE  | -22.07                          | 22.85                           | 31                                  | -1.63              | 4.52                             | 0.04     | -0.11    |
| (b) Distribution for avian ovomucoid third domains subject to restrictions*. |                                 |                                 |                                     |                    |                                  |          |          |
| CHYM   | -16.25                          | 1.16                            | 93                                  | -11.30             | 3.90                             | 1.19     | 0.83     |
| PPE  | -15.90                          | 1.87                            | 95                                  | -12.21             | 3.79                             | 2.14     | 4.35     |
| CARL   | -15.23                          | 0.81                            | 95                                  | -11.73             | 3.58                             | 1.56     | 1.68     |
| SGPA   | -16.37                          | 0.35                            | 98                                  | -12.94             | 3.25                             | 1.78     | 3.33     |
| SGPB   | -15.18                          | 1.50                            | 98                                  | -11.78             | 3.00                             | 1.79     | 3.80     |
| HLE  | -15.52                          | -0.91                           | 96                                  | -11.20             | 3.26                             | 1.65     | 2.20     |

\*Parameters for the distribution function of the standard free energies of interaction of the six selected enzymes with (a) all possible Kazal inhibitors and (b) avian ovomucoid third domains from 153 species of birds. Figure 3 is a plot of the same data for SGPB.  $\Delta G_{\min}^{\circ}$  is for the strongest inhibitor in the group. In (a), this is the strongest possible inhibitor.  $\Delta G_{\max}^{\circ}$  is for the worst inhibitor in the group. The difference between these is the range. The active fraction is the fraction with  $\Delta G^{\circ}$  below  $-4$  kcal/mol. The mean  $\Delta G^{\circ}$  is much more negative for (b) than for (a), indicating that ovomucoid third domains evolved to inhibit enzymes with similar specificity to those we selected. The relatively high standard deviation in (b) is an indication of hypervariability (see text). In other systems, such as triosephosphate isomerase, the standard deviation in  $k_{\text{cat}}/K_m$  of a natural set is about zero. Ovomucoid third domains show a very pronounced skewness (easily seen in the red curve of Figure 3). We have not as yet analyzed the kurtosis, which is very large but varies from enzyme to enzyme. Unfortunately, all the results are still subject to restriction (see text). <sup>†</sup> $\Delta G < -4$  kcal/mol.

was a sidechain–sidechain hydrogen bond, we concluded that both  $P_2$  and  $P'_1$  must be polar for nonadditivity to manifest itself. We therefore constructed several additivity cycles without any polar  $P_2$ –polar  $P'_1$  pairs. They were found to be additive. Figure 4a shows such an additivity cycle where  $\Delta G_1^{\circ} = 130$  cal/mol. Errors less than 400 cal/mol in such a cycle indicate additivity. We then started the cycle with our wild-type  $P_2$  Thr17– $P'_1$  Glu19 pair in OMTKY3 and in two steps changed it to a  $P_2$  nonpolar Val– $P'_1$  nonpolar Val pair (Figure 4b). This is clearly highly nonadditive.

### Application to other systems

There are currently 18 known families of standard-mechanism canonical protein inhibitors of serine proteinases [7]. It seems quite likely that, ultimately, similar algorithms will be developed for all of them. Large amounts of equilibrium constant data were acquired by Miura and co-workers [33,34] for the *Streptomyces* subtilisin inhibitor (SSI) family, by Otlewski and co-workers [19,35] for the BPTI family and by the Leatherbarrow group [36\*,37,38] for Bowman–Birk inhibitors. In our opinion, the biggest advance towards an additivity-based

**Figure 4**

Double-variant cycles for  $P_2$ – $P'_1$ . The additivity criterion for such cycles is  $|\Delta G_1^{\circ}| < 0.4$  kcal/mol. All the cycles discussed are for CHYM. (a) The cycle involves nonpolar–nonpolar pairs (e.g.  $V^{17}$ – $V^{19}$  and  $V^{17}$ – $I^{19}$ ) and mixed pairs (e.g.  $T^{17}$ – $V^{19}$  and  $T^{17}$ – $I^{19}$ ). As can be seen, this cycle is additive. Several cycles without polar–polar pairs were constructed and all were additive. (b) This cycle contains wild-type OMTKY3, which has a  $T^{17}$ – $E^{19}$  pair. The sidechains of these two residues are hydrogen bonded to each other in unbound OMTKY3. This hydrogen bond becomes much stronger in complex with CHYM. As expected, this cycle is strongly nonadditive (see text).

algorithm is the analysis of BPTI by alanine shaving [39•]. Additivity was also used in a study of the binding of peptides to MHC type 2 antigen [40•].

The development of an algorithm for the other 17 families of protein inhibitors of serine proteinases would be a trivial task if the data set already determined for Kazal inhibitors could be directly used for them at all the corresponding Schechter–Berger positions (Figure 2). Such an outcome would be called complete interscaffolding additivity. Alas, it is not completely true. It was most thoroughly tested for the P<sub>1</sub> residue, which was most likely to be successful. This is because, in all inhibitor families, the P<sub>1</sub> residue is very highly exposed in the unbound form. In complex, it becomes buried in the S<sub>1</sub> cavity of the proteinase. Extensive comparisons were carried out between the Kazal family and the potato I family [20] for the six enzymes listed in Figure 1. Even more extensive comparisons were done between the Kazal and BPTI (Kunitz) families for CHYM and human leukocyte elastase (HLE) [19]. For most residues,  $\Delta\Delta G^\circ$  values relative to glycine agree within experimental error. Proline is a significant exception in the potato I/Kazal comparison. Even larger exceptions are arginine and lysine when their binding to CHYM is compared for the Kazal and BPTI families. A structural explanation for this was provided by X-ray crystallography of complexes of BPTI (a trypsin inhibitor) with CHYM [41,42] and of a complex of a OMTKY3 mutant with lysine at P<sub>1</sub> interacting with CHYM [43]. The P<sub>1</sub> lysine in the Lys18 OMTKY3 mutant has its lysine residue in the ‘down’ position, extending into the S<sub>1</sub> pocket of CHYM. In the BPTI complex, the lysine residue is in the ‘up’ position, with the lysine positive charge interacting with the enzyme–inhibitor interface [41,42].

### Why was the Kazal family so successful?

Why did the Kazal inhibitor family interacting with enzymes of chymotrypsin-like, elastase-like and subtilisin-like specificity prove so receptive to the development of the additivity-based sequence to reactivity algorithm? The disingenuous answer is that comparable effort was not expanded on any other system. However, that is not all. This was and is a lucky system. But because comparisons are scarce, our guesses about the reasons seem tentative. One reason is clearly the rigid-body association between enzymes and Kazal inhibitors. The ability to reduce the problem to only ten variable contact residues seems lucky (see [18•] for more typical numbers). The very open reactive site region of Kazal inhibitors is in contrast to the more crowded one in the BPTI (Kunitz) family and, therefore, eliminates interference between contact residues. Replacement of P<sub>1</sub> Gly16 or P<sub>1</sub> Ala16 by larger residues in BPTI leads to a large decrease in inhibitory activity due to steric conflict between the large P<sub>1</sub> residues and the P<sub>3</sub> Ile18 of the inhibitor [35]. An X-ray structure shows that, in complex with trypsin, the effect of P<sub>1</sub> Leu16 propagates by

interfering with P<sub>3</sub> Ile18 and, in turn, with Tyr39, indicating a crowded reactive site region.

In the Kazal inhibitors associating with the enzymes studied, the interactions are predominantly hydrophobic and therefore, in contrast to purely electrostatic interactions, short range. In a very recently submitted paper (MA Qasim *et al.*, unpublished data), we show that electrostatic interactions between clusters of negative charges on one protein and positive charges on another are highly nonadditive. Probably the most important reason for the success of the algorithm is the very large changes in the energy of interaction upon single substitutions. The average range at the ten variable contact positions is 4–5 kcal/mol per position. We believe that this is much larger than for other protein–protein associations, but few investigators report all 20 amino acid residues at each of several positions.

### Conclusions

Additivity-based prediction is still in its early stages. However, it appears that most protein chemists over-rate the role of nonadditivity. It occurs, but seldom enough that additivity-based systems can be developed and then improved by making nonadditivity corrections. We learned the most from comparing the standard free energy of association distributions for natural proteins with those for all possible proteins. It seems to us difficult to learn much about protein reactivity without keeping these two sets separated.

### Acknowledgements

We are grateful to MC Laskowski, C Bailey-Kellogg and K Dill for suggesting approximations for the distribution function, for implementing such approximations and for discussion of the comparison of all possible distributions versus natural variant distributions. Work in our laboratory was supported by National Institutes of Health grants GM10831 and GM63539.

### References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Anfinsen CB: **Principles that govern the folding of protein chains.** *Science* 1973, **181**:223-230.
  2. Merrifield B: **Solid phase synthesis.** *Science* 1986, **232**:341-347.
  3. Im H, Woo M-S, Hwang KY, Yu M-H: **Interactions causing the kinetic trap in serpin protein folding.** *J Biol Chem* 2002, **277**:46347-46354.
  4. Jaswal SS, Sohl JL, Davis JH, Agard DA: **Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability.** *Nature* 2002, **415**:343-346.
- A beautiful analysis of the folding of  $\alpha$ -lytic proteinase. The authors state that the unfolded mature protein is not only kinetically trapped, but is slightly more stable than the folded form.
5. Galperin MY, Koonin EV: **Comparative genome analysis.** *Methods Biochem Anal* 2001, **43**:359-392.
  6. Blacklow SC, Liu KD, Knowles JR: **Stepwise improvements in catalytic effectiveness: interdependence in combinations of point mutations of a sluggish triosephosphate isomerase.** *Biochemistry* 1991, **30**:8470-8476.

7. Laskowski M Jr, Qasim MA: **What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?** *Biochim Biophys Acta* 2000, **1477**:324-337.
8. Laskowski M Jr, Kato I: **Protein inhibitors of proteinases.** *Annu Rev Biochem* 1980, **49**:593-626.
9. Bode W, Huber R: **Natural protein proteinase inhibitors and their interaction with proteinases.** *Eur J Biochem* 1992, **204**:433-451.
10. Schechter I, Berger A: **On the size of the active site in proteases I papain.** *Biochem Biophys Res Commun* 1967, **27**:157-162.
11. Fitch WM: **Homology, a personal view on some of the problems.** *Trends Genet* 2000, **16**:227-231.
12. Conant JB, Kistiakowsky GB: **Energy changes involved in the addition reactions of unsaturated hydrocarbons.** *Chem Rev* 1937, **20**:181-194.
13. Benson SW: *Thermochemical Kinetics. Methods for the Estimation of Thermochemical Data and Rate Parameters.* John Wiley and Sons; 1968.
14. Serrano L, Horowitz A, Avron B, Bycroft M, Fersht AR: **Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double mutant cycles.** *Biochemistry* 1990, **29**:9343-9352.
15. Schreiber G, Fersht AR: **Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles.** *J Mol Biol* 1995, **248**:478-486.
16. Roisman LC, Piehler J, Trosset J-Y, Scheraga HA, Schreiber G:
  - **Structure of the interferon-receptor complex determined by distance constraints from double-mutant cycles and flexible docking.** *Proc Natl Acad Sci USA* 2001, **98**:13231-13236.
 This is a powerful application of double-mutant cycles whereby each partner has one mutation. However, the final product is not the energy of interaction, but the structure of the receptor to which a protein ligand is bound.
17. Lu SM, Lu W, Qasim MA, Anderson S, Apostol I, Ardelt W, Bigler T,
  - Yi WC, Cook J, James MNG *et al.*: **Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases.** *Proc Natl Acad Sci USA* 2001, **98**:1410-1415.
 The additivity-based prediction of standard free energies of association is proposed and applied to a restricted subset of all possible members of the Kazal inhibitor family. The results are extensively tested. Some natural multiple variants are clearly nonadditive, but a very large fraction are additive or partially additive.
18. Janin J: **Kinetics and thermodynamics of protein-protein interactions.** In *Protein-Protein Recognition*. Edited by Kleanthous C. Oxford: Oxford University Press; 2000:1-32.  
 This chapter reviews mainly X-ray crystallographic findings on protein-protein interactions. A distribution function in it (Figure 1) neatly divides rigid-body associations from those involving a conformational change in one or in both protein partners. The estimate of contact residues is 20 for each partner. Our 10 made our graduate students happy.
19. Krowarsch D, Dadlez M, Buczek O, Krokszyńska I, Smalas AO, Otlewski J: **Interscaffolding additivity: binding of P<sub>1</sub> variants of bovine pancreatic trypsin inhibitor to four serine proteinases.** *J Mol Biol* 1999, **289**:175-186.
20. Qasim MA, Ganz PJ, Saunders CW, Bateman KS, James MNG, Laskowski M Jr: **Interscaffolding additivity. Association of P<sub>1</sub> variants of eglin c and of turkey ovomucoid third domain with serine proteinases.** *Biochemistry* 1997, **36**:1598-1607.
21. Apostol I, Giletto A, Komiyama T, Zhang W, Laskowski M Jr: **Amino acid sequences of ovomucoid third domains from 27 additional species of birds.** *J Prot Chem* 1993, **12**:419-433.
22. Qasim MA, Ranjbar MR, Wynn R, Anderson S, Laskowski M Jr: **Ionizable P<sub>1</sub> residues in serine proteinase inhibitors undergo large pK shifts on complex formation.** *J Biol Chem* 1995, **270**:27419-27422.
23. Laskowski M, Qasim MA, Lu SM: **Interaction of standard mechanism, canonical protein inhibitors with serine proteinases.** In *Protein-Protein Recognition*. Edited by Kleanthous C. Oxford: Oxford University Press; 2000:228-279.  
 A comprehensive review dealing with the energetics of binding of OMTKY3 and a large number of its variants with different serine proteinases. Both intrascaffolding and interscaffolding additivities are discussed, and the use of the former in the design of the strongest possible and most specific inhibitor is discussed in detail.
24. Huang K, Lu W, Anderson S, Laskowski M Jr, James MNG: **Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu<sup>18</sup>, Ala<sup>18</sup>, and Gly<sup>18</sup> variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B.** *Protein Sci* 1995, **4**:1985-1997.
25. Bateman KS, Huang K, Anderson S, Lu W, Qasim MA, Laskowski M Jr, James MNG: **Contribution of peptide bonds to inhibitor-protease binding: crystal structures of the turkey ovomucoid third domain backbone variants OMTKY3-Pro18I and OMTKY3-Ψ [COO] – Leu18I in complex with *Streptomyces griseus* proteinase B (SGPB) and the structure of the free inhibitor, OMTKY3 - Ψ [CH<sub>2</sub> NH<sub>2</sub><sup>+</sup>] – Asp19I.** *J Mol Biol* 2001, **305**:839-849.
26. Helland R, Otlewski J, Sundheim O, Dadlez M, Smalas AO: **The crystal structures of the complexes between bovine beta trypsin and ten P<sub>1</sub> variants of BPTI.** *J Mol Biol* 1999, **287**:923-942.
27. Lorber DM, Udo MK, Shoichet BK: **Protein-protein docking with multiple residue conformations and residue substitutions.** *Protein Sci* 2002, **11**:1393-1408.  
 Massive docking calculations for several widely studied protein-protein interactions are carried out to test, among other effects, the docking of different conformers of contact residues. Additivity of conformer contributions was assumed. The method produces a correlation between the docking score and the experimental standard free energy of interaction. The results are excellent for X<sup>15</sup> BPTI-trypsin, very good for X<sup>18</sup> OMTKY3-chymotrypsin and not good at all for X<sup>18</sup> OMTKY3-SGPB.
28. Laskowski M Jr, Kato I, Ardelt W, Cook J, Denton A, Empie MW, Kohr WJ, Park SJ, Parks K, Schatzley BL *et al.*: **Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues.** *Biochemistry* 1987, **26**:202-221.
29. Graur D, Li W-H: **Evolution of protein inhibitors of serine proteinases: positive Darwinian selection or compositional effects.** *J Mol Evol* 1988, **28**:131-135.
30. Brooijmans N, Sharp KA, Kuntz ID: **Stability of macromolecular complexes.** *Proteins* 2002, **48**:645-653.  
 This data-mining paper examines natural protein-protein association and finds lower than expected strength of binding for the strongest interacting pairs. This is attributed to extrathermodynamic reasons. This is similar to the analysis of why the strongest observed ovomucoid third domains bind to their target enzymes less strongly than the best possible Kazal inhibitor.
31. Bode W, Epp O, Huber R, Laskowski M Jr, Ardelt W: **The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3).** *Eur J Biochem* 1985, **147**:387-395.
32. Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M Jr, James MNG: **Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution.** *J Mol Biol* 1987, **195**:397-418.
33. Kojima S, Nishiyama Y, Kumagai I, Miura K-I: **Inhibition of subtilisin BPN' by reactive site P<sub>1</sub> mutants of *Streptomyces subtilisin* inhibitor.** *J Biochem* 1991, **109**:377-382.
34. Masuda-Momma K, Hatanaka T, Inouye K, Kanaori K, Tamura A, Akasaka K, Kojima S, Kumagai I, Miura K-I, Tonomura B: **Interaction of subtilisin BPN' and recombinant *Streptomyces subtilisin* inhibitors with substituted P<sub>1</sub> site residues.** *J Biochem* 1993, **114**:553-559.
35. Grzesiak A, Helland R, Smalas AO, Krowarsch D, Dadlez M, Otlewski J: **Substitutions at the P-1' position in BPTI strongly affect the association energy with serine proteinases.** *J Mol Biol* 2000, **301**:205-217.
36. McBride JD, Watson EM, Brauer ABE, Jaulent AM, Leatherbarrow RJ: **Peptide mimics of the Bowman-Birk inhibitor reactive site loop.** *Biopolymers* 2002, **66**:79-92.

Recent advances in research work on synthetic cyclic peptides that mimic the reactive site of Bowman-Birk inhibitors are summarized. Comparison of the energetics of interaction of these peptides with serine proteinases shows that, in general, they correlate well with similar data obtained on serine proteinase inhibitors. This is also supported by their canonical conformation in complex with proteinases, which is characteristic of standard-mechanism inhibitors.

37. McBride JD, Brauer ABE, Nievo M, Leatherbarrow RJ: **The role of threonine in the P<sub>2</sub> position of Bowman-Birk proteinase inhibitors studies on P<sub>2</sub> variation in cyclic peptides encompassing the reactive site loop.** *J Mol Biol* 1998, **282**:447-458.
38. Gariani T, McBride JD, Leatherbarrow RJ: **The role of the P<sub>2</sub>' position of Bowman-Birk proteinase inhibitor in the inhibition of trypsin. Studies on P<sub>2</sub>' variation in cyclic peptides encompassing the reactive site loop.** *Biochim Biophys Acta* 1999, **1431**:232-237.
39. Buczek O, Koscielska-Kasprzak K, Krowarsch D, Dadlez M, Otlewski J: **Analysis of serine proteinase – inhibitor interaction by alanine shaving.** *Protein Sci* 2002, **11**:806-819.
- In this paper, up to eight contact residues of BPTI are changed to alanine. The free energy of association of these variants with chymotrypsin and trypsin is then studied. In seven out of ten of these variants, including the one with eight mutations, the measured free energy of association is the same, within experimental error, as the one calculated from the free energy of association of single alanine variants. The results clearly demonstrate that additivity holds well in BPTI-CHYM association. This paper is also a great rarity. Generally, equilibrium constants from different laboratories are not consistent. The single variants to alanine at eight positions were made and measured by Castro and Anderson [44]. The multiple variants were made and measured by the authors themselves. The consistency of the results is impressive.
40. McFarland BJ, Beeson C: **Binding interactions between peptides and proteins of the class II major histocompatibility complex.** *Med Res Rev* 2002, **22**:168-203.
- A comprehensive summary of the available literature on the interaction between peptides and MHCII proteins is presented in this article. Although many aspects of the interaction are poorly understood, it appears that the binding contribution of various contact residues in this interaction is independent, like inhibitor–proteinase interaction but not like antigen–antibody interaction.
41. Capasso C, Rizzi M, Menegatti E, Ascenzi P: **Bolognesi: Crystal structure of the bovine alpha-chymotrypsin: Kunitz inhibitor complex. An example of multiple protein:protein recognition sites.** *J Mol Recog* 1997, **10**:26-35.
42. Scheidig AJ, Hynes RT, Pelletier LA, Wells JA, Kossiakoff AA: **Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APP) and basic pancreatic trypsin inhibitor (BPTI).** *Protein Sci* 1997, **6**:1806-1824.
43. Qasim MA, Lu SM, Ding J, Bateman KS, James MNG, Anderson S, Song J, Markley JL, Ganz PJ, Saunders CW *et al.*: **Thermodynamic criterion for the conformation of P<sub>1</sub> residues of substrates and of inhibitors in complexes with serine proteinases.** *Biochemistry* 1999, **38**:7142-7150.
44. Castro MJ, Anderson S: **Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin.** *Biochemistry* 1996, **35**:11435-11446.