Higher-Order Packing Interactions in Triple and Quadruple Mutants of Staphylococcal Nuclease†

Junmei Chen‡ and Wesley E. Stites*

Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701-1201

Received June 19, 2001; Revised Manuscript Received September 11, 2001

ABSTRACT: Sixty-four triple and 32 quadruple mutants were constructed in the core of staphylococcal nuclease. This is the first time that a large number of multiple mutants with all possible variations and all possible lower-order mutants has been systematically constructed in any protein core. Stabilities were determined by solvent denaturation. The energetic effects of these multiple mutants have been analyzed in combination with the stability data from the component single and double mutants. It was found that most of the stability changes in triple and quadruple mutants cannot be correctly predicted from stability effects of component single mutants. However, if the interaction energy between pairs of side chains in the component double mutants is taken into account, correct stability prediction can be made for most triple and quadruple mutants. The data further show that while packing interactions unique to triple and quadruple mutations do occur, they are of much less energetic significance than interactions between pairs of residues. The results presented here show that the packing of a protein interior can be closely approximated in most cases as a series of short-range, nearest-neighbor interactions. This has profound implications for rational protein design and structure prediction.

Packing has an importance equal to that of hydrophobicity in determination of protein stability (*1*−*5*). Although stability effects caused by single and double packing mutants have been studied, the design of a protein with reasonable stability will require calculation of packing interactions between many residues. Therefore, it seems reasonable to examine the effects of simultaneous side chain substitutions at multiple sites in a protein.

Previously, multiple mutants (i.e., three or more simultaneous mutations) have been constructed and studied in the hydrophobic core of the following proteins: T4 lysozyme (*6*−*9*), the N-terminal domain of lambda repressor (*10*−*13*), C-terminal domains of lambda Cro repressor (*14*), Rop (*15*), ubiquitin (*16*), and chicken egg-white lysozyme (*17*).

Although several of the experiments have examined large numbers of multiple mutants in protein cores, these proteins were often examined by only a qualitative functional assay (*10*, *11*, *14*). Thus, only a few multiple mutations have been fully examined thermodynamically to determine the change in the free energy difference between the denatured and native states (*6*−*9*, *12*, *15*−*17*). Among those few multiple packing mutants which have been studied, nonadditive stability effects are common. In other words, the stability of the multiple mutants generally is not predicted successfully by summing the effects of the component single mutants (*6*, *7*, *9*).

This would seem to be an indication that predicting the stability of novel packing arrangements may be difficult. Several groups have attempted modeling the effects of multiple substitutions in the protein core with some success at predicting the conformation of side chains (*7*, *16*, *18*−*28*). However, much remains to be done. In particular, predicting precise energetic effects, a critical step for the design of new proteins, has not been terribly successful. So while it is clear that substantial changes can be made to the packing of a protein core, there are not yet good methods for predicting the effects of such packing changes. One of the most significant problems preventing improvements in models is the lack of large numbers of thermodynamically well-characterized multiple mutants. A large set of multiple mutant test cases with complementary data about the effects of all component single, double, and multiple mutants would likely speed progress on this problem.

For this reason, a large number of triple and quadruple mutants have been systematically constructed in the hydrophobic core of staphylococcal nuclease in this work. Only three possible side chain substitutions (isoleucine, leucine, or valine) were used to generate the triple and quadruple mutants at the six selected core positions (23, 25, 66, 72, 92, and 99). All possible triple and quadruple mutants were constructed at two quartets of sites: positions 23, 25, 66, and 72 as well as positions 66, 72, 92, and 99. Since there is a relatively large distance between the pairs of positions 23 and 25 and 92 and 99 and to limit the numbers of mutants to manageable numbers, no systematic effort to construct triple and quadruple mutants at all six sites was made. Nevertheless, there is still a large number of possible mutants, 64 and 32 possible triple and quadruple mutants, respectively, and all of them have been made. To our knowledge, this is

† Funding for this work was provided by NIH Grants GM52714 and 1P20RR15569-1.

* To whom correspondence should be addressed. Phone: (501) 575-7478. Fax: (501) 575-4049. E-mail: wstites@uark.edu.

‡ Current address: N1302 Alkek Research Building, One Baylor Plaza, Baylor College of Medicine, Houston, TX 77030.

10.1021/bi011269d CCC: $20.00
Published on Web 10/26/2001

the largest set of higher-order packing mutants ever made and the only set for which all component lower-order mutations have had stabilities determined as well. This allows us to analyze the interactions between these side chains in triple and quadruple mutant cycles. To the best of our knowledge, only one triple mutant cycle has ever been done (*29*, *30*) and no quadruple mutant cycles have ever been performed.

EXPERIMENTAL PROCEDURES

Mutagenesis, Protein Expression, and Protein Purification. All triple and quadruple mutants were prepared by multiple cycles of Kunkel mutagenesis (*31*, *32*) by using the proper Kunkel templates containing single or double mutations as described previously (*33*, *34*). Protein expression and purification followed the procedures previously described (*33*, *35*).

Protein Stability Determination. The stabilities of each of the mutants in this study were characterized by guanidine hydrochloride denaturation at 20 °C using fluorescence as a probe of structure as previously described (*33*, *36*, *37*).

Deviation between the Observed and Predicted Stabilities: Calculation of Interaction Energies. Most readers are likely familiar with the idea of double mutant cycles, in which the interaction of two side chains is examined by comparing the effects of the component single mutants to that of the double mutant (*39*). In other words, the deviation between predicted and observed stability changes is taken as the value of the interaction energy. A general theory for examining the energetics of side chain interactions between multiple residues has been laid out in detail by Horovitz and Fersht (*29*). We briefly review it here in slightly different form. The deviations between the observed and predicted stabilities for triple and quadruple mutants can be calculated in several ways. In the simplest calculation, shown here for triple mutants, the predicted stability changes are calculated only on the basis of the stability effects caused by the component single mutants.

$$\Delta^3G_{\mathrm{int}-1} = \Delta\Delta G_{\mathrm{triple}} - \sum\Delta\Delta G_{\mathrm{single}} = \Delta\Delta G_{\mathrm{triple}} - (\Delta\Delta G_{\mathrm{single1}} + \Delta\Delta G_{\mathrm{single2}} + \Delta\Delta G_{\mathrm{single3}})$$

where $\Delta\Delta G_{\mathrm{triple}}$ is the stability change of a triple mutant relative to that of the wild type, $\sum\Delta\Delta G_{\mathrm{single}}$ is the sum of the stability changes caused by the constituent single mutants, and $\Delta\Delta G = \Delta G_{\mathrm{H_2O,mutant}} - \Delta G_{\mathrm{H_2O,WT}}$. The term $\Delta^nG_{\mathrm{int}-x}$ is derived from the nomenclature of Horovitz and Fersht (*29*), where the superscript *n* indicates the total number of side chains being mutated. The subscript *x*, 1 this case, indicates that only the effects of the component single mutants are taken into account.

Thus, calculation of $\Delta^3G_{\mathrm{int}-1}$ removes most of the simple effects due the single mutations, such as changes in hydrophobicity. The remaining energetic effect is due to the packing interaction of the mutated residues and their environment. While this assesses the overall interactions between the residues of a multiple mutant, it fails to use all the information at our disposal. For example, it is clear from the double mutant cycles in the preceding paper that there can be significant interaction between pairs of side chains, while other pairs of side chains might have little interaction. We have measured these interactions already. It remains possible that three or four side chains might also have higher-order interactions which are not captured by examination of pairwise interactions.

The stability changes in a multiple mutant that cannot be explained by the effects of the component single mutants or by the interactions in the constituent lower-order mutants (i.e., three single and three double mutants in the case of a triple mutant) define an interaction energy, Δ^nG_{int}, where all interactions are examined. For a triple mutant, this interaction energy is equal to

$$\Delta^3G_{\mathrm{int}} = \Delta^3G_{\mathrm{int}-1} - \sum\Delta^2G_{\mathrm{int}}$$

The lack of subscript number in Δ^3G_{int} indicates that all possible lower-order interactions are taken into account. As shown below, $\sum\Delta^2G_{\mathrm{int}}$ is the sum of interaction energies from all the component double mutants of that triple mutant. The values of $\sum\Delta^2G_{\mathrm{int}}$ are calculated for triple mutants as follows:

$$\sum\Delta^2G_{\mathrm{int}} = \Delta^2G_{\mathrm{intA}} + \Delta^2G_{\mathrm{intB}} + \Delta^2G_{\mathrm{intC}}$$

The subscript A, B, or C in Δ^2G_{int} indicates the three different component double mutants of the triple mutant. The values of Δ^2G_{int} for all component double mutants in this study are published in the preceding paper (*34*). Since $\Delta^3G_{\mathrm{int}-1} = \Delta\Delta G_{\mathrm{triple}} - \sum\Delta\Delta G_{\mathrm{single}}$ and since the general expression for Δ^2G_{int} is equivalent to $\Delta\Delta G_{\mathrm{double}} - \Delta\Delta G_{\mathrm{single1}} - \Delta\Delta G_{\mathrm{single2}}$, the above equation for Δ^3G_{int} can be rearranged as

$$\Delta^3G_{\mathrm{int}} = \Delta\Delta G_{\mathrm{triple}} - \sum\Delta\Delta G_{\mathrm{double}} + \sum\Delta\Delta G_{\mathrm{single}}$$

where $\sum\Delta\Delta G_{\mathrm{double}}$ and $\sum\Delta\Delta G_{\mathrm{single}}$ are the sums of the stability changes caused by the constituent double and single mutants, respectively.

For quadruple mutants, the calculation of interaction energy is similar to that in triple mutants. If the predicted stability of a quadruple mutant is calculated just on the basis of the stability effects caused by the component single mutants or single and double mutants, the interaction energies are calculated with these first and second equations, respectively,

$$\Delta^4G_{\mathrm{int}-1} = \Delta\Delta G_{\mathrm{quadruple}} - \sum\Delta\Delta G_{\mathrm{single}}$$

$$\Delta^4G_{\mathrm{int}-2} = \Delta\Delta G_{\mathrm{quadruple}} - \sum\Delta^2G_{\mathrm{int}} - \sum\Delta\Delta G_{\mathrm{single}} = \Delta\Delta G_{\mathrm{quadruple}} - \sum\Delta\Delta G_{\mathrm{double}} + 2\sum\Delta\Delta G_{\mathrm{single}}$$

As above, we use the term $\Delta^4G_{\mathrm{int}-1}$ where the subscript 1 indicates that only the stability effects of the four component single mutants are considered. In the term $\Delta^4G_{\mathrm{int}-2}$, the subscript 2 indicates that, in addition to the effects of the single mutants, the pairwise interactions of six component double mutants are taken into account. However, in a quadruple mutant, besides the four constituent single and six double mutants, there are four component triple mutants. The interactions within the component triple mutants are not necessarily predicted by the stability effects of the single and double mutants. Indeed, this deviation between the observed stability and predicted stability was just calculated above for triple mutants ($\Delta^3G_{\mathrm{int}} = \Delta\Delta G_{\mathrm{triple}} - \sum\Delta\Delta G_{\mathrm{double}}$

$+ \sum\Delta\Delta G_{single}$). If these triplet interactions are taken into account, then the residual interaction energy between the observed and predicted stability for a quadruple mutant is given by

$$\begin{aligned}\Delta^4 G_{int} &= \Delta\Delta G_{quadruple} - (\sum\Delta\Delta G_{triple} - 2\sum\Delta\Delta G_{double} + 3\sum\Delta\Delta G_{single}) - \sum\Delta\Delta G_{double} + 2\sum\Delta\Delta G_{single} \\ &= \Delta\Delta G_{quadruple} - \sum\Delta\Delta G_{triple} + \sum\Delta\Delta G_{double} - \sum\Delta\Delta G_{single}\end{aligned}$$

where $\Delta\Delta G_{quadruple}$ is the stability change of a quadruple mutant relative to the wild type and $\sum\Delta\Delta G_{triple}$, $\sum\Delta\Delta G_{double}$, and $\sum\Delta\Delta G_{single}$ are the sums of the stability changes found for all the component triple, double, and single mutants, respectively.

Criteria for Additivity and Error Estimation. Error in the various energetic terms can be estimated by standard methods for propagation of error. A few examples are given in detail. Recall that $\Delta^3 G_{int-1}$ is calculated as

$$\Delta^3 G_{int-1} = \Delta\Delta G_{triple} - (\Delta\Delta G_{single1} + \Delta\Delta G_{single2} + \Delta\Delta G_{single3})$$

Since $\Delta\Delta G = \Delta G_{H_2O,mutant} - \Delta G_{H_2O,WT}$, each value of $\Delta\Delta G_{triple}$, $\Delta\Delta G_{single1}$, $\Delta\Delta G_{single2}$, and $\Delta\Delta G_{single3}$ contains a common term, $\Delta G_{H_2O,WT}$. Expansion of the above equation gives

$$\Delta^3 G_{int-1} = \Delta G_{H_2O,triple} - \Delta G_{H_2O,WT} - (\Delta G_{H_2O,single1} - \Delta G_{H_2O,WT} + \Delta G_{H_2O,single2} - \Delta G_{H_2O,WT} + \Delta G_{H_2O,single3} - \Delta G_{H_2O,WT})$$

After two $\Delta G_{H_2O,WT}$ terms have been canceled and rearrangement

$$\Delta^3 G_{int-1} = \Delta G_{H_2O,triple} - \Delta G_{H_2O,single1} - \Delta G_{H_2O,single2} - \Delta G_{H_2O,single3} + 2\Delta G_{H_2O,WT}$$

The experimental error for each determination of ΔG_{H_2O} is random and independent from each other except the error of the two $\Delta G_{H_2O,WT}$ terms is the same, while still random and independent from other errors. Therefore, the square of the sum of $\Delta G_{H_2O,WT}$ errors is used in the error estimation of $\Delta^3 G_{int-1}$:

$$\sigma_{\Delta^3 G_{int-1}} = \sqrt{(\sigma_{\Delta G_{H2O,triple}})^2 + (\sigma_{\Delta G_{H2O,single1}})^2 + (\sigma_{\Delta G_{H2O,single2}})^2 + (\sigma_{\Delta G_{H2O,single3}})^2 + (2\sigma_{\Delta G_{H2O,WT}})^2}$$

The magnitude of experimental error for each determination of ΔG_{H_2O} is estimated to be ± 0.1 kcal/mol on the basis of the extensive experience of experimental reproducibility. Therefore

$$\sigma_{\Delta^3 G_{int-1}} = \sqrt{4 \times (0.1)^2 + (2 \times 0.1)^2} = \pm 0.28 \text{ kcal/mol}$$

which we round to ± 0.3 kcal/mol. The above equation can be generalized as

$$\sigma_{\Delta^n G_{int-1}} = \sqrt{(n+1) \times (0.1)^2 + [(n-1)0.1]^2}$$

where n is the number of mutations in a protein. Therefore, the estimated error for quadruple mutants is equal to

$$\sigma_{\Delta^4 G_{int-1}} = \sqrt{(4+1) \times (0.1)^2 + [(4-1)0.1]^2} = \pm 0.37 \text{ kcal/mol}$$

which we round to ± 0.4 kcal/mol. While the fact that the estimated error is higher for the quadruple mutants is kept in mind, for the sake of simplicity and consistency, if the value of $\Delta^3 G_{int-1}$ or $\Delta^4 G_{int-1}$ is 0 ± 0.3 kcal/mol, the multiple mutant is considered to have additive stability effects. The same criterion was applied to $\Delta^2 G_{int}$ in previous studies (*34*, *38*).

A similar expansion, followed by cancelation of wild-type terms, can be done to show that the error is highest for $\Delta^4 G_{int-2}$, where many terms are repeated, at ± 0.57 kcal/mol, rounded to ± 0.6 kcal/mol. In the same manner, it can be shown that the error for $\Delta^4 G_{int}$ is ± 0.40 kcal/mol and for $\Delta^3 G_{int}$ ± 0.28 kcal/mol, rounded to ± 0.3 kcal/mol.

RESULTS

Guanidine Hydrochloride Denaturation. The stability data of 64 triple mutants from guanidine hydrochloride denaturation are summarized in Table 1. Only one triple mutant has stability similar to that of the wild type, that being 66L/92V/99L ($\Delta G_{H_2O} = 5.5$ kcal/mol). The remaining triple mutants are less stable than the wild type, with stabilities ranging from 2.2 to 5.0 kcal/mol. The midpoint concentrations (C_m) ranged from 0.32 to 0.90 M. The slope values (m_{GuHCl}) of the triple mutants expressed relative to that of the wild type varied from 0.79 to 1.15.

The stabilities of 32 quadruple mutants have also been determined by guanidine hydrochloride denaturation and are listed in Table 2. The stabilities (ΔG_{H_2O}) and midpoint concentrations (C_m) of all quadruple mutants are lower than that of the wild type. The values ranged from 2.0 to 4.7 kcal/mol and 0.28 to 0.71 M, respectively. The slope values varied from 0.80 to 1.08 relative to the value of the wild type.

The average stability loss for the quadruple mutants was 2.2 kcal/mol, for triple mutants 1.7 kcal/mol, and, in comparison, for the double mutants in the preceding article (*34*) 1.3 kcal/mol.

Additive and Nonadditive Stability Effects. Stability changes in multiple mutants ($\Delta\Delta G_{multiple}$) can be compared to the sum of the stability changes of the corresponding single mutants ($\sum\Delta\Delta G_{single}$). If the value of $\Delta\Delta G_{multiple}$ is close (i.e., within the estimated error) to $\sum\Delta\Delta G_{single}$, the multiple mutant has additive stability effects; otherwise, its stability effects are nonadditive.

The stability changes of triple ($\Delta\Delta G_{triple}$) and quadruple mutants ($\Delta\Delta G_{quadruple}$) relative to the wild type are plotted against the sum of stability changes of the component single mutants ($\sum\Delta\Delta G_{single}$) in panels A and B of Figure 1, respectively. The dotted line is the theoretical line with a slope of unity that mutants of perfect additivity would fall upon, and the solid line is the actual regression line through

Table 1: Solvent Denaturation Data for Triple Packing Mutants

	Δ-G_{H_2O}[a]	C_m[b]	m_{GuHCl}[c]	ΔΔ-G_{triple}[d]	∑ΔΔ-G_{single}[e]	Δ^3-G_{int-1}[f]	Δ-CH_2[g]
23/25/66							
23I/25I/66I	2.5	0.43	0.87	−2.9	−2.9	0.0	2
23I/25I/66L	2.9	0.56	0.79	−2.5	−2.2	−0.3	2
23I/25V/66I	2.7	0.46	0.89	−2.7	−3.0	0.3	1
23I/25V/66L	3.3	0.60	0.84	−2.1	−2.3	0.2	1
23L/25V/66L	3.4	0.56	0.94	−2.0	−2.2	0.2	1
23L/25V/66I	2.6	0.40	0.99	−2.8	−2.9	0.1	1
23L/25I/66I	2.3	0.40	0.90	−3.1	−2.8	−0.3	2
23L/25I/66L	3.0	0.51	0.89	−2.4	−2.1	−0.3	2
23/25/72							
23I/25I/72V	2.6	0.38	1.05	−2.8	−3.1	0.3	0
23I/25I/72L	3.7	0.60	0.95	−1.7	−2.1	0.4	1
23I/25V/72V	3.1	0.44	1.08	−2.3	−3.2	0.9	−1
23I/25V/72L	4.3	0.67	1.00	−1.1	−2.2	1.1	0
23L/25V/72V	2.6	0.35	1.15	−2.8	−3.1	0.3	−1
23L/25V/72L	3.9	0.57	1.05	−1.5	−2.1	0.6	0
23L/25I/72V	2.9	0.41	1.08	−2.5	−3.0	0.5	0
23L/25I/72L	4.0	0.60	1.04	−1.4	−2.0	0.6	1
66/92/99							
66I/92V/99I	4.9	0.75	1.00	−0.5	−1.6	1.1	1
66L/92V/99I	5.0	0.85	0.91	−0.4	−0.9	0.5	1
66I/92V/99L	4.5	0.65	1.05	−0.9	−1.7	0.8	1
66L/92V/99L	5.5	0.90	0.94	0.1	−1.0	1.1	1
66L/92L/99L	4.7	0.79	0.90	−0.7	−1.2	0.5	2
66I/92L/99L	3.3	0.53	0.95	−2.1	−1.9	−0.2	2
66L/92L/99I	4.0	0.71	0.85	−1.4	−1.1	−0.3	2
66I/92L/99I	3.3	0.57	0.88	−2.1	−1.8	−0.3	2
72/92/99							
72V/92V/99I	4.0	0.56	1.10	−1.4	−1.8	0.4	−1
72L/92V/99I	4.7	0.68	1.05	−0.7	−0.8	0.1	0
72V/92V/99L	3.4	0.47	1.11	−2.0	−1.9	−0.1	−1
72L/92V/99L	4.3	0.63	1.05	−1.1	−0.9	−0.2	0
72L/92L/99L	4.0	0.60	1.04	−1.4	−1.1	−0.3	1
72V/92L/99L	3.0	0.41	1.13	−2.4	−2.1	−0.3	0
72L/92L/99I	4.0	0.65	0.95	−1.4	−1.0	−0.4	1
72V/92L/99I	3.3	0.47	1.08	−2.1	−2.0	−0.1	0
23/66/72							
23I/66I/72L	3.4	0.62	0.83	−2.0	−1.4	−0.6	2
23I/66I/72V	3.8	0.63	0.93	−1.6	−2.4	0.8	1
23I/66L/72L	4.0	0.69	0.88	−1.4	−0.7	−0.7	2
23I/66L/72V	4.5	0.76	0.90	−0.9	−1.7	0.8	1
23L/66I/72L	3.4	0.58	0.91	−2.0	−1.3	−0.7	2
23L/66I/72V	3.8	0.57	1.00	−1.6	−2.3	0.7	1
23L/66L/72L	4.1	0.65	0.96	−1.3	−0.6	−0.7	2
23L/66L/72V	4.1	0.64	0.99	−1.3	−1.6	0.3	1
25/66/72							
25I/66I/72L	2.4	0.38	0.97	−3.0	−2.9	−0.1	1
25I/66I/72V	2.2	0.32	1.07	−3.2	−3.9	0.7	0
25I/66L/72L	3.1	0.49	0.95	−2.3	−2.2	−0.1	1
25I/66L/72V	3.2	0.51	0.98	−2.2	−3.2	1.0	0
25V/66I/72L	2.4	0.40	0.93	−3.0	−3.0	0.0	0
25V/66I/72V	2.3	0.32	1.09	−3.1	−4.0	0.9	−1
25V/66L/72L	3.2	0.51	0.97	−2.2	−2.3	0.1	0
25V/66L/72V	3.4	0.51	1.02	−2.0	−3.3	1.3	−1
66/72/92							
66I/72L/92L	3.1	0.58	0.83	−2.3	−1.8	−0.5	1
66I/72V/92L	3.4	0.50	1.04	−2.0	−2.8	0.8	0
66L/72L/92L	3.9	0.66	0.92	−1.5	−1.1	−0.4	1
66L/72V/92L	4.1	0.65	0.96	−1.3	−2.1	0.8	0
66I/72L/92V	4.0	0.61	1.00	−1.4	−1.6	0.2	0
66I/72V/92V	4.0	0.59	1.04	−1.4	−2.6	1.2	−1
66L/72L/92V	4.6	0.69	1.01	−0.8	−0.9	0.1	0
66L/72V/92V	4.9	0.75	1.01	−0.5	−1.9	1.4	−1
66/72/99							
66I/72L/99L	3.5	0.58	0.92	−1.9	−1.5	−0.4	2
66I/72V/99L	3.8	0.60	0.98	−1.6	−2.5	0.9	1
66L/72L/99L	4.6	0.78	0.90	−0.8	−0.8	0.0	2
66L/72V/99L	4.9	0.79	0.94	−0.5	−1.8	1.3	1
66I/72L/99I	3.5	0.63	0.85	−1.9	−1.4	−0.5	2
66I/72V/99I	4.6	0.73	0.98	−0.8	−2.4	1.6	1
66L/72L/99I	4.4	0.75	0.90	−1.0	−0.7	−0.3	2
66L/72V/99I	4.6	0.75	0.95	−0.8	−1.7	0.9	1
wild type	5.4	0.82	1.00	–	–	–	–

[a] Free energy difference between the native and denatured states in the absence of denaturant in units of kilocalories per mole. The error is estimated to be ±0.1 kcal/mol. [b] Midpoint concentration (concentration of guanidine hydrochloride at which half of the protein is denatured) in units of molar. The error is estimated to be ±0.01 M. [c] Slope value (change in free energy with respect to the change in guanidine hydrochloride concentration) expressed relative to the wild-type value of 6.53 kcal mol^{-1} M^{-1}. The error is estimated to be ±0.02. [d] Difference in free energy between the free energy of the protein with three substitutions and the free energy of the wild-type protein. $\Delta\Delta G_{triple} = \Delta G_{H_2O}$ (triple mutant) − 5.4 (WT). The error is estimated to be ±0.2 kcal/mol. [e] The sum of the $\Delta\Delta G_{single}$ values of corresponding single substitutions. [f] $\Delta^3 G_{int-1} = \Delta\Delta G_{triple} - \sum\Delta\Delta G_{single}$. The error is estimated to be ±0.3 kcal/mol. [g] Difference in the number of methylene groups between the mutant and the wild type.

Table 2: Solvent Denaturation Data for Quadruple Packing Mutants

	Δ-G_{H_2O}[a]	C_m[b]	m_{GuHCl}[c]	ΔΔ-$G_{quadruple}$[d]	∑ΔΔ-G_{single}[e]	Δ^4-G_{int-1}[f]	Δ-CH_2[g]
23/25/66/72							
23I/25I/66I/72V	2.2	0.35	0.94	−3.2	−4.1	0.9	1
23I/25I/66I/72L	2.2	0.40	0.85	−3.2	−3.1	−0.1	2
23I/25I/66L/72L	2.7	0.47	0.88	−2.7	−2.4	−0.3	2
23I/25I/66L/72V	2.8	0.48	0.90	−2.6	−3.4	0.8	1
23I/25V/66I/72V	2.4	0.37	0.98	−3.0	−4.2	1.2	0
23I/25V/66I/72L	2.6	0.45	0.89	−2.8	−3.2	0.4	1
23I/25V/66L/72L	2.9	0.49	0.91	−2.5	−2.5	0.0	1
23I/25V/66L/72V	3.3	0.52	0.97	−2.1	−3.5	1.4	0
23L/25I/66I/72V	2.0	0.32	0.98	−3.4	−4.0	0.6	1
23L/25I/66I/72L	2.2	0.37	0.90	−3.2	−3.0	−0.2	2
23L/25I/66L/72L	3.0	0.48	0.96	−2.4	−2.3	−0.1	2
23L/25I/66L/72V	2.8	0.41	1.03	−2.6	−3.3	0.7	1
23L/25V/66I/72V	2.0	0.28	1.08	−3.4	−4.1	0.7	0
23L/25V/66I/72L	2.6	0.41	0.95	−2.8	−3.1	0.3	1
23L/25V/66L/72L	3.4	0.52	1.00	−2.0	−2.4	0.4	1
23L/25V/66L/72V	2.9	0.42	1.04	−2.5	−3.4	0.9	0
66/72/92/99							
66I/72V/92V/99I	4.4	0.67	1.01	−1.0	−2.8	1.8	0
66I/72V/92V/99L	3.7	0.54	1.06	−1.7	−2.9	1.2	0
66I/72V/92L/99I	3.4	0.51	1.02	−2.0	−3.0	1.0	1
66I/72V/92L/99L	3.0	0.42	1.08	−2.4	−3.1	0.7	1
66I/72L/92V/99I	3.8	0.59	1.00	−1.6	−1.8	0.2	1
66I/72L/92V/99L	3.6	0.52	1.05	−1.8	−1.9	0.1	1
66I/72L/92L/99I	2.1	0.41	0.80	−3.3	−2.0	−1.3	2
66I/72L/92L/99L	2.4	0.42	0.88	−3.0	−2.1	−0.9	2
66L/72V/92V/99I	4.6	0.69	1.02	−0.8	−2.1	1.3	0
66L/72V/92V/99L	4.7	0.71	1.02	−0.7	−2.2	1.5	0
66L/72V/92L/99I	3.6	0.57	0.97	−1.8	−2.3	0.5	1
66L/72V/92L/99L	4.0	0.62	1.00	−1.4	−2.4	1.0	1
66L/72L/92V/99I	4.4	0.69	0.97	−1.0	−1.1	0.1	1
66L/72L/92V/99L	4.7	0.71	1.01	−0.7	−1.2	0.5	1
66L/72L/92L/99I	3.5	0.58	0.92	−1.9	−1.3	−0.6	2
66L/72L/92L/99L	4.1	0.65	0.98	−1.3	−1.4	0.1	2
wild type	5.4	0.82	1.00	–	–	–	–

[a] Free energy difference between the native and denatured states in the absence of denaturant in units of kilocalories per mole. The error is estimated to be ±0.1 kcal/mol. [b] Midpoint concentration (concentration of guanidine hydrochloride at which half of the protein is denatured) in units of molar. The error is estimated to be ±0.01 M. [c] Slope value (change in free energy with respect to the change in guanidine hydrochloride concentration) expressed relative to the wild-type value of 6.53 kcal mol^{-1} M^{-1}. The error is estimated to be ±0.02. [d] Difference in free energy between the free energy of the protein with four substitutions and the free energy of the wild-type protein. $\Delta\Delta G_{quadruple} = \Delta G_{H_2O}$ (quadruple mutant) − 5.4 (WT). The error is estimated to be ±0.2 kcal/mol. [e] The sum of the $\Delta\Delta G_{single}$ values of corresponding single substitutions. [f] $\Delta^4 G_{int-1} = \Delta\Delta G_{quadruple} - \sum\Delta\Delta G_{single}$. The error is estimated to be ±0.4 kcal/mol. [g] Difference in the number of methylene groups between the mutant and the wild type.

the data points. The correlation coefficients (r) of triple and quadruple mutants are 0.7233 and 0.6581, respectively. Compared to the double mutants in previous study (*34*), with a correlation coefficient (r) of 0.8572 for a similar plot, nonadditive stability effects are obviously more dominant among the triple and quadruple mutants.

DISCUSSION

A nonadditive stability effect is defined as an experimentally significant deviation between the stability change of a multiple mutant relative to the wild type and the sum of the stability changes caused by the corresponding single mutants. Nonadditive stability effects have been found in multiple packing mutants in previous studies (*7*), but without a systematic effort to examine all possible or at least a large number of multiple mutants, it was not clear how common such effects are (*6*, *9*). In our study, 36 of 64 triple mutants (56%) and 22 of 32 quadruple mutants (69%) have nonad-

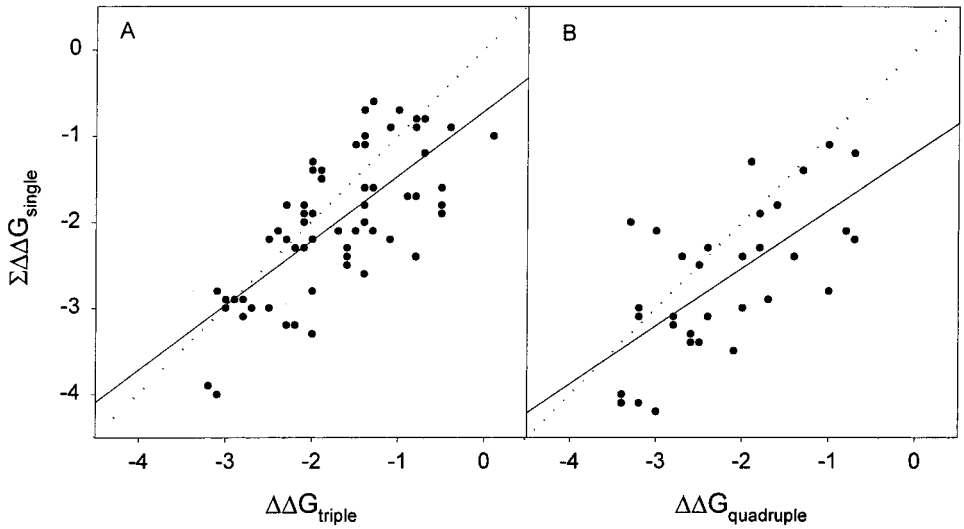

FIGURE 1: (A) Plot of $\sum\Delta\Delta G_{single}$ (in kilocalories per mole) as calculated from single mutants vs $\Delta\Delta G_{triple}$ of triple mutants (in kilocalories per mole). (B) Plot of $\sum\Delta\Delta G_{single}$ (in kilocalories per mole) as calculated from single mutants vs $\Delta\Delta G_{quadruple}$ of quadruple mutants (in kilocalories per mole). The dotted line is the theoretical line with a slope of unity. Points below this line are mutants that are more stable than predicted from the single mutants. Those above are mutants that are less stable than predicted. The solid line is the regression line through the data points. The correlation coefficient (r) for triple mutants is 0.7233 and for quadruple mutants is 0.6581.

ditive stability effects (Figure 1). Therefore, nonadditive stability effects are very common among multiple mutants, and their prevalence increases as the number of mutated side chains in a protein core increases.

Correlation of Nonadditive Stability Effects with ΔCH_2. Evidence that these nonadditive effects are due to packing interactions is found by examining the effect of changing the number of methylene units in the core of the protein. The difference between the stability change of a multiple mutant relative to the wild type ($\Delta\Delta G_{multiple}$) and the sum of the stability effects of the component single mutants ($\sum\Delta\Delta G_{single}$) is Δ^3G_{int-1} or Δ^4G_{int-1} for the triple or quadruple mutant, respectively. If Δ^3G_{int-1} or Δ^4G_{int-1} is positive, the multiple mutant in question is more stable than expected from the sum of the changes from single mutants. This is presumably due to favorable interactions among the mutated side chains. If Δ^3G_{int-1} or Δ^4G_{int-1} is negative, the strains among the mutated side chains caused the multiple mutant to be less stable than the sum of the single mutant effects. The values of Δ^3G_{int-1} and Δ^4G_{int-1} are plotted in Figures 2A and 3A, respectively.

The triple and quadruple mutants have been divided into different groups on the basis of the change in the number of methylene units relative to the wild type (ΔCH_2). Figure 2A shows that when $\Delta CH_2 = -1$ and 0, most values of Δ^3G_{int-1} are positive (i.e., favorable). When the number of methylene groups in the core is increased by one, although favorable interactions are still dominant, unfavorable interactions appear in some mutants. Furthermore, when two methylene groups are added, unfavorable interactions are dominant. Similarly for the quadruple mutant shown in Figure 3A, all values of Δ^4G_{int-1} are positive when $\Delta CH_2 = 0$ and 1. Again, the values of Δ^4G_{int-1} change to negative when one more methylene group is added. The correlation of Δ^3G_{int-1} and Δ^4G_{int-1} and ΔCH_2 among the triple and quadruple mutants implies that more strains are introduced among the mutant side chains as the overall side chain volume increases. This is precisely what one would predict from a packing effect. While individual mutants can vary, the overall trend is absolutely clear. There can be no doubt that the nonadditive effects are due to packing.

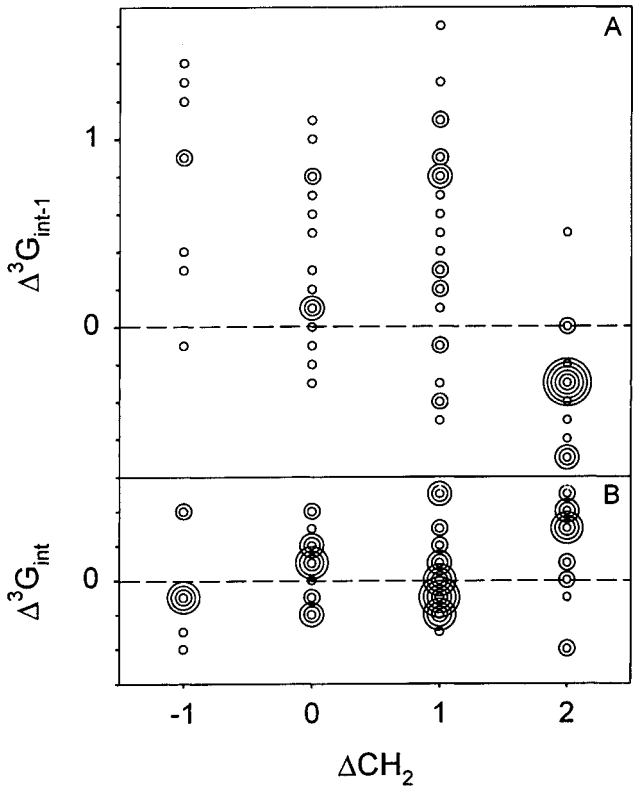

FIGURE 2: Bullseye-bubble plot of interaction energies (in kilocalories per mole) in triple mutants organized by the value of ΔCH_2 (the change in the number of methylene groups upon mutation). In the case of coincident data points, the size of the bubble and the number of rings indicate the number of mutants with that particular interaction energy. (A) Only the energetic effects caused by constituent single mutants are considered in calculations of the predicted stability ($\Delta^3G_{int-1} = \Delta\Delta G_{triple} - \sum\Delta\Delta G_{single}$). The estimated error for each point is ±0.3 kcal/mol. (B) The energetic effects caused by both component single and double mutants are considered in calculations of the predicted stability ($\Delta^3G_{int} = \Delta\Delta G_{triple} - \sum\Delta\Delta G_{double} + \sum\Delta\Delta G_{single}$). The estimated error for each point is ±0.3 kcal/mol.

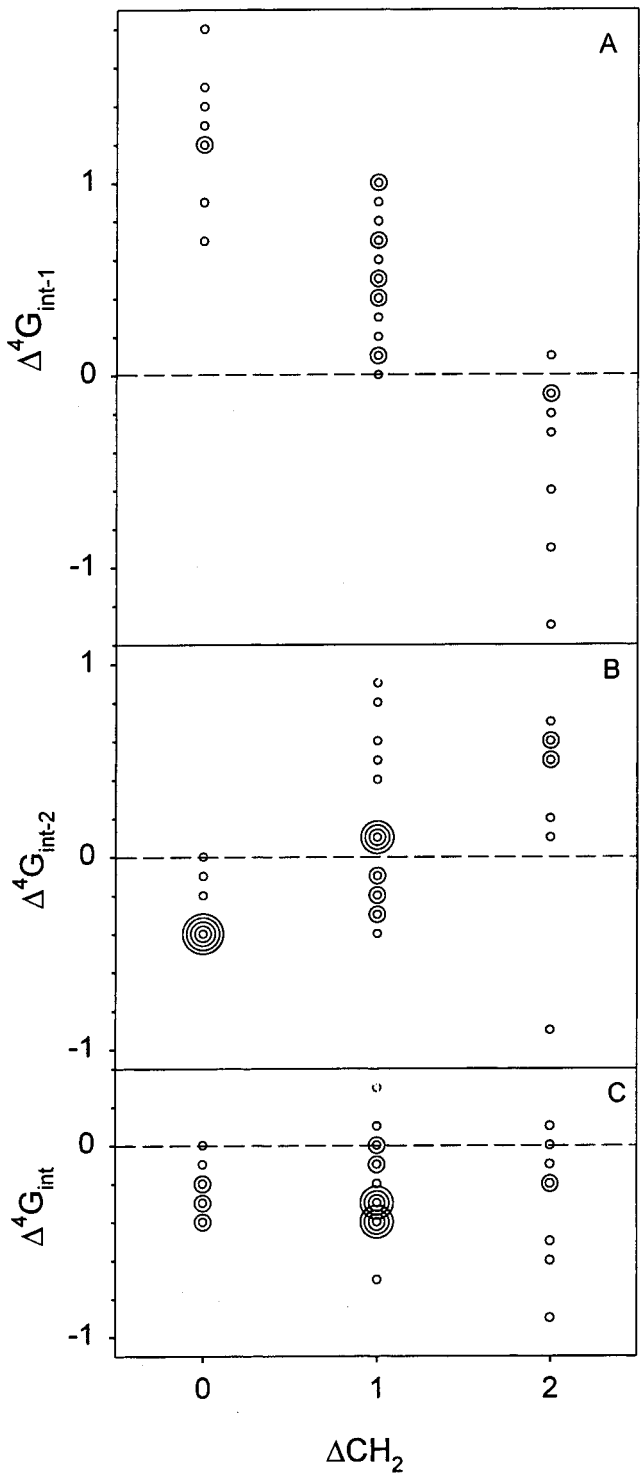

FIGURE 3: Bullseye-bubble plot of interaction energies (in kilocalories per mole) in quadruple mutants organized by the value of ΔCH_2 (the change in the number of methylene groups upon mutation). In the case of coincident data points, the size of the bubble and the number of rings indicate the number of mutants with that particular interaction energy. (A) Only the energetic effects caused by constituent single mutants are considered in calculations of the predicted stability ($\Delta^4G_{int-1} = \Delta\Delta G_{quadruple} - \sum\Delta\Delta G_{single}$). The estimated error for each point is ±0.4 kcal/mol. (B) The energetic effects caused by both component single and double mutants are considered in calculations of the predicted stability ($\Delta^4G_{int-2} = \Delta\Delta G_{quadruple} - \sum\Delta\Delta G_{double} + 2\sum\Delta\Delta G_{single}$). The estimated error for each point is ±0.6 kcal/mol. (C) The energetic effects caused by component single, double, and triple mutants are considered in calculations of the predicted stability ($\Delta^4G_{int} = \Delta\Delta G_{quadruple} - \sum\Delta\Delta G_{triple} + \sum\Delta\Delta G_{double} - \sum\Delta\Delta G_{single}$). The estimated error for each point is ±0.4 kcal/mol.

Hierarchy of Nonadditive Stability Effects. It is easy to see how changes in packing can lead to nonadditive stability effects. Side chains will adopt different packing arrangements depending on what the neighboring side chains are. Change two or more neighboring side chains simultaneously, and the multiple mutant may well find a packing arrangement for each side chain that is different from what each component side chain substitution would adopt in the absence of a nearby mutation.

What is less obvious is the level at which these interactions between side chains must be considered. Clearly, the entire protein acts as an integrated whole in optimizing packing to minimize the overall free energy of the protein. This was recently demonstrated by our discovery that experimentally significant structural changes in the packing of the core of staphylococcal nuclease occurred when mutations were made in surface residues far removed from the core (*40*). However, it is not clear how energetically significant these large-scale, longer-range interactions are.

If two residues in the proximity of one another are mutated, they will reach a mutually satisfactory packing arrangement that may be affected little by other mutational changes elsewhere in the protein, especially if those changes are relatively further away. Can interactions in a large group of residues be well-approximated by using just the interactions between smaller subsets of residues? This is a critical question for protein design efforts.

The answer to this question is demonstrated in Figures 2 (triple mutants) and 3 (quadruple mutants), where we show the interaction energies due to various levels of interaction between side chains. The details of the calculation of the interaction energies (Δ^nG_{int-x}), slightly extended beyond the proposal of Horovitz and Fersht (*29*, *30*), are given in Experimental Procedures. In essence, the various interaction energies allow us to quantify the degree to which the stability change of a multiple mutant is caused by the component single side chain substitutions or by interactions between those side chains. Further, we can examine the degree to which interactions must be considered between pairs, triplets, or even quadruplets of residues.

First, we consider the triple mutants. In Figure 2A, only the energetic effects caused by single mutants are considered in calculations of the predicted stability. Clearly, there is a great deal of energetically significant interaction between the three residues of these mutants which is not simply due to the effects of the single mutants. In Figure 2B, the stability effects caused by both single mutants and the stability effects of interactions found in the double mutants are considered ($\Delta^3G_{int} = \Delta\Delta G_{triple} - \sum\Delta\Delta G_{double} + \sum\Delta\Delta G_{single}$). There is an obvious improvement. In fact, relatively few of the values of Δ^3G_{int} are larger than the likely errors in measurement. Another important point to note is that the correlation of decreasing stability with increasing numbers of methylene units, so prominent in panel A, is now gone. The interaction energies measured in the double mutants account quite nicely for most packing effects.

Similarly in Figure 3, the degree to which stability effects in quadruple mutants are accounted for by the effects of the single mutants is shown in panel A ($\Delta^4G_{int-1} = \Delta\Delta G_{quadruple} - \sum\Delta\Delta G_{single}$). If the effects of the single mutants and the interactions between pairs of residues measured in double mutants are taken into account, the remaining interaction

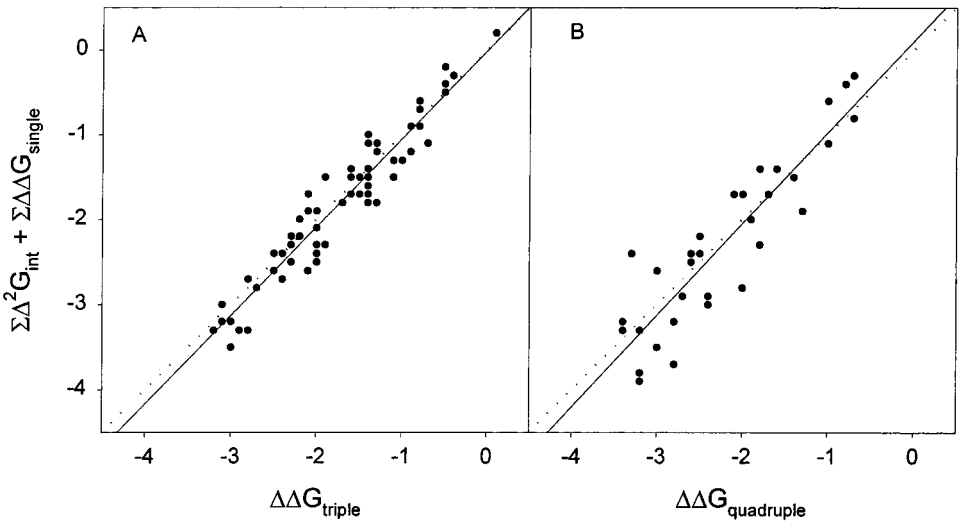

FIGURE 4: Plot of the pairwise interaction energies calculated from component double mutants plus the stability effects of the component single mutants ($\Sigma\Delta^2G_{int} + \Sigma\Delta\Delta G_{single}$ in kilocalories per mole) vs the observed stability effects of triple mutants [$\Delta\Delta G_{triple}$ (A)] and of quadruple mutants [$\Delta\Delta G_{quadruple}$ (B)]. The dotted line is the theoretical line with a slope of unity on which all points would fall if Δ^3G_{int} and Δ^4G_{int-2} equaled zero. The solid line is the regression line through the data points. The correlation coefficient (r) is 0.9566 for the triple mutants and 0.9012 for the quadruple mutants.

energy is shown in panel B ($\Delta^4G_{int-2} = \Delta\Delta G_{quadruple} - \Sigma\Delta\Delta G_{double} + 2\Sigma\Delta\Delta G_{single}$). The correlation with the change in the number of methylene units is removed, or perhaps even slightly reversed, although the combination of relatively high scatter in the values and error in Δ^4G_{int-2} makes an actual reverse correlation doubtful. Finally, if the effects of the single mutants, the interactions between pairs of residues, and the interactions unique to triplets of residues measured in triple mutants are all accounted for, the remaining interaction energy unique to the interaction of all four residues in the quadruple mutant is shown in panel C ($\Delta^4G_{int} = \Delta\Delta G_{quadruple} - \Sigma\Delta\Delta G_{triple} + \Sigma\Delta\Delta G_{double} - \Sigma\Delta\Delta G_{single}$).

The trend is clear. Even after the stability effects of the single mutants are accounted for, the remaining energy of all the packing interactions between the residues of a multiple mutant is significant. If the packing interaction energy measured in the component double mutants is subtracted, the residual interaction energy is decreased dramatically and the correlation of energy with the number of methylene units added is lost. If the energies of the interactions unique to the three side chains of the corresponding triple mutants are also considered, the energy that remains is due to interactions found only in the complete quadruple mutant. That residual interaction energy, Δ^4G_{int}, shown in panel C, is rather small and is obviously uncorrelated to the change the number of methylene units.

The major conclusion we draw from these two figures is that higher-order interactions, i.e., interactions that occur only in complete multiple mutant, are much less important than the lower-order interaction between pairs of side chains. Many, if not most, of the packing interactions in quadruple mutants are reflected in the corresponding double mutants. Pairwise interactions dominate in the hierarchy of protein packing.

This idea that double mutants accurately reflect the interactions in higher-order mutants can be shown in a different way. The observed stability effects of triple and quadruple mutants ($\Delta\Delta G_{quadruple}$ and $\Delta\Delta G_{triple}$) are plotted in panels A and B of Figure 4, respectively, against the sum of interaction energies from all component double mutants and the effects of the component single mutations ($\Sigma\Delta^2G_{int} + \Sigma\Delta\Delta G_{single}$). Expecting a good correlation between these two quantities is equivalent to expecting Δ^nG_{int-2} to equal zero since, for example, $\Delta^4G_{int-2} = \Delta\Delta G_{quadruple} - \Sigma\Delta\Delta G_{double} + 2\Sigma\Delta\Delta G_{single}$. Significant correlations are found in both cases, much improved over the case where merely the effects of the single mutants are taken into account (Figure 1). The correlation coefficient (r) is 0.9566 for triple mutants and 0.9012 for quadruple mutants. This demonstrates that the complicated interactions among multiple core side chains can be broken down and closely approximated, to ~90%, by considering the effects of the single mutants (hydrophobic transfer free energy changes, entropic factors, and the like) and only packing interactions between pairs of side chains. This is in agreement with our earlier conclusion (*34*) that nonadditive effects are largely limited to side chains in actual van der Waals contact and do not propagate beyond these nearest neighbors.

We hasten to add that while this is the general trend not every mutant obeys it perfectly. In Figure 4, some data points deviate from both the regression line and the theoretical line, which must be due to changes in the interactions found in double mutants when a third or fourth mutation is introduced. However, the magnitudes of these changes are relatively small compared to the magnitudes of the packing effects captured by consideration of pairwise interactions. A further caution is that these conclusions are drawn from a series of aliphatic substitutions and might not hold true for other types of substitutions.

Regardless, two lines of evidence, the additivity of double mutants and the low residual interaction energies once double mutant interactions are taken into account in higher-order mutations, both indicate the same thing. The energetically significant packing interactions in the core of staphylococcal nuclease are largely short-range, pairwise interactions.

Implications for Rational Protein Design. Protein engineering would be simplified if the energetic effects caused by single mutants are simply additive in multiple mutants.

However, this is not the case as the results from our studies show that the stabilities of many proteins with two or more mutations deviate from the sum of stability changes of the component single mutants. The more substitutions in the protein core, the more likely the stability effects will be nonadditive with 32% of double mutants (*34*), 56% of triple mutants, and 69% of quadruple mutants having nonadditive stability effects. This clearly highlights the importance of packing to protein stability, but is bad, although not unexpected, news from the standpoint of protein design. Also bad news is the fact that very few of the triple or quadruple mutants have energies similar to that of the wild type and the more mutations, the lower the average stability of the protein. One of the two most destabilizing mutations here, 23L/25V/66I/72V, does not add or subtract a single atom relative to the wild type, but destabilizes the protein by 3.4 kcal/mol. In other words, there are many more ways to get packing wrong, than right.

The good, and somewhat surprising, news is that stabilities in triple and quadruple mutants are closely approximated by considering the effects of single mutants and the packing interactions between pairs of residues. Further, the effects of the single mutants are largely additive in those pairs unless the two residues are in direct van der Waals contact. Longer-range effects involving larger numbers of residues can be seen in our data, but they are of much less energetic significance. This allows an important simplification of the packing interactions which must be considered when attempting to computationally predict the stability of a protein.

ACKNOWLEDGMENT

We thank the referees for useful comments.

REFERENCES

1. Richards, F. M., and Lim, W. A. (1993) *Q. Rev. Biophys. 26*, 423−498.
2. Hurley, J. H. (1994) in *The Protein Folding Problem and Tertiary Structure Prediction*, pp 549−578, Birkhäuser, Boston.
3. Hubbard, S. J., and Argos, P. (1995) *Curr. Opin. Biotechnol. 6*, 375−381.
4. Lee, B., and Vasmatzis, G. (1997) *Curr. Opin. Biotechnol. 8*, 423−428.
5. Lazar, G. A., and Handel, T. M. (1998) *Curr. Opin. Chem. Biol. 2*, 675−679.
6. Daopin, S., Alber, T., Baase, W. A., Wozniak, J. A., and Matthews, B. W. (1991) *J. Mol. Biol. 221*, 647−667.
7. Hurley, J. H., Baase, W. A., and Matthews, B. W. (1992) *J. Mol. Biol. 224*, 1143−1159.
8. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1993) *Science 262*, 1715−1718.
9. Gassner, N. C., Baase, W. A., and Matthews, B. W. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12155−12158.
10. Lim, W. A., and Sauer, R. T. (1989) *Nature 339*, 31−36.
11. Lim, W. A., and Sauer, R. T. (1991) *J. Mol. Biol. 219*, 359−376.
12. Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry 31*, 4324−4333.
13. Lim, W. A., Hodel, A., Sauer, R. T., and Richards, F. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 423−427.
14. Mollah, A. K., Aleman, M. A., Albright, R. A., and Mossing, M. C. (1996) *Biochemistry 35*, 743−748.
15. Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., and Regan, L. (1996) *Protein Sci. 5*, 1584−1593.
16. Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1997) *Protein Sci. 6*, 1167−1178.
17. Shih, P., Holland, D. R., and Kirsch, J. F. (1995) *Protein Sci. 4*, 2050−2062.
18. Lee, C., and Levitt, M. (1991) *Nature 352*, 448−451.
19. Lee, C., and Subbiah, S. (1991) *J. Mol. Biol. 217*, 373−388.
20. Ponder, J. W., and Richards, F. M. (1987) *J. Mol. Biol. 193*, 775−791.
21. Dahiyat, B. I., and Mayo, S. L. (1996) *Protein Sci. 5*, 895−903.
22. Kono, H., Nishiyama, M., Tanokura, M., and Doi, J. (1997) *Pac. Symp. Biocomput. '97*, 210−221.
23. Desjarlais, J. R., and Handel, T. M. (1995) *Protein Sci. 4*, 2006−2018.
24. Bryson, J. W., Desjarlais, J. R., Handel, T. M., and DeGrado, W. F. (1998) *Protein Sci. 7*, 1404−1414.
25. Kono, H., Nishiyama, M., Tanokura, M., and Doi, J. (1998) *Protein Eng. 11*, 47−52.
26. Kono, H., and Doi, J. (1994) *Proteins 19*, 244−255.
27. Wilson, C., Gregoret, L. M., and Agard, D. A. (1993) *J. Mol. Biol. 229*, 996−1006.
28. Tuffery, P., Etchebest, C., Hazout, S., and Lavery, R. (1991) *J. Biomol. Struct. Dyn. 8*, 1267−1289.
29. Horovitz, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 733−740.
30. Horovitz, A., and Fersht, A. R. (1990) *J. Mol. Biol. 214*, 613−617.
31. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.
32. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
33. Byrne, M. P., Manuel, R. L., Lowe, L. G., and Stites, W. E. (1995) *Biochemistry 34*, 13949−13960.
34. Chen, J., and Stites, W. E. (2001) *Biochemistry 40*, 14004−14011.
35. Shortle, D., Stites, W. E., and Meeker, A. K. (1990) *Biochemistry 29*, 8033−8041.
36. Stites, W. E., Byrne, M. P., Aviv, J., Kaplan, M., and Curtis, P. M. (1995) *Anal. Biochem. 227*, 112−122.
37. Schwehm, J. M., and Stites, W. E. (1998) *Methods Enzymol. 295*, 150−170.
38. Green, S. M., and Shortle, D. (1993) *Biochemistry 32*, 10131−10139.
39. Carter, P. J., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1984) *Cell 38*, 835−840.
40. Chen, J., Lu, Z., Sakon, J., and Stites, W. E. (2000) *J. Mol. Biol. 303*, 125−130.

BI011269D