Packing Is a Key Selection Factor in the Evolution of Protein Hydrophobic Cores[†]

Junmei Chen[‡] and Wesley E. Stites*

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

Received September 10, 2001; Revised Manuscript Received October 9, 2001

ABSTRACT: The energy derived from optimized van der Waals interactions in closely packed, folded proteins has been proposed to be of similar energetic magnitude to hydrophobicity in stabilizing the native state. If packing is this energetically important, it should influence the evolution of protein core sequences. To test this hypothesis, the occurrence of various amino acid side chains in the major hydrophobic core of staphylococcal nuclease and 42 homologous proteins was determined. Most such positions in this protein family are usually isoleucine, leucine, or valine. Previously we have constructed and measured the stabilities of 12 single, 44 double, 64 triple, and 32 quadruple mutants, representing all possible permutations of these three side chains at two overlapping sets of four positions in the core of staphylococcal nuclease. The stabilities and interaction energies of those mutants with various combinations of the most common, or consensus, sequence were compared to the stabilities of all other mutants. Mutants which had the consensus side-chain combinations were not necessarily the most stable, but usually were found to have the best interaction energies. In other words, these proteins were far more stable than would be predicted from simply summing the observed energetic effects of the component single mutations, apparently reflecting particularly favorable packing interactions that are possible for the most common side chains. An additional 12 mutants which tested possible alternate explanations of the results were constructed. The stabilities and interaction energies of these mutants also support the conclusion that packing is a crucial determinant guiding the sequence evolution of protein cores.

Hydrophobicity has long been thought to be the major driving force for protein folding (*1*). Recently, close packing of the hydrophobic core to optimize van der Waals contacts and minimize cavities has been proposed to be of roughly equal importance energetically (*2−11*). During molecular evolution, important residues in protein sequences are highly conserved for functional and structural reasons, often reasons of structural stability. Indeed, Maxwell and Davidson (*12*) recently showed that the rate of occurrence of different amino acids at two interacting positions in over 350 SH3 domain homologues was very highly related to the stability of the protein with that side chain substituted. In keeping with their importance to protein stability, buried hydrophobic residues are generally well conserved during evolution (*13*, *14*).

Since stability has been demonstrated as a selection pressure during protein evolution and if packing of cores is actually as energetically important as hydrophobicity, then sequences which pack the hydrophobic core well should also be conserved. Therefore, it follows that the combinations of protein core residues which are more commonly found in a group of homologous proteins should have some stability advantage over other possible permutations of hydrophobic residues because of the particular favorable packing arrangements available to these conserved sets of residues.

To test this hypothesis, the consensus side chains have been identified at residues 23, 25, 66, 72, 92, and 99 in the cores of staphylococcal nuclease and 42 homologous proteins found in Genbank as of August 1, 2001. The stabilities of double, triple, and quadruple mutants at these sites with elements of the consensus core sequences are compared with those of random hydrophobic core sequences. In addition, staphylococcal nuclease mutants with all six consensus core residues were generated, their stabilities were determined by guanidine hydrochloride denaturation, and their stabilities are compared to random hydrophobic core sequences.

MATERIALS AND METHODS

Sequence Alignment. PSI-BLAST 2.2 (*15*) (http://www.ncbi.nlm.nih.gov/BLAST/) was used to identify and initially align sequences in our study. First, the sequence of staphylococcal nuclease or a homologue was used to search the Genbank database for all homologous proteins. Once they were identified, related sequences were aligned manually, utilizing our knowledge of important structural features in addition to simple sequence homology.

Mutagenesis, Protein Expression, and Purification. All mutants were prepared by multiple cycles Kunkel mutagenesis by using the special Kunkel templates with desired mutations as previously described (*10*). Proteins were expressed and purified as previously described (*16*, *17*).

Guanidine Hydrochloride Denaturation. The stabilities of the multiple mutants were determined by guanidine hydrochloride denaturation in an AVIV model ATF-101 fluorometer as previously described (*9*, *17−19*). Data analysis was carried out assuming a two-state model for the reversible protein unfolding reaction.

RESULTS

Occurrence of Various Amino Acid Types in the Core of Staphylococcal Nuclease Homologues. The sequence of

Table 1: Amino Acid Types Found in Each of the Major Hydrophobic Cores of Staphylococcal Nuclease and Its Homologs

protein[b]	23	25	66	72	92	99	14	17	34	36	39	62	65	74	103	protein[b]	23	25	66	72	92	99	14	17	34	36	39	62	65	74	103
1	V	L	V	I	I	V	L	A	F	L	V	T	M	V	L	23	I	V	V	L	V	V	L	A	V	Y	I	N	L	L	L
2	I	I	L	V	V	F	V	V	V	L	V	T	H	L	L	24	F	L	I	V	L	V	C	C	G	L	V	L	V	I	M
3	I	V	L	V	V	F	V	V	L	M	V	T	H	L	L	25	I	L	L	I	V	V	V	V	I	F	V	L	V	V	L
4	I	V	V	V	V	V	V	I	V	L	I	L	M	V	M	26	I	V	L	V	V	I	V	I	V	L	L	L	Q	L	M
5	I	V	T	V	V	I	V	I	V	L	I	L	L	V	M	27	I	C	L	V	V	A	S	V	V	I	I	A	R	I	L
6	I	V	V	V	V	V	V	V	I	L	I	M	L	V	M	28	L	V	I	V	V	F	V	W	I	I	I	A	L	I	M
7	V	V	L	I	I	F	V	V	I	L	V	A	E	V	L	29	P	I	I	V	L	F	I	W	L	I	I	A	L	V	M
8	I	V	C	V	V	L	C	V	I	L	V	V	M	L	L	30	F	V	I	V	V	F	M	Y	I	I	I	V	L	V	M
9	F	C	I	V	V	L	V	V	V	L	I	S	L	L	L	31	F	S	I	V	I	L	I	V	Y	V	I	V	K	V	L
10	L	V	T	V	M	I	V	V	V	L	V	L	I	L	L	32	L	V	A	V	V	L	V	V	I	M	I	L	A	V	Q
11	V	V	T	V	L	V	V	V	V	L	I	L	Y	L	L	33	I	V	V	L	L	V	P	V	K	N	I	L	I	V	M
12	I	A	I	I	I	I	V	I	I	M	I	L	A	V	M	34	V	V	V	L	I	V	A	V	I	L	I	L	V	V	M
13	V	V	L	V	L	Y	V	V	V	I	I	A	A	L	A	35	V	V	V	L	I	I	V	V	I	L	I	L	L	V	M
14	V	L	A	L	L	L	V	V	V	L	I	A	S	L	L	36	F	F	L	V	V	I	V	V	V	L	I	L	K	I	M
15	F	A	L	V	V	Y	V	V	V	L	V	S	T	L	L	37	F	F	I	V	V	L	V	V	I	L	I	L	R	V	M
16	I	F	L	F	V	L	V	V	I	I	I	L	L	L	c	38	I	I	S	T	C	V	Q	V	V	L	V	D	L	C	L
17	I	V	L	V	V	I	V	I	I	F	I	L	I	I	L	39	I	I	S	W	C	V	A	I	V	L	I	L	H	C	M
18	I	V	V	I	V	V	V	V	I	L	I	M	L	V	M	40	I	M	A	V	C	V	V	V	V	F	I	D	L	C	M
19	L	C	V	V	V	I	V	I	V	L	I	L	L	L	M	41	I	L	I	V	C	L	A	I	I	M	I	L	L	C	Q
20	I	I	V	V	V	A	V	V	V	L	I	L	L	V	M	42	I	I	L	I	C	I	I	V	I	L	I	L	F	C	M
21	I	V	V	L	V	V	L	A	V	Y	V	N	L	L	L	43	W	Y	I	Q	V	S	T	F	H	L	I	L	L	Y	R
22	L	C	V	V	V	V	V	I	V	L	I	L	L	L	M																

[a] Numbers refer to the sequence numbering of wild-type staphylococcal nuclease. [b] The Genbank accession number for each protein follows. Many sequences are available in Genbank under multiple accession numbers, in which case only one arbitrarily chosen accession number is given here. 1 (staphylococcal nuclease), gi224650; 2, gi1171860; 3, gi1171859; 4, gi152543; 5, gi266681; 6, gi3337063; 7, gi2501230; 8, gi2621565; 9, gi2983292; 10, gi3257629; 11, gi5458481; 12, gi10957342; 13, gi10581248; 14, gi11351545; 15, gi14247095; 16, gi8572690; 17, gi6968414; 18, gi14422500; 19, gi2501229; 20, gi9971685; 21, gi1750126; 22, gi12721193; 23, gi3025484; 24, gi13242638; 25, gi1724012; 26, gi4894356; 27, gi10956900; 28, gi1351504; 29, gi2496325; 30, gi14089438; 31, gi14089552; 32, gi7380351; 33, gi4586057; 34, gi14030876; 35, gi2501231; 36, gi3150137; 37, gi1723879; 38, gi7387647; 39, gi12620760; 40, gi13470703; 41, gi13470768; 42, gi14027483; 43, gi14028038. [c] This putative protein is truncated relative to other nuclease homologs, ending before residue 103.

Table 2: Number of Times Each Side Chain Is Found at a Given Major Hydrophobic Core Position in Staphylococcal Nuclease and Homologs

amino acid	occurrence of each amino acid at position number														
	23	25	66	72	92	99	14	17	34	36	39	62	65	74	103
I	22	6	10	6	6	9	3	9	16	6	31	0	3	5	0
L	5	5	12	6	6	8	3	0	2	26	1	23	18	15	17
V	7	20	12	27	25	15	27	26	20	1	11	3	3	17	0
F	7	3	0	1	0	6	0	1	1	3	0	0	1	0	0
Y	0	1	0	0	0	2	0	1	1	2	0	0	1	1	0
M	0	1	0	0	1	0	1	0	0	4	0	2	3	0	21
C	0	4	1	0	5	0	2	1	0	0	0	0	0	5	0
W	1	0	0	1	0	0	0	2	0	0	0	0	0	0	0
A	0	2	3	0	0	2	3	3	0	0	0	6	3	0	1
D	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
G	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	1	0	0	0	3	0	0
K	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0
N	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0
P	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Q	0	0	0	1	0	0	1	0	0	0	0	0	1	0	2
R	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1
S	0	1	2	0	0	1	1	0	0	0	0	2	1	0	0
T	0	0	3	0	1	0	1	0	0	0	0	3	1	0	0
consensus	I	V	L/V/I	V	V	V	V	V	V/I	L	I	L	L	V/L	M/L
staph. nuclease	V	L	V	I	I	V	L	A	F	L	V	T	M	V	L

staphylococcal nuclease was aligned with 42 homologous proteins found in Genbank as of August 1, 2001. These proteins are identified by their Genbank accession numbers in Table 1. The complete sequence alignment is available electronically as Supporting Information on the World Wide Web edition. Several sequences which differ by only a few amino acids from another sequence were not included. A few very distantly related homologues are difficult to align with certainty and so were not included. Most such excluded homologues belong to a eukaryotic transcription factor family, each having five nuclease domains (20, 21).

The residues found in the major hydrophobic cores of staphylococcal nuclease and its homologues are summarized in Table 2, which include the residues equivalent to the wild-type staphylococcal nuclease positions 14, 17, 23, 25, 34, 36, 39, 62, 65, 66, 72, 74, 92, 99, and 103. These 15 residues represent a broad definition of the major hydrophobic core in nuclease. Most of these residues are fully buried or at least largely so. A few are somewhat more solvent exposed but are included because they are consistently hydrophobic and pack up against other members of the major hydrophobic core. (Another smaller cluster of buried hydrophobic residues,

Table 3: Solvent Denaturation Data of the Consensus Core and Related Mutants

mutants	$\Delta G_{H_2O}{}^a$	$C_m{}^b$	$m_{GuHCl}{}^c$	$\Delta\Delta G^d$	$\Sigma\Delta\Delta G_{singles}{}^e$	$\Delta^n G_{int-1}{}^f$	$\Delta CH_2{}^g$
full consensus core mutants with no deletion							
23I/25V/72V/92V	3.0	0.41	1.10	−2.4	−3.6	1.2	−2
23I/25V/66L/72V/92V	3.5	0.53	1.00	−1.9	−3.9	2.0	−1
full consensus core mutants with deletion of residue 70							
23I/25V/72V/92V/69Δ71	1.8	0.25	1.13	−3.6	−5.0	1.4	−2
23I/25V/66L/72V/92V/69Δ71	2.3	0.33	1.06	−3.1	−5.3	2.2	−1
other related mutants							
69Δ71	4.0	0.60	1.01	−1.4			0
I72V/69Δ71	2.8	0.39	1.10	−2.6	−2.6	0.0	−1
I72V/I92V/69Δ71	2.4	0.33	1.14	−3.0	−3.0	0.0	−2
23I/66L/72V/92V	4.8	0.74	1.00	−0.6	−2.1	1.5	0
23I/25V/72V/92V/99L	2.4	0.33	1.11	−3.0	−3.9	0.9	−1
23I/25V/72V/92V/99I	2.7	0.39	1.08	−2.7	−3.8	1.1	−1
23I/25I/66L/72V/92V	3.1	0.47	1.00	−2.3	−3.8	1.5	0
23L/25V/66L/72V/92V	3.2	0.45	1.09	−2.2	−3.8	1.6	−1
Wild-type	5.4	0.82	1.00				

^{a} The free-energy difference between native and denatured states in the absence of denaturant in units of kcal/mol. 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 M. Error is estimated to be ±0.01 M. ^{c} Slope value (change in free energy with respect to change in guanidine hydrochloride concentration) expressed relative to wild-type value of 6.53 kcal/(mol M). Error is estimated to be ±0.02. ^{d} Difference in free energy between the free energy of the mutant protein and the free energy of the wild-type protein. $\Delta\Delta G = \Delta G_{H_2O}$ (mutant) − 5.4 (WT). Error is estimated to be ±0.2 kcal/mol. ^{e} The sum of the $\Delta\Delta G$ values of corresponding single substitution mutants. ^{f} $\Delta^n G_{int-1} = \Delta\Delta G - \Sigma\Delta\Delta G_{singles}$. Error is estimated to be ±0.2 kcal/mol for double mutants, ±0.3 kcal/mol for triple mutants, ±0.4 kcal/mol for quadruple mutants, ±0.5 kcal/mol for quintuple mutants, and ±0.6 kcal/mol for sextuple mutants. See reference (*11*) for details of error calculation. ^{g} Difference in the number of methylene groups between the mutant and wild-type.

the minor hydrophobic core, exists elsewhere in the protein but is not considered further here.) The stabilities of 152 single, double, triple, and quadruple mutants permutating isoleucine, leucine, and valine at positions 23, 25, 66, 72, 92, and 99 have been previously published (*9−11*). These positions are the focus of this study as well and are, therefore, listed first in Table 2.

The number of times a given residue type occurs at each position in the cores of staphylococcal nuclease and its homologues is summarized in Table 2. Isoleucine, leucine, and valine dominate the composition of the hydrophobic core, except for position 103. At position 103, in addition to leucine, another hydrophobic amino acid, methionine, occurs with high frequency.

Many positions, including 23, 25, 72, and 92, strongly favor a particular amino acid, although no core position shows the nearly absolute conservation of, for example, some active site residues or some other structurally important sites elsewhere in the protein. Indeed, some positions, including position 66, appear to have no marked preference between two or three of the aliphatic hydrophobic amino acids. Other positions, such as 99, appear to have only a slight preference for a particular amino acid. However, all six positions strongly favor the three aliphatic hydrophobic side chains, isoleucine, leucine, and valine, over the other side chains. The most frequently occurring residue(s) for each position are listed at the bottom of Table 2 as the consensus sequence along with the sequence of wild-type staphylococcal nuclease. The same results were also found when we first examined in 1998 the 15 staphylococcal nuclease homologues then in Genbank. Even though the number of homologues has nearly tripled from 15 to 42, the consensus sequence at these six positions has changed only in that a slight preference for valine or leucine over isoleucine at position 66 has grown less pronounced. This perhaps indicates that the consensus core sequence in the nuclease family is not likely to change dramatically as more homologues are found.

Stability of Mutant Proteins. Twelve mutant proteins which test the importance of conserving various aspects of the core packing in staphylococcal nuclease were constructed, and protein stabilities were determined. These previously unpublished stabilities are given in Table 3. The mutants are all less stable than wild-type staphylococcal nuclease. Among the consensus core mutants with no deletion, 23I/25V/66L/72V/92V has the highest stability, 0.5 kcal/mol higher than 23I/25V/72V/92V. When isoleucine or leucine is substituted at position 99 in 23I/25V/72V/92V, the stabilities of 23I/25V/72V/92V/99I and 23I/25V/72V/92V/99L are decreased further by 0.3 and 0.6 kcal/mol, respectively. As discussed below, some homologues of staphylococcal nuclease have the equivalent of residue 70 deleted. All the mutants with residue 70 deleted have lower stability than those without deletion. There are three other related multiple mutants with the consensus core sequence at positions of 66, 72, 92, and 99, but with different sequence at positions 23 and 25. The stability (ΔG_{H_2O}) of 23I/25V/66L/72V/92V is higher than that of 23I/25I/66L/72V/92V and that of 23L/25V/66L/72V/92V.

The stability changes of these multiple mutants relative to wild-type ($\Delta\Delta G$) are compared to the sum of the stability effects from the corresponding single mutants ($\Sigma\Delta\Delta G_{singles}$). The interaction energy, $\Delta^n G_{int-1}$, introduced previously (*10, 11, 22*), is the difference between $\Delta\Delta G$ and $\Sigma\Delta\Delta G_{singles}$. The superscript *n* indicates the number of mutant side chains in the multiple mutant, and the subscript 1 indicates that the interaction energy takes into account only the effects of single mutants and not, for example, the interactions between component triple mutants in a quadruple mutant. In the case of double mutants, the subscript 1 is usually dropped for $\Delta^2 G_{int}$, since there are no lower order component multiple mutants. When $\Delta^n G_{int-1}$ is positive, there are favorable interactions among the substitutions, and the multiple mutant is more stable than would be predicted from the effects of the single mutants. Similarly, a negative $\Delta^n G_{int-1}$ is presumably due to strains introduced by the mutations and indicates

that the multiple mutant is less stable than expected from the effects of the component single mutations.

DISCUSSION

The six residues at positions 23, 25, 66, 72, 92, and 99 of staphylococcal nuclease are close to each other in the tertiary structure and form an extensive hydrophobic contact network among themselves and with other neighboring residues. Hence, these six residues can be viewed as the center of the major hydrophobic core of nuclease. These six residues are fairly well conserved in related protein sequences but have enough variability to allow a practical test of the influence of packing energetics upon evolution of protein cores. Generally, these positions are restricted to the aliphatic hydrophobic amino acids: isoleucine, leucine, or valine.

Previously we have described the construction and stability determination of 152 single, double, triple, and quadruple packing mutants systematically covering all possible combinations of these three amino acid types at two overlapping sets of four positions drawn from these six positions (*9−11*). Therefore, in this present study, we focus on combinations of the amino acids which are commonly found in the nuclease family of proteins at positions 23, 25, 66, 72, 92, and 99 and examine the correlation between residue conservation and the contribution of packing to protein stability.

It is unmistakable from Table 2 that different positions show different preferences for the various amino acid types. In all likelihood, a particular amino acid is found more commonly at a certain position because it confers a selective advantage to organisms which have it there. Ultimately, this evolutionary advantage must relate to the function of the protein, and since most proteins do not function if unfolded, protein stability should be a selection factor.

We start with the assumption that the advantage at these six positions in the core of nuclease homologues relates principally to the stability of the protein and not directly to function. This seems reasonable since these six core sites are not directly in the active site. Position 23 is closest in both primary and tertiary structure to the active site, being near residues D19, G20, D21, and T22. (The two aspartates help bind the calcium required for activity.) It is likely that mutations at 23 perturb the local conformation of these residues to some degree, but in contrast to the near absolute conservation of residues 19 (42/43 D), 20 (43/43 G), 21 (43/43 D), and 22 (39/43 T), the degree of variability at 23 is much higher and is comparable to other core residues well removed from the active site. Encouragingly, it has been previously demonstrated in another protein that even highly mutated cores often retain enzymatic activity (*23*). So while we are aware that alterations of the position 23 side chain in particular or the packing of the core as a whole might subtly alter active site geometries and hence evolutionary fitness, the hypothesis we test here is that selection at these sites is driven primarily by stability and, specifically, by the stability effects of hydrophobic core packing.

The most frequently occurring residues define the consensus sequence, shown at the bottom of Table 2. Isoleucine is found most frequently at position 23, while valine is the consensus residue at positions 25, 72, 92, and 99, although the preference at 99 is not especially strong. At position 66, leucine and valine are equally common and only slightly more common than isoleucine. As also shown at the bottom of Table 2, at positions 99 and 66, where consensus is weakest, staphylococcal nuclease does have one of the residues commonly found there. However, nuclease does not have the consensus residues at positions 23, 25, 72, or 92.

In this, nuclease is not an exception. Although they are the most frequently observed side chains at a particular site, as a combination occurring simultaneously, the complete consensus core is more rare. If we allow any of the three aliphatic hydrophobes at position 66 and require the most common amino acid at the other five positions, only two proteins, numbers 4 and 6, have what might be called a consensus core. If we further relax the criteria to allow any of the three aliphatic hydrophobes at position 99, we only add two more proteins, numbers 17 and 26, to the list of those with consensus cores. Therefore, the selection favoring the complete consensus core is not overwhelming. On the other hand, we point out that if isoleucine, leucine, and valine are allowed at four different sites, then the total number of possible permutations of these residues is 3^4 or 81. Four occurrences of a particular combination out of 43 examples are clearly much greater than chance would predict.

Nevertheless, the lack of agreement between the nuclease sequence and the consensus sequence makes an important point. More than one reasonably stable packing arrangement is possible. Taken as a whole, the wild-type nuclease core is an energetically stable arrangement. None of the 152 packing mutants we have made at these six positions significantly increases protein stability, and most decrease it appreciably. As Tables 1 and 2 also show, the more broadly defined hydrophobic core in nuclease also deviates from the consensus. As we shall demonstrate, there is an advantage to the consensus residues at the six positions relative to all other possibilities, but that advantage is probably only fully realized in a context of multiple changes to the entire protein core.

In this manner, the situation is remarkably similar to that which is likely to occur in the natural process of protein evolution. Suppose a protein with a stable packing arrangement is mutated. A single mutational event is unlikely to alter even two or three positions widely separated in primary structure, let alone the 15 residues of the entire hydrophobic core. As discussed further below, our experimental evidence shows that single mutations in the core that alter packing are likely to have unfavorable energetic effects even when they are conservative hydrophobic substitutions (*9*). Therefore, it seems likely that the evolutionary migration from one stable packing arrangement to another is through a series of mutants, most or all of which are less stable than the starting point, until at last another local free-energy minima in a new stable packing arrangement is found.

While disfavored over the long term, a short term loss of protein stability can be tolerated in the process of evolution. For example, wild-type staphylococcal nuclease has a stability of 5.4 kcal/mol, meaning a 10 000:1 ratio of folded to unfolded molecules. A reduction in stability of about 1.35 kcal/mol shifts that ratio to 1000:1. While this means a marginally less active enzyme more prone to proteolytic degradation and is probably less favorable over many generations of selection, it is also obviously not instantly lethal. The larger the loss of stability, the less likely it will be tolerated even briefly. As we will demonstrate below,

Table 4: Stabilities of Single Mutants at Six Selected Positions in the Major Hydrophobic Core of Staphylococcal Nuclease

mutant	$\Delta G_{H_2O}{}^a$	mutant	$\Delta G_{H_2O}{}^a$
V23I	5.2	I92L	4.8
V23L	5.3	**I92V**	5.0
L25I	3.7	V99I	5.2
L25V	3.6	V99L	5.1
V66I	4.4	Wild-type	5.4
V66L	5.1		
I72L	5.2		
I72V	4.2		

^{a} Free-energy difference between native and denatured states in the absence of denaturant in units of kcal/mol. Error is estimated to be ±0.1 kcal/mol. Data from reference (9).

various combinations of the consensus sequence residues are often the least disruptive mutations in the core of staphylococcal nuclease, and these mutants are among the least disruptive because they tend to be those with the most favorable packing interactions with their neighbors. However, as we will also show, stability is not the whole picture.

Before we examine the hypothesis that packing energetics drive the selection of residues, we first must examine an alternative hypothesis that factors intrinsic to each side chain rather than its interactions with other side chains drive selection. Even though the three aliphatic hydrophobes are similar, they differ in the precise amount of hydrophobic transfer free energy, secondary structure preference, and in the degrees of freedom lost when the side chains are frozen upon folding. At a fully buried position, these factors do not depend on the neighboring residues. Might such simple factors intrinsic to only the mutated side chain itself explain the rate of occurrence of each side chain at these sites?

The answer to this question is no, as an examination of our previously published data (9) on the stability of nuclease mutants at these six positions reveals. In Table 4, we show the stabilities of the two possible aliphatic mutants at each site. At the four positions with clear preferences, we highlight the consensus mutant in bold type. None of these single mutations is overwhelmingly destabilizing, but all are destabilizing. More importantly, note that the single mutants which are the consensus residues do not appear to be energetically more favorable than the mutants which are not. Indeed, if anything, consensus residues are less favored.

Correlation between the Consensus Core Sequence and the Stabilities of Packing Mutants. Is there a stability advantage to be found in the combinations of consensus residues that explains their evolutionary advantage? Our data on the stability of double, triple, and quadruple mutants (10, 11) show that there is.

In Table 5, we show the double, triple, and quadruple mutants grouped by position of mutation and then ranked in order of descending stability. The interaction energy, $\Delta^n G_{int-1}$, a quantity introduced previously (10, 11, 22), is also shown, where the superscript n indicates the number of mutant side chains, and the subscript 1 indicates that the interaction energy takes into account only the effects of the single mutants. The interaction energy has been discussed extensively elsewhere (10, 11, 22) and will be discussed further below, but in short it is the difference between the observed change in stability of a multiple mutant and the changes in

stabilities of the component single mutations. It is, as the name implies, a measure of how the side chains of a multiple mutant interact. A positive number indicates a favorable interaction; in other words, the multiple mutant is more stable than predicted on the basis of the effects of the component single mutations.

Another piece of information is provided in Table 5. Several mutants are highlighted in bold type. This indicates that these mutants have consensus sequence at the mutated positions. Recall that the consensus side chain at positions 23, 25, 72, and 92 is a mutation relative to wild-type nuclease. Valine is the amino acid found at positions 66 and 99 in wild-type nuclease. The occurrence of valine, isoleucine, and leucine in the entire nuclease family is essentially identical at 66, so any one of these could be fairly called a consensus choice. At position 99 valine is preferred, but not tremendously, and the occurrence of isoleucine and leucine is nearly identical. Consensus combinations which match the wild-type nuclease sequence are not considered since those are, not surprisingly, always the most stable and, more importantly, do not illuminate the process by which sequential mutation leads to a new stable packing arrangement. Therefore, isoleucine and leucine are both considered equally evolutionarily favored (or disfavored) at positions 66 and 99 in the mutant side-chain combination comparison, and either isoleucine or leucine is accepted as a consensus residue at positions 66 and 99.

Forty-four double, 64 triple, and 32 quadruple mutants are listed in Table 5 (10, 11). There are four double mutants at each of 11 pairs of positions, eight triple mutants at eight position triplets, and 16 quadruple mutants at two quadruplets of positions. This is every possible permutation of isoleucine, leucine, and valine at each particular set of positions. The mutants are listed in order of decreasing stability. Again, any mutants with the same amino acid composition as the consensus core sequence at the given positions are designated with bold type.

The consensus core side-chain combinations are often the mutants of the highest stability. This, at first, seems intuitive. More stable mutants should be selected for. However, closer examination shows that there is more going on than a simple relationship between mutant stability and evolutionary selection. First, in many cases, the consensus sequence mutant is far from the most stable. Additionally, in mutants that include 66 or 99, there are two possible consensus choices, four possible consensus choices if both are in a set, and these choices often have fairly different stabilities.

It is impossible not to notice that interaction energies of the consensus core mutants tend to be among the most favorable. Again, interaction energy is the energy difference between the observed stability change in a multiple mutant and the summed stability changes observed in that multiple's component single mutant (10, 11, 22), and a positive value indicates a favorable interaction. Admittedly, the correlation between the consensus side chains and interaction energy is not perfect, but it does seem far better than the correlation with protein stability. This is especially noticeable in the quadruple mutants.

Correlation of Consensus Sequence with Favorable Interaction Energies Shows the Importance of Packing as an Evolutionary Selection Factor. It seems obvious why mutants with the consensus side chains are often the most stable. As

Table 5: Stabilities of Double, Triple, and Quadruple Mutants at Six Selected Positions in the Major Hydrophobic Core of Staphylococcal Nuclease

Double mutants

double mutants	$\Delta G_{H_2O}{}^a$	$\Delta^2 G_{int}{}^b$
23/25		
V23I/L25V	4.0	0.6
V23L/L25I	3.9	0.3
V23I/L25I	3.7	0.2
V23L/L25V	3.6	0.1
66/72		
V66L/I72V	4.9	1.0
V66L/I72L	4.5	−0.4
V66I/I72V	4.2	1.0
V66I/I72L	3.7	−0.5
92/99		
I92V/V99I	4.7	−0.1
I92L/V99I	4.5	−0.1
I92V/V99L	4.5	−0.2
I92L/V99L	4.1	−0.4
23/66		
V23I/V66L	4.4	−0.5
V23L/V66L	4.3	−0.7
V23I/V66I	3.9	−0.3
V23L/V66I	3.9	−0.4
25/66		
L25V/V66L	3.6	0.3
L25I/V66L	3.5	0.1
L25V/V66I	2.5	−0.1
L25I/V66I	2.4	−0.3
23/72		
V23L/I72L	5.0	−0.1
V23I/I72L	4.9	−0.1
V23I/I72V	4.3	0.3
V23L/I72V	4.3	0.2
25/72		
L25V/I72L	3.8	0.4
L25I/I72L	3.7	0.2
L25V/I72V	2.5	0.1
L25I/I72V	2.4	−0.1
66/92		
V66L/I92V	5.5	0.8
V66I/I92V	4.7	0.7
V66L/I92L	4.4	−0.1
V66I/I92L	3.5	−0.3
66/99		
V66L/V99L	5.4	0.6
V66L/V99I	4.8	−0.1
V66I/V99I	4.7	0.5
V66I/V99L	4.1	0.0
72/92		
I72L/I92L	4.7	0.1
I72L/I92V	4.7	−0.1
I72V/I92V	3.7	−0.1
I72V/I92L	3.6	0.0
72/99		
I72L/V99I	4.9	−0.1
I72L/V99L	4.6	−0.3
I72V/V99I	4.2	0.2
I72V/V99L	3.7	−0.2

Triple mutants

triple mutants	$\Delta G_{H_2O}{}^a$	$\Delta^3 G_{int-1}{}^b$
23/25/66		
23L/25V/66L	3.4	0.2
23I/25V/66L	3.3	0.2
23L/25I/66L	3.0	−0.3
23I/25I/66L	2.9	−0.3
23I/25V/66I	2.7	0.3
23L/25V/66I	2.6	0.1
23I/25I/66I	2.5	0.0
23L/25I/66I	2.3	−0.3
23/25/72		
23I/25V/72L	4.3	1.1
23L/25I/72L	4.0	0.6
23L/25V/72L	3.9	0.6
23I/25I/72L	3.7	0.4
23I/25V/72V	3.1	0.9
23L/25I/72V	2.9	0.5
23I/25I/72V	2.6	0.3
23L/25V/72V	2.6	0.3
66/92/99		
66L/92V/99L	5.5	1.1
66L/92V/99I	5.0	0.5
66I/92V/99I	4.9	1.1
66L/92L/99L	4.7	0.5
66I/92V/99L	4.5	0.8
66L/92L/99I	4.0	−0.3
66I/92L/99L	3.3	−0.2
66I/92L/99I	3.3	−0.3
72/92/99		
72L/92V/99I	4.7	0.1
72L/92V/99L	4.3	−0.2
72V/92V/99I	4.0	0.4
72L/92L/99L	4.0	−0.3
72L/92L/99I	4.0	−0.4
72V/92V/99L	3.4	−0.1
72V/92L/99I	3.3	−0.1
72V/92L/99L	3.0	−0.3
23/66/72		
23I/66L/72V	4.5	0.8
23L/66L/72V	4.1	0.3
23L/66L/72L	4.1	−0.7
23I/66L/72L	4.0	−0.7
23I/66I/72V	3.8	0.8
23L/66I/72V	3.8	0.7
23I/66I/72L	3.4	−0.6
23L/66I/72L	3.4	−0.7
25/66/72		
25V/66L/72V	3.4	1.3
25I/66L/72V	3.2	1.0
25V/66L/72L	3.2	0.1
25I/66L/72L	3.1	−0.1
25V/66I/72L	2.4	0.0
25I/66I/72L	2.4	−0.1
25V/66I/72V	2.3	0.9
25I/66I/72V	2.2	0.7
66/72/92		
66L/72V/92V	4.9	1.4
66L/72L/92V	4.6	0.1
66L/72V/92L	4.1	0.8
66I/72V/92V	4.0	1.2
66I/72L/92V	4.0	0.2
66L/72L/92L	3.9	−0.4
66I/72V/92L	3.4	0.8
66I/72L/92L	3.1	−0.5
66/72/99		
66L/72V/99L	4.9	1.3
66I/72V/99I	4.6	1.6
66L/72V/99I	4.6	0.9
66L/72L/99L	4.6	0.0
66L/72L/99I	4.4	−0.3
66I/72V/99L	3.8	0.9
66I/72L/99L	3.5	−0.4
66I/72L/99I	3.5	−0.5

Quadruple mutants

quadruple mutants	$\Delta G_{H_2O}{}^a$	$\Delta^4 G_{int-1}{}^b$
23/25/66/72		
23L/25V/66L/72L	3.4	0.4
23I/25V/66L/72V	3.3	1.4
23L/25I/66L/72L	3.0	−0.1
23L/25V/66L/72V	2.9	0.9
23I/25V/66I/72L	2.9	0.0
23I/25I/66L/72V	2.8	0.8
23L/25I/66L/72V	2.8	0.7
23I/25I/66L/72L	2.7	−0.3
23L/25V/66I/72L	2.6	0.4
23L/25V/66I/72L	2.6	0.3
23I/25V/66I/72V	2.4	1.2
23I/25I/66I/72V	2.2	0.9
23I/25I/66I/72L	2.2	−0.1
23L/25I/66I/72L	2.2	−0.2
23L/25V/66I/72V	2.0	0.7
23L/25I/66I/72V	2.0	0.6
66/72/92/99		
66L/72V/92V/99L	4.7	1.5
66L/72L/92V/99L	4.7	0.5
66L/72V/92V/99I	4.6	1.3
66I/72V/92V/99I	4.4	1.8
66L/72L/92V/99I	4.4	0.1
66L/72L/92L/99L	4.1	0.1
66L/72V/92L/99L	4.0	1.0
66I/72L/92V/99I	3.8	0.2
66I/72V/92V/99L	3.7	1.2
66L/72V/92L/99I	3.6	0.5
66L/72L/92L/99L	3.6	0.1
66L/72L/92L/99I	3.5	−0.6
66I/72V/92L/99I	3.4	1.0
66I/72V/92L/99L	3.0	0.7
66I/72L/92L/99L	2.4	−0.9
66I/72L/92L/99I	2.1	−1.3

[a] The free energy difference between native and denatured states in the absence of denaturant in units of kcal/mol. Error is estimated to be ±0.1 kcal/mol. Data from references (*10, 11*). [b] $\Delta^n G_{int-1} = \Delta\Delta G - \Sigma\Delta\Delta G_{singles}$. Error is estimated to be ±0.4 kcal/mol for $\Delta^4 G_{int-1}$, ±0.3 kcal/mol for $\Delta^3 G_{int-1}$, and ±0.2 kcal/mol for $\Delta^2 G_{int}$. Data from references (*10, 11*).

argued above, stability must be a selection factor; indeed other workers' experiments determining mutant stability clearly show that it is (*12*). Since a multiple mutant with a favorable interaction energy is likely to be more stable than

one without such favorable interactions, it might be argued that the association of favorable interaction energies in mutants with consensus side chains is fortuitous. This argument ignores several key points.

First, as already noted, the correlation of consensus side chains with favorable interaction energies is stronger than the correlation with stability. Second, speaking of the most stable mutant may obscure the fact that all these mutants are less stable than wild-type, sometimes significantly so. If stability were the key factor, the wild-type sequence has a great advantage. Third, and most importantly, we must consider what having a favorable interaction energy physically means.

There are several different ways to formulate interaction energies (*10, 11, 22*), but the one used here is the simplest: the difference between observed change in stability of a multiple mutant and the sum of the changes in stabilities of the component single mutations that make up that multiple mutant. It should be noted that interaction energies are empirical, not a fundamental thermodynamic property, and that several assumptions are made in their calculation (*24*). As discussed above, the single mutants can alter protein energetics in many ways. Factors such as changes in hydrophobicity and secondary structure preference are to a first approximation independent of the identity of the surrounding residues. By summing the effects of the single mutants and subtracting them from the stability change of a multiple mutant, we can capture, again to a first approximation, the degree to which these side chains interact with each other. For other sets of residues, these interactions can include other factors; for example, the interaction of two charged residues would certainly include electrostatic interactions. A further caution is necessary since it has recently been shown by the elegant work of Taniuchi et al. (*25*) that alterations in the core of yeast iso-2-cytochrome *c* altered stability in part by altering interactions between charged residues on the surface of the protein. Nevertheless, it seems reasonable to make the assumption to a first approximation that among these buried aliphatic residues the principal interaction is packing, which is to say optimizing van der Waals interactions and minimizing cavities.

To put it another way, a favorable interaction energy for a multiple mutant in Table 5 means that those side chains in that mutant have a favorable packing interaction. This does not guarantee that the mutant will be especially stable. There may be unfavorable interactions with other, unmutated residues. In addition, optimizing packing may involve a trade off against other favorable factors such as maximizing hydrophobic transfer free energies. Nevertheless, it is fair to say that in a universe of possible hydrophobic mutants at a given set of buried positions, the multiple mutant with the highest interaction energy has optimized the packing in that local pocket of structure.

Considered in these terms, an apparent selection for favorable interaction energies makes perfect sense. In the final analysis, it must be simple, raw stability that acts as the ultimate selection factor. Yet the mutants here represent that intermediate stage in the evolution of a protein core discussed earlier. None are more stable than wild-type. However, some of them have the *potential* to be more stable, and the interaction energy measures that potential. Those with the best interaction energy at a set of positions have found

the optimal structural solution for that locality. That local optimized structure may interact unfavorably with surrounding residues resulting in a less stable protein, but that extremely favorable local interaction is still present. This means there is a kernel of optimal structure around which further mutation can occur. These further mutations can resolve other unfavorable interactions and allow progress toward another global structural solution, one that is potentially more stable than the starting configuration. These global solutions, rather than the unstable intermediate steps, are fixed in genomes by natural selection and are what we presumably found by examining the nuclease family.

This also explains why some mutants that are relatively stable do not have consensus sequence. These mutants have a low interaction energy. Their good stabilities mean that they have found a reasonable global solution to the packing problem, but the low interaction energies mean they have not reached a local structural optimum and have little room for improvement without radical alteration. In other words, they represent an evolutionary dead end.

This correlation between interaction energy and the consensus sequence can only be interpreted as evidence that optimization of packing is a critical, perhaps even *the* critical, factor guiding the evolution of protein cores. Stability depends on numerous factors as noted before. However, the interaction energies at these sites depend principally on packing, and interaction energy in turn is clearly linked to evolutionary fitness.

Several potential objections can be raised to this idea that we need to address. One possible concern is whether the interaction energies could also be affected by interactions in the denatured state, not simply packing of the native state. We have discussed this issue elsewhere at length (*10*) and feel that this is not an issue with the mutants discussed here.

It might be argued that the correlation is accidental given the fairly high errors estimated for some of these multiterm interaction energies. While this cannot be ruled out, it is most unlikely such a correlation arose by chance among such a large number of mutants. Indeed, we feel the excellent correlation is evidence that our error estimates are well on the conservative side.

Another possible concern lies in the peculiar nature of position 72. Most of the exceptions to the correlation either of stability or, to a lesser degree, of favorable interaction energies with having the consensus side-chain sequence involve this position. The consensus at position 72 is valine. However, in staphylococcal nuclease the observed, most stable, double mutants have a leucine substitution at 72, with the single exception of V66L/I72V. This is less noticeable in the triples and quadruples, where 72 is much more likely to be a valine in the most stable mutant.

Nevertheless, the evolutionary preference observed in the entire nuclease family appears to predict the relative stability among four possible double mutants at most position pairs, except for the four pairs containing position 72 where the evolutionary consensus is for a valine rather than the observed leucine. Valine usually has a better interaction energy at 72, in line with our belief in the importance of packing interaction energy in guiding evolution. Regardless, this still begs the question of whether there is something special about staphylococcal nuclease relative to other homologues that requires a larger side chain at position 72.

Indeed, staphylococcal nuclease does have another difference in this region of sequence. In the wild-type structure, position 72 is the first residue that forms the hydrogen-bonding network in a β-sheet strand. Positions 70 and 71 are in a turn that connects an upstream α-helix and this β-strand. However, examination of the sequence alignment of staphylococcal nuclease and homologues shows that nuclease has an extra amino acid in this turn relative to other homologues. In fact, 23 out of 27 proteins that have valines at the 72 position also have a deletion at position 70. The additional residue at position 70 in staphylococcal nuclease perhaps makes the turn larger and, consequently, may move the main chain at position 72 away from the major hydrophobic core. The extra methylene in the leucine side chain at position 72 may compensate for this effect and stabilize the protein better than a valine residue. This idea is supported by the stabilities of the valine and leucine single mutants at position 72. The leucine substitution at position 72 is 1.0 kcal/mol more stable than valine mutation. Hence, an argument can be made that the increased stability of isoleucine or leucine relative to valine at position 72 is due to a single amino acid insertion at position 70 absent in the usual sequence of nuclease homologues.

On the other hand, if one considers the nine homologues which have the extra amino acid in the turn, there is no preference for isoleucine (4 of 9) versus valine (4 of 9) at 72. Isoleucine is more common if there is an extra amino acid in the turn, but not required. Further, as already noted, the relative stability advantage of a leucine at 72 is clear in the double mutants, but fades away in the triples and the quadruple mutants. This is difficult to understand if there truly is a special need for a larger amino acid in proteins with the insertion. Conversely, this is consistent with the idea that interaction energy measures the potential for further evolution, since as more and more changes are made, unfavorable interactions with the remainder of the protein are lost.

Another potential concern is that the consensus mutants are those which have either lost or at least not gained atoms (expressed in terms of methylene units) relative to other mutants (data not shown). This stands to reason since in our earlier packing study (*10*, *11*), stability effects were found to correlate with the change in methylene groups upon mutation (ΔCH_2), and a loss of methylene units generally leads to a more positive value for the interaction energy. Is the consensus sequence one that merely maintains overall hydrophobic core volume? To some degree, this is a moot question. Volume of the core only matters from a packing perspective. Further, among the quadruple mutants at 23/25/66/72, there are several mutants which do not alter side-chain volume overall, but which are not consensus and do not have particularly favorable interaction energies.

One last possible concern is that accepting the use of isoleucine or leucine as consensus residue at position 99 as laid out above might be inappropriate. There was little choice about this if we want to test against mutant stabilities since the true consensus residue was the side chain present in wild-type, but it is a concern that should be addressed.

While the evidence for the relationship of evolution with interaction energies and thus with packing appears sound, a final experimental test to examine all these potential issues seemed in order.

Staphylococcal Nuclease Mutants with the Consensus Core Sequences. Positions 23, 25, 72, and 92 in staphylococcal nuclease were all substituted simultaneously with the side chains found in the consensus core sequence (23I/25V/72V/92V). Not every possible variant of this set of four residues was constructed as we have detailed elsewhere (*10*, *11*) for other residue sets, but 11 variants were made. In addition to 23I/25V/72V/92V itself, the leucine mutation at position 66 is also made, which at this variable position must also be considered a consensus side chain. To facilitate comparison to Table 5, isoleucine and leucine substitutions at position 99 were also added, although these are not strictly speaking true consensus residues. Further, consensus core mutants with a residue at position 70 deleted were constructed. The mutant 23I/25V/72V/92V/69Δ71 has the same core sequence at these six positions as homologous proteins no. 4 and no. 6 in Table 1. A few other proteins which have changes at the same positions, but which have a single side-chain difference from consensus sequence, were also constructed as controls. The stabilities of these mutants were determined by guanidine hydrochloride denaturation (Table 3). The mutants are all less stable than wild-type staphylococcal nuclease.

Tellingly, all the mutants with residue 70 deleted have lower stability than those without deletion. This disposes of the possibility that there is something truly extraordinary about the relationship of the deletion, the side chain occupying position 72, and protein stability. It is interesting, however, to note that if we count the deletion itself as an interacting mutation (which is in fact what we did in calculating the interaction energies shown in Table 3 for all mutants with the deletion), a mutant with the deletion has slightly higher interaction energies than the corresponding mutant without the deletion. This increase is within the estimated error, but still intriguing.

A more important observation is that all these mutants are much more stable than the stability predicted from summing the effects of the corresponding single mutants. Comparison of the "full" consensus quintuple mutant 23I/25V/66L/72V/92V with the three related mutants, 23I/66L/72V/92V, 23I/25I/66L/72V/92V, and 23L/25V/66L/72V/92V, is interesting. In each of these three mutants, a single position is nonconsensus. In each, the interaction energy is markedly lowered than the full consensus, even though stabilities are nearly as good as or even better than the full consensus quintuple mutant.

The mutants 23I/25V/66L/72V/92V and 23I/25V/66L/72V/92V/69Δ71 have the highest $\Delta^n G_{int-1}$ values: 2.0 and 2.2 kcal/mol, respectively. These interaction energies are extraordinarily high values, well outside any possible error. A comparison with other interaction energies shows this clearly. The $\Delta^n G_{int-1}$ values of the full consensus core mutants (Table 3) are compared to those of packing mutants in Tables 3 and 5 meeting the less rigorous definition of consensus core previously used for Table 5 as well as to all other packing mutants (Tables 3 and 5) in Figure 1. It is strikingly obvious that the full consensus mutants in Table 3 have considerably more favorable interaction energies than those with random sequences at core positions and are generally better than those identified as consensus in Table 5. The only nonconsensus mutants that come close to having such favorable interaction energies are the two quintuple
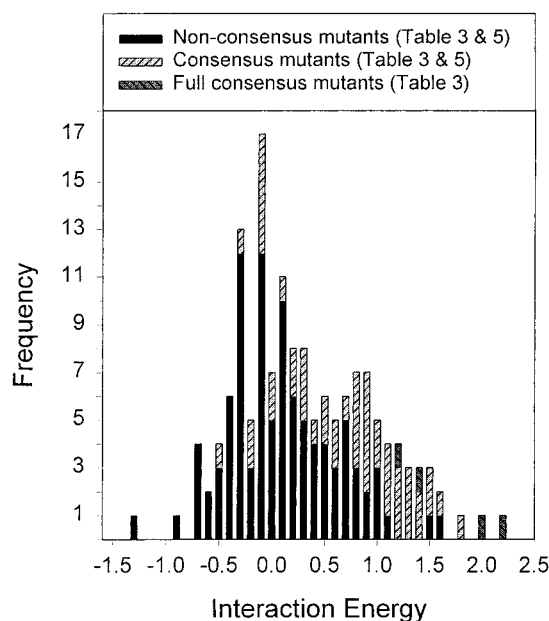
FIGURE 1: Histogram of interaction energies ($\Delta^n G_{int-1}$) for various packing mutants of staphylococcal nuclease. The number of times mutants are found with a given interaction energy is indicated by the frequency axis. The four quadruple and quintuple mutants from Table 3, which have all the side chains at the six positions as a consensus residue, are indicated in dark gray as full consensus. Mutants which are identified in bold type in Table 5 as consensus sequence or those in Table 3 which have partial elements of the full consensus sequence or meet the criteria applied to Table 5 are plotted in light gray. All other mutants from Tables 3 and 5, which all have at least one mutated residue which is a nonconsensus side chain, are indicated in black.

mutants in Table 3 which differ from consensus by only one side chain.

As noted above, it could be argued that this comparison is biased, since in our earlier packing study (*10, 11*), stability effects were found to correlate with the change in methylene groups upon mutation (ΔCH_2), and a loss of methylene units generally leads to a more positive value for the interaction energy. The full consensus core has lost one or two methylene units depending on whether valine 66 is substituted with a leucine or not. Because of the prevalence of valine at the six positions in wild-type nuclease, there are relatively few mutants that have not gained a methylene or two. However, even if the comparison is restricted to mutants that have lost either one or two methylene units (Figure 2A and B), many of which are lesser consensus mutants themselves, it is clear that these new mutants which have the full consensus core are exceptional in their packing as assessed by interaction energies. Increasing the numbers of consensus residues clearly translates into better interaction energies. However, it is also clear that wholesale changes of side chains identity in the core are needed to find a new and reasonably stable packing solution since even the best of these mutants still compares poorly to wild-type in outright, actual protein stability.

CONCLUSION

The above discussion clearly shows that favorable packing interaction energies are found in mutants of staphylococcal
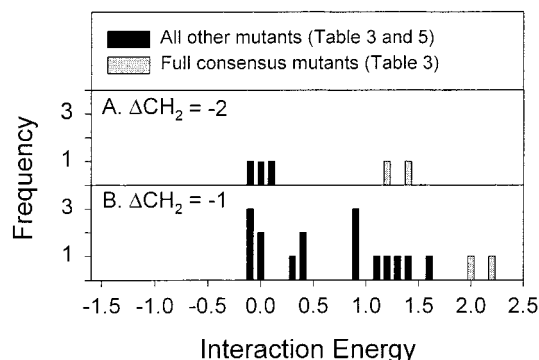

FIGURE 2: Histogram of interaction energies ($\Delta^n G_{int-1}$) for packing mutants of staphylococcal nuclease sorted by the change in the number of methylene units with respect to wild type. The four full consensus core mutants from Table 3 are plotted in light gray. All other mutants from Tables 3 and 5 which have lost two or one methylene units are plotted in black in panels A and B, respectively.

nuclease which have the most common side-chain combinations found in homologous proteins. Packing interaction energies in this sample of mutants appear to be a better predictor of side-chain identity in the hydrophobic core in this family of proteins than other factors, including protein stability. The systematic approach in which every possible permutation is explored leaves little room for alternate interpretations. The final experimental test where an extraordinary interaction energy was found for the complete core appears to rule out all the potential objections we could think of to this interpretation. This is compelling experimental evidence that the packing of protein hydrophobic cores, or more precisely, the energetic effect of packing, is a critical factor in determining the evolutionary fitness of protein hydrophobic cores. This further supports the idea that the importance of packing has been underappreciated relative to other factors that influence the stability of proteins.

ACKNOWLEDGMENT

SUPPORTING INFORMATION AVAILABLE

Complete sequence alignment of staphylococcal nuclease and homologues. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Dill, K. A. (1990) *Biochemistry 29*, 7133−7155.
2. Richards, F. M., and Lim, W. A. (1993) *Q. Rev. Biophys. 26*, 423−498.
3. Hubbard, S. J., and Argos, P. (1995) *Curr. Opin. Biotechnol. 6*, 375−381.
4. Szilagyi, A., and Zavodszky, P. (2000) *Structure Fold. Des. 8*, 493−504.
5. Baldwin, E. P., and Matthews, B. W. (1994) *Curr. Opin. Biotechnol. 5*, 396−402.
6. Lee, B., and Vasmatzis, G. (1997) *Curr. Opin. Biotechnol. 8*, 423−428.
7. Lazar, G. A., and Handel, T. M. (1998) *Curr. Opin. Chem. Biol. 2*, 675−679.
8. Chen, H. M., Li, Y., Panda, T., Buehler, F. U., Ford, C., and Reilly, P. J. (1996) *Protein Eng. 9*, 499−505.
9. Holder, J., Bennett, A. F., Chen, J., Spencer, D. S., Byrne, M. P., and Stites, W. E. (2001) *Biochemistry* (in press).
10. Chen, J., and Stites, W. E. (2001) *Biochemistry* (in press).

11. Chen, J., and Stites, W. E. (2001) *Biochemistry* (in press).
12. Maxwell, K. L., and Davidson, A. R. (1998) *Biochemistry 37*, 16172−16182.
13. Koshi, J. M., and Goldstein, R. A. (1997) *Proteins 27*, 336−344.
14. Finucane, M. D., and Woolfson, D. N. (1999) *Biochemistry 38*, 11613−11623.
15. Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. (2001) *Nucleic Acids Res. 29*, 2994−3005.
16. Shortle, D., Stites, W. E., and Meeker, A. K. (1990) *Biochemistry 29*, 8033−8041.
17. Byrne, M. P., Manuel, R. L., Lowe, L. G., and Stites, W. E. (1995) *Biochemistry 34*, 13949−13960.
18. Stites, W. E., Byrne, M. P., Aviv, J., Kaplan, M., and Curtis, P. M. (1995) *Anal. Biochem. 227*, 112−122.
19. Schwehm, J. M., and Stites, W. E. (1998) *Methods Enzymol. 295*, 150−170.
20. Callebaut, I., and Mornon, J. P. (1997) *Biochem. J. 321*, 125−132.
21. Ponting, C. P. (1997) *Protein Sci. 6*, 459−463.
22. Horovitz, A., and Fersht, A. R. (1990) *J. Mol. Biol. 214*, 613−617.
23. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5590−5594.
24. Mark, A. E., and van Gunsteren, W. F. (1994) *J. Mol. Biol. 240*, 167−176.
25. Taniuchi, H., Shi, Y., Miguel, G. I. S., Ferretti, J. A., Mack, J. W., Fisher, A., Shah, M., Schechter, A. N., and Shiloach, J. (2001) *J. Protein Chem. 20*, 203-215.