THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 31, pp. 22378–22385, August 4, 2006

Comprehensive and Quantitative Mapping of Energy Landscapes for Protein-Protein Interactions by Rapid Combinatorial Scanning*[S]◆

Received for publication, April 20, 2006, and in revised form, June 6, 2006 Published, JBC Papers in Press, June 8, 2006, DOI 10.1074/jbc.M603826200

Gábor Pál[‡1], Jean-Louis K. Kouadio[‡], Dean R. Artis[§], Anthony A. Kossiakoff[‡2], and Sachdev S. Sidhu[§3]

From the [§]Department of Protein Engineering, Genentech Inc., South San Francisco, California 94080 and the [‡]Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, Cummings Life Sciences Center, University of Chicago, Chicago, Illinois 60637

A novel, quantitative saturation (QS) scanning strategy was developed to obtain a comprehensive data base of the structural and functional effects of all possible mutations across a large protein-protein interface. The QS scan approach was applied to the high affinity site of human growth hormone (hGH) for binding to its receptor (hGHR). Although the published structure-function data base describing this system is probably the most extensive for any large protein-protein interface, it is nonetheless too sparse to accurately describe the nature of the energetics governing the interaction. Our comprehensive data base affords a complete view of the binding site and provides important new insights into the general principles underlying protein-protein interactions. The hGH binding interface is highly adaptable to mutations, but the nature of the tolerated mutations challenges generally accepted views about the evolutionary and biophysical pressures governing protein-protein interactions. Many substitutions that would be considered chemically conservative are not tolerated, while conversely, many non-conservative substitutions can be accommodated. Furthermore, conservation across species is a poor predictor of the chemical character of tolerated substitutions across the interface. Numerous deviations from generally accepted expectations indicate that mutational tolerance is highly context dependent and, furthermore, cannot be predicted by our current knowledge base. The type of data produced by the comprehensive QS scan can fill the gaps in the structure-function matrix. The compilation of analogous data bases from studies of other protein-protein interactions should greatly aid the development of computational methods for explaining and designing molecular recognition.

Protein-protein interactions are essential for most biological processes, and they are often characterized by a striking structural plasticity that allows contact points to adapt to conformational changes and multiple amino acid substitutions (1–5). As a result, the biophysics governing protein-protein interactions is extremely complex, and an area of extensive investigation is concerned with establishing a detailed knowledge base that will enhance our understanding of protein-protein associations and enable the development of predictive criteria for engineering novel protein functions.

In this regard, the extensive structure-function data base characterizing the mechanism and energetics of the association of human growth hormone (hGH)[4] with its receptor (hGHR) has provided fundamental insights into general features inherent to protein-protein interactions (1–4, 6–17). The structure-function data base for the hGH-hGHR interaction is probably the most extensive available for any large protein-protein interaction, but nonetheless, it is not comprehensive, and attempts to extract universal trends have been hindered by the incompleteness of the data set.

What is needed to address these issues is a full sampling of all the available structural and chemical diversity afforded by the 20 amino acids that can be genetically encoded at each position in the binding interface. However, for a large protein interface, a comprehensive approach leads to two serious practical demands. It requires the production of hundreds of individual variants by classical mutagenesis methods and the analysis of each protein variant with biophysical techniques. To overcome these technical barriers, we have developed a combinatorial "quantitative saturation (QS) scanning" strategy that enables rapid and facile assessment of the structural and functional effects of all possible point mutations across a large protein-protein interface. We applied the strategy to the interaction between hGH and the hGHR, and we believe that the resulting data base provides the most comprehensive picture of adaptability in a large protein-protein interface that has ever been achieved. Importantly, this comprehensive data base can be collected and analyzed in a small lab setting using available methods and reagents. In practice, QS scanning allows a single scientist to accomplish more than has been possible for a large group using classical mutagenesis methods.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
◆ This article was selected as a Paper of the Week.
[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1 and S2.
[1] Present address: Dept. of Biochemistry, Eötvös University, 1/C Pázmány P. Street, Budapest, Hungary.
[2] To whom correspondence may be addressed: Dept. of Biochemistry and Molecular Biology and Inst. for Biophysical Dynamics, Cummings Life Sciences Center, University of Chicago, 920 East 58th St., Chicago, IL 60637. Tel.: 773-702-9257; Fax: 773-702-0439; E-mail: koss@cummings.uchicago.
[3] To whom correspondence may be addressed: Dept. of Protein Engineering, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080. Tel.: 650-225-1056; Fax: 650-225-3734; E-mail: sidhu@gene.com.

[4] The abbreviations used are: hGH, human growth hormone; hGHR, hGH receptor; ECD, extracellular domain; wt, wild-type; QS, quantitative saturation; SI, specificity index.

With this comprehensive data base in hand, we focused on several universal features of protein-protein interactions that are not satisfactorily addressed by classical methods. For instance, which types of amino acids are the most versatile in forming productive contacts in different packing environments and also which types are the most specific? What is the relative mutational tolerance of the hot spot residues compared with that of bystander residues? Does sequence conservation among species provide insights into the steric and chemical restrictions imposed on side chains at interfaces? Are there functionally homologous hydrophilic side chains, and if so, in which circumstances can they be interchanged? The insights gained from the QS scan data base greatly enhance our understanding of adaptability in protein-protein interfaces, and the results demonstrate that the QS scanning strategy is a powerful, general method for exploring molecular recognition.

EXPERIMENTAL PROCEDURES

Library Construction and Analysis—Wild-type hGH was displayed on M13 bacteriophage as a fusion to the major coat protein. Phage-displayed hGH libraries were constructed, sorted, and analyzed, essentially as described in Ref. 18 and references therein. A hGH-displaying phagemid vector containing TAA stop codons at all positions to be mutated was used as the template for the Kunkel mutagenesis method (19). Mutagenic oligonucleotides were designed to replace the stop codons with degenerate NNK (N = A/C/G/T; K = G/C) codons that collectively code for all 20 natural amino acid residues. Six libraries were designed, with each combining four mutagenic oligonucleotides to introduce the mutations into the hGH sequence. The six mutagenesis reactions were electroporated separately into *Escherichia coli* SS320 (18) and each yielded a library of $\sim 2 \times 10^{10}$ unique members. The libraries were handled separately and each was independently sorted on two different immobilized targets: the extracellular domain (ECD) of the hGHR (hGHR-ECD) or an anti-hGH monoclonal antibody (3F6.B1.4B1) (20). Individual clones from each round of selection were grown in a 96-well format, and the culture supernatants were used directly in phage enzyme-linked immunosorbent assays (18) to detect hormone-phage that bound to either monoclonal antibody 3F6.B1.4B1 or the hGHR-ECD. After two rounds of binding selections, an average of 180 enzyme-linked immunosorbent assay-positive binding clones from each selection were sequenced and subjected to statistical analysis.

Statistical Analysis—The sequences from each selection were aligned and the occurrence of each amino acid at each position was tabulated (supplemental Tables S1 and S2). At each position, the normalized frequency of each amino acid (p) was calculated by correcting for bias in the NNK degenerate codon (21). Shannon Entropy (H) is defined for protein sites by the formula: $H = -[\text{sum}]_{i=1-20} p_i \ln p_i$, where p_i is the fraction of residues at the site that are of type i (22).

The transformed Shannon entropy (TH) was calculated as e^H (23). The specificity index (SI) for each position was calculated as the difference between the TH values for the antibody binding and receptor binding selections.

Sequence conservation statistics were derived using the ConSurf server.

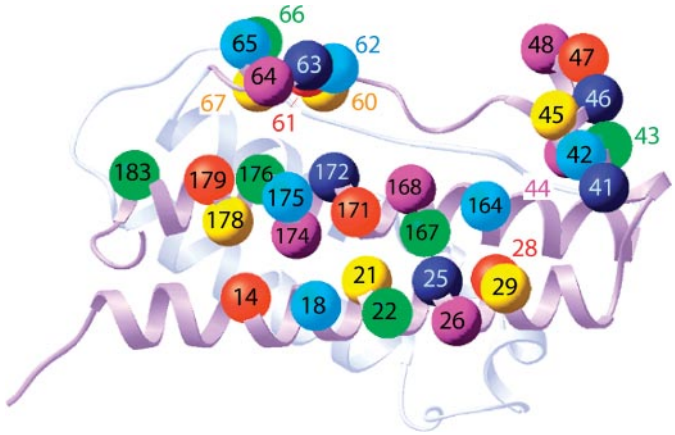
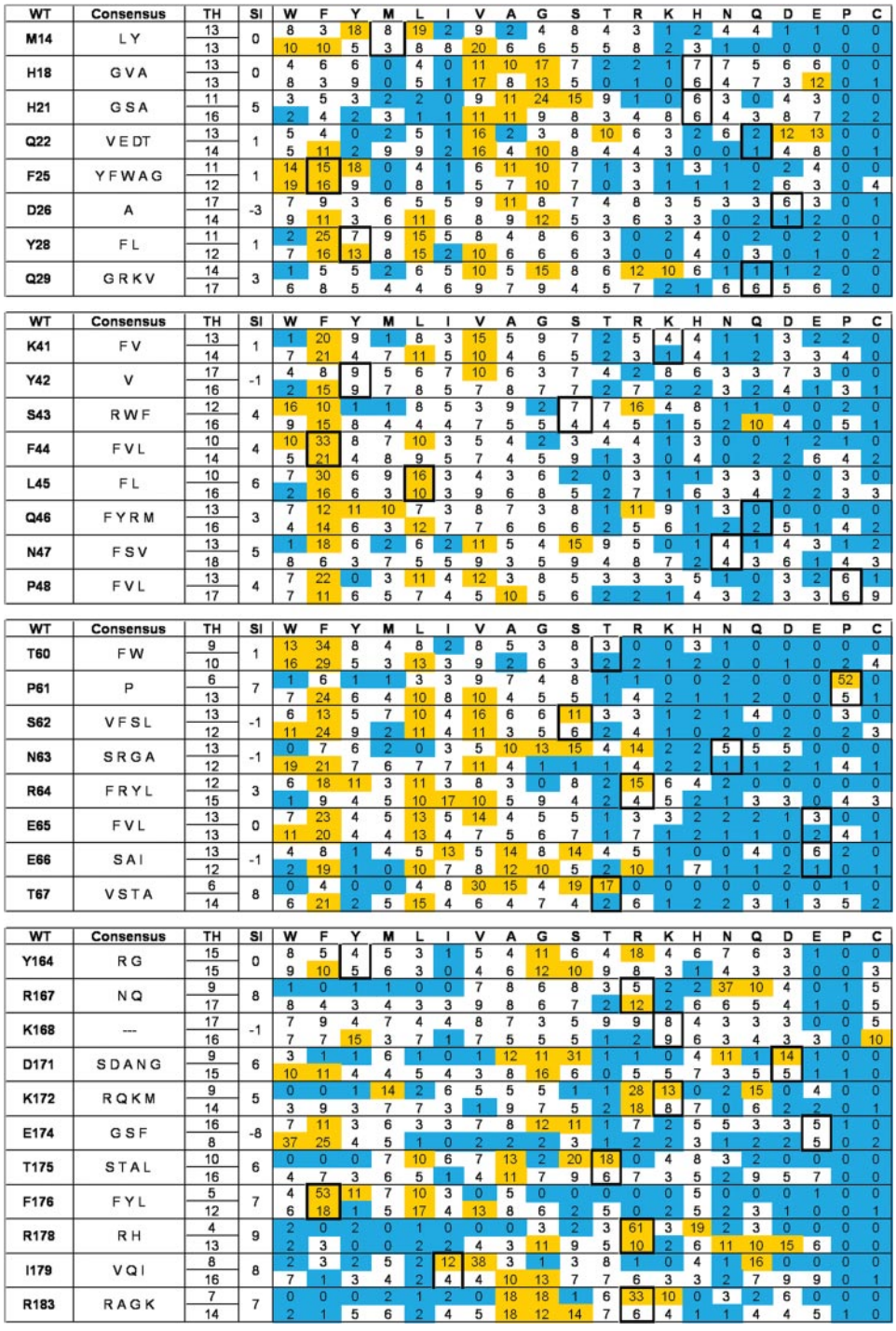


FIGURE 1. **QS scan library design.** The hGH site 1 for binding to the hGHR contains 35 residues distributed across four regions: helices 1 and 4 of the four-helix bundle (residues 14–29 and 164–183) and two connecting loops (residues 41–48 and 60–67). Six libraries were designed to group together five or six residues each in a manner that maximized the distances between residues in any one library (the closest Cα-Cα approach was 7.6 Å between Phe[44] and Pro[48]). The hGH structure is shown as a schematic with the main chain depicted as a ribbon and scanned residues depicted as spheres. Each library contained only one of the six energetically most important residues (Pro[61], Arg[64], Lys[172], Thr[175], Phe[176], or Arg[178]) and residues are color-coded according to the library they share. Structures were derived from the coordinates of the 2:1 hGHR-hGH complex (11) and were rendered in Ribbons (24).

Protein Purification and Affinity Analysis—Individual mutant hGH genes were produced by the Kunkel mutagenesis method (19). Mutant hGH proteins were expressed and purified as described (8, 13). A truncated form of the hGHR-ECD (residues 29–238) with a S237C mutation was expressed and purified as described (25). The purities of all protein samples exceeded 95% as judged by RP-HPLC and SDS-PAGE, and the identities of all mutants were confirmed by mass spectrometry. Receptor binding affinities were measured at 25 °C by surface plasmon resonance on a Biacore 2000 instrument (Biacore Inc., Piscataway, NJ). The hGHR-ECD was immobilized through the engineered cysteine residue on a Pioneer C1 sensor chip at ~50 response units, and unreacted functional groups were blocked with glutathione, as described (25). For kinetic analysis, 2-fold serial dilutions of hGH variants in HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Tween 20, pH 7.4) were injected, and binding responses to the hGHR-ECD were corrected by subtracting responses from running buffer injections (no hGH injected) and by subtracting responses on a blank flow cell (no hGHR-ECD immobilized). The measurements were done in triplicate. For kinetic analysis, a 1:1 Langmuir model with separate fitting of k_{on} and k_{off} was used. The K_d value was estimated from the ratios of k_{on} and k_{off}.

RESULTS

Principles of QS Scanning—We aimed to develop a saturation scanning strategy that would provide quantitative information about the structural and functional effects of all possible amino acid substitutions within a protein binding site. For the high affinity receptor-binding site of hGH (site 1), this involves 35 residues (11). Thus, the site was broken down into six libraries covering five or six residues each (Fig. 1), which could be completely represented by the diversities obtainable by phage display (10^{10}). So as to minimize potential cooperative interactions between mutations, the members within each library were cho-



WT	Consensus	TH	SI	W	F	Y	M	L	I	V	A	G	S	T	R	K	H	N	Q	D	E	P	C
M14	LY	13	0	8	3	18	8	19	2	9	2	4	8	4	3	1	2	4	4	1	1	0	0
		13		10	10	5	3	8	8	20	6	6	5	5	8	2	3	1	0	0	0	0	0
H18	GVA	13	0	4	6	6	0	4	0	11	10	17	7	2	2	1	7	7	5	6	6	0	0
		13		8	3	9	0	5	1	17	8	13	5	0	1	0	6	4	7	3	12	0	1
H21	GSA	11	5	3	5	3	2	2	0	9	11	24	15	9	1	0	6	3	0	4	3	0	0
		16		2	4	2	3	1	1	11	11	9	8	3	4	8	6	4	3	8	7	2	2
Q22	VEDT	13	1	5	4	0	2	5	1	16	2	3	8	10	6	3	2	6	2	12	13	0	0
		14		5	11	2	9	9	2	16	4	10	8	4	4	3	0	0	1	4	8	0	1
F25	YFWAG	11	1	14	15	18	0	4	1	6	11	10	7	1	3	1	3	1	0	2	4	0	0
		12		19	16	9	0	8	1	5	7	10	7	0	3	1	1	1	2	6	3	0	4
D26	A	17	-3	7	9	3	6	5	5	9	11	8	7	4	8	3	5	3	3	6	3	0	1
		14		9	11	3	6	11	6	8	9	12	5	3	6	3	3	0	2	1	2	0	0
Y28	FL	11	1	2	25	7	9	15	5	8	4	8	6	3	0	2	4	0	2	0	2	0	1
		12		7	16	13	8	15	2	10	6	6	6	3	0	0	4	0	3	0	1	0	2
Q29	GRKV	14	3	1	5	5	2	6	5	10	5	15	8	6	12	10	6	1	1	1	2	0	0
		17		6	8	5	4	4	6	9	7	9	4	5	7	2	1	6	6	5	6	2	0

WT	Consensus	TH	SI	W	F	Y	M	L	I	V	A	G	S	T	R	K	H	N	Q	D	E	P	C
K41	FV	13	1	1	20	9	1	8	3	15	5	9	7	2	5	4	4	1	1	3	2	2	0
		14		7	21	4	7	11	5	10	4	6	5	2	3	1	4	1	2	3	3	4	0
Y42	V	17	-1	4	8	9	5	6	7	10	6	3	7	4	2	8	6	3	3	7	3	0	0
		16		2	15	9	7	8	5	7	8	7	7	2	7	2	2	3	2	4	1	3	1
S43	RWF	12	4	16	10	1	1	8	5	3	9	2	7	7	16	4	8	1	1	0	0	2	0
		16		9	15	8	4	4	4	7	5	5	4	4	5	1	5	2	10	4	0	5	1
F44	FVL	10	4	10	33	8	7	10	3	5	4	2	3	4	4	1	3	0	0	1	2	1	0
		14		5	21	4	8	9	5	7	4	5	9	1	3	0	4	0	2	2	6	4	2
L45	FL	10	6	7	30	6	9	16	3	4	3	6	2	0	3	1	1	3	3	0	0	3	0
		16		2	16	6	3	10	3	9	6	8	5	2	7	1	6	3	4	2	2	3	3
Q46	FYRM	13	3	7	12	11	10	7	3	8	7	3	8	1	11	9	1	3	0	0	0	0	0
		16		4	14	6	3	12	7	7	6	6	6	2	5	6	1	2	2	5	1	4	2
N47	FSV	13	5	1	18	6	2	6	2	11	5	4	15	9	5	0	1	4	1	4	3	1	2
		18		8	6	3	7	5	5	9	3	5	9	4	8	7	2	4	3	6	1	4	3
P48	FVL	13	4	7	22	0	3	11	4	12	3	8	5	3	3	3	5	1	0	3	2	6	1
		17		7	11	6	5	7	4	5	10	5	6	2	2	1	4	3	2	3	3	6	9

WT	Consensus	TH	SI	W	F	Y	M	L	I	V	A	G	S	T	R	K	H	N	Q	D	E	P	C
T60	FW	9	1	13	34	8	4	8	2	8	5	3	8	3	0	0	3	1	0	0	0	0	0
		10		16	29	5	3	13	3	9	2	6	3	2	2	1	2	0	0	1	0	2	4
P61	P	6	7	1	6	1	1	3	3	9	7	4	8	1	1	0	0	2	0	0	0	52	0
		13		7	24	6	4	10	8	10	4	5	5	1	4	2	1	1	2	0	0	5	1
S62	VFSL	13	-1	6	13	5	7	10	4	16	6	6	11	3	3	1	2	1	4	0	0	3	0
		12		11	24	9	2	11	4	11	3	5	6	2	4	1	0	2	0	2	0	2	3
N63	SRGA	13	-1	0	7	6	2	0	3	5	10	13	15	4	14	2	2	5	5	5	0	0	0
		12		19	21	7	6	7	7	11	4	1	1	1	4	2	2	1	1	2	1	4	1
R64	FRYL	12	3	6	18	11	3	11	3	8	3	0	8	2	15	6	4	2	0	0	0	0	0
		15		1	9	4	5	10	17	10	5	9	4	2	4	5	2	1	3	3	0	4	3
E65	FVL	13	0	7	23	4	5	13	5	14	4	5	5	1	3	3	2	2	2	1	3	0	0
		13		11	20	4	4	13	4	7	5	6	7	1	7	1	2	1	1	0	2	4	1
E66	SAI	13	-1	4	8	1	4	5	13	5	14	8	14	4	5	1	0	0	4	0	6	2	0
		12		2	19	1	0	10	7	8	12	10	5	2	10	1	7	1	1	2	1	0	1
T67	VSTA	6	8	0	4	0	0	4	8	30	15	4	19	17	0	0	0	0	0	0	0	1	0
		14		6	21	2	5	15	4	6	4	7	4	2	6	1	2	2	3	1	3	5	2

WT	Consensus	TH	SI	W	F	Y	M	L	I	V	A	G	S	T	R	K	H	N	Q	D	E	P	C
Y164	RG	15	0	8	5	4	5	3	1	5	4	11	6	4	18	4	6	7	6	3	1	0	0
		15		9	10	5	6	3	0	4	6	12	10	9	8	3	1	4	3	3	0	0	3
R167	NQ	9	8	1	0	1	1	0	0	7	8	6	8	3	5	2	2	37	10	4	0	1	5
		17		8	4	3	4	3	3	9	8	6	7	2	12	2	6	6	5	4	1	0	5
K168	---	17	-1	7	9	4	7	4	4	8	7	3	5	9	9	8	4	3	3	3	0	0	5
		16		7	7	15	3	7	1	7	5	5	5	1	2	9	6	3	4	3	3	0	10
D171	SDANG	9	6	3	1	1	6	1	0	1	12	11	31	1	1	0	4	11	1	14	1	0	0
		15		10	11	4	4	5	4	3	8	16	6	0	5	5	7	3	5	5	1	1	0
K172	RQKM	9	5	0	0	1	14	2	6	5	5	5	1	1	28	13	0	2	15	0	4	0	0
		14		3	9	3	7	7	3	1	9	7	5	2	18	8	7	0	6	2	2	0	1
E174	GSF	16	-8	7	11	3	6	3	3	7	8	12	11	1	7	2	5	5	3	3	5	1	0
		8		37	25	4	5	1	0	2	2	2	3	1	2	2	3	1	2	2	5	0	2
T175	STAL	10	6	0	0	0	7	10	6	7	13	2	20	18	0	4	8	3	2	0	0	0	0
		16		4	7	3	6	5	1	4	11	7	9	6	7	3	5	2	9	5	7	0	0
F176	FYL	5	7	4	53	11	7	10	3	0	5	0	0	0	0	0	5	0	0	0	1	0	0
		12		6	18	1	5	17	4	13	8	6	2	5	0	2	5	2	3	1	0	0	1
R178	RH	4	9	2	0	2	0	1	0	0	0	3	2	3	61	3	19	2	3	0	0	0	0
		13		2	3	0	0	2	2	4	3	11	9	5	10	2	6	11	10	15	6	0	0
I179	VQI	8	8	2	3	2	5	2	12	38	3	1	3	8	1	0	4	1	16	0	0	0	0
		16		7	1	3	4	2	4	4	10	13	7	7	6	3	3	2	7	9	9	0	1
R183	RAGK	7	7	0	0	0	2	1	2	0	18	18	1	6	33	10	0	3	2	6	0	0	0
		14		2	1	5	6	2	4	5	18	12	14	7	6	4	1	1	4	4	5	1	0

FIGURE 2. **The QS scan data base.** At each position (*WT*), following selection for binding to the hGHR (*top row*) or an anti-hGH antibody (*bottom row*), the percent occurrence of each amino acid type was calculated after normalization for codon bias. The wt occurrences are *boxed* and values are colored as follows: *yellow*, ≥10%; *blue*, ≤2%. The consensus lists the amino acids over-represented (≥10%) in the receptor binding selection data set. *TH* diversity (*TH*) and SI values were calculated as described under "Materials and Methods."

sen to be as far apart as possible in the three-dimensional structure. In addition, each library contained only one of the six energetically most important residues, as determined by alanine scanning (13, 14). With these principles, we ensured that the interface was comprehensively scanned and also that the libraries were designed to minimize the likelihood of interactions that would be very different from the wild-type (wt) mechanism of binding.

Structural Constraints on Sequence Space—To determine the constraints on sequence space imposed by the structural requirements of the hGH four-helix bundle fold, we selected each library for binding to a structure-specific antibody that recognizes an epitope on the side of hGH opposite to site 1 (20). Variants from each library selection were sequenced, and the amino acid frequency distribution at each randomized position was determined (Fig. 2). The data were used to compute the transformed Shannon entropy (*TH*) value at each position, which measures the degree of randomness within a population and has been used to quantify diversity within antibodies, T-cell receptors, and other proteins (23, 26–28). For a frequency distribution of 20 amino acids, the *TH* varies between values of 1 and 20 for positions that are completely conserved or completely random, respectively. Following selection for the native fold, the average *TH* value for the 35 positions is quite high (14 ± 2). This indicates that the hGH fold is highly tolerant to mutations at these solvent-exposed positions, and the low standard deviation suggests that there is little difference between the mutational tolerance of the individual positions.

The deviation from complete randomness can be explained by three major biases. First, cysteine is scarce at almost all positions, and second, proline is scarce within the α-helical regions. Both of these findings are in accord with general views of protein stability, as cysteines are likely to interfere with the native disulfides of hGH and prolines are known to destabilize alpha-helices. A third bias is an overabundance of hydrophobic residues at the expense of hydrophilic residues. This finding is unexpected, as it contradicts popular views about protein folding and stability, which hold that solvent-exposed hydrophobes should be disfavored (29).

Overall, the distributions are not biased in favor of the wild-type, which shows average abundance at 19 positions and above or below average abundance at 7 or 9 positions, respectively.

Functional Constraints on Sequence Space—We next determined the constraints on sequence space imposed by selection for the binding of hGH to the hGHR. This selection combines

TABLE 1
Kinetic analysis of hGH mutants binding to the hGHR

hGH mutant	$k_{on} \times 10^5$	$k_{off} \times 10^{-4}$	K_d	$K_{d,wt}/K_{d,mut}$ [a]	
				Calculated	Predicted
	$M^{-1} s^{-1}$	s^{-1}	nM		
wt	2.34 ± 0.07	3.64 ± 0.08	1.56 ± 0.08	1	
R167N	2.78 ± 0.16	0.32 ± 0.11	0.12 ± 0.05	13	15
D171S	2.52 ± 0.17	1.57 ± 0.03	0.63 ± 0.06	2.5	2
E174G	2.73 ± 0.09	2.75 ± 0.03	1.01 ± 0.04	1.5	6
E174A	3.57 ± 0.11	1.24 ± 0.11	0.35 ± 0.03	4.5	4
E174S	2.82 ± 0.12	1.07 ± 0.08	0.38 ± 0.04	4.2	4
I179V	1.83 ± 0.09	1.14 ± 0.10	0.62 ± 0.03	2.5	3
R167N/E174G/I179V	1.97 ± 0.15	0.12 ± 0.05	0.06 ± 0.02	27	
R167N/E174A/I179V	2.61 ± 0.11	0.12 ± 0.01	0.05 ± 0.01	34	
R167N/D171S/E174G/I179V	2.38 ± 0.16	0.15 ± 0.06	0.06 ± 0.03	24	

[a] Surface plasmon resonance was used to determine k_{on} and k_{off} values. The K_d values were calculated as k_{off}/k_{on} and were used to derive the calculated $K_{d,wt}/K_{d,mut}$ values. The predicted $K_{d,wt}/K_{d,mut}$ values were derived from the QS scan data, as described for shotgun alanine-scanning (6), from the observed frequencies of the wt (p_{wt}) or mutant (p_{mut}) after selection for binding to the hGHR or the anti-hGH antibody (Fig. 2), as follows: $K_{d,wt}/K_{d,mut} = (p_{mut,hGHR}/p_{wt,hGHR})(p_{wt,antibody}/p_{mut,antibody})$.

two types of constraints: a structural constraint for a stable native fold and a functional constraint for a binding site capable of receptor recognition. As the structural constraints had been independently assessed from the antibody binding selection (see above), we could quantify the additional constraints imposed by function with the "SI," a metric that we defined as the difference between the *TH* values for the antibody binding and receptor binding selections (Fig. 2). As expected, a positive mean SI value (3 ± 4) across the 35 scanned positions indicates that receptor binding function imposes additional constraints above and beyond those imposed by structural demands, and the large standard deviation indicates that these functional constraints are position-specific. Interestingly, the correlation between TH values and sequence conservation across species is not strong. For instance, Lys^{168}, which is highly conserved across species has the largest value ($TH = 17$), indicating that most amino acid types can substitute equally well for lysine without affecting the overall binding.

Kinetic Measurements Validate the Accuracy of QS Scanning—We tested the accuracy with which the QS scanning strategy predicts the functional effects of single amino acid substitutions. The statistical analysis suggested that the wild-type residue is suboptimal for receptor binding at a number of positions and that affinity could be improved with one or several types of amino acid substitutions. We purified six hGH protein variants, each of which contained a single point mutation predicted to improve affinity. The kinetics of wild-type hGH and each of the mutants for binding to the hGHR were determined by surface plasmon resonance (Table 1). The results unequivocally demonstrate that the QS scan data are highly reliable, as the affinity of each mutant was improved over that of wt hGH and, for five of the six mutants, the measured improvements were virtually identical with the predictions. In addition, the QS scan is highly sensitive, as even effects as small as 2-fold were predicted accurately.

We also constructed two triple mutants in which the mutated sites were buffered by at least two intervening residues in the three-dimensional structure. The affinities of these mutants were improved ~30-fold relative to that of wt hGH (Table 1). Although the effects of the mutations did not combine in a completely additive manner, the affinities of the triple mutants exceeded those of the single mutants. Interestingly, the addition of a fourth mutation (D171S) did not improve affinity further suggesting that additivity principles begin to deteriorate when too many mutations are incorporated in a proximal area.

DISCUSSION

There exists an extensive literature describing structure-function relationships for a large number of protein-protein interactions. In particular, the use of alanine scanning has been extremely powerful for determining the binding contributions of individual residues (3, 6, 8, 9, 13, 14, 17, 25, 30–38). The compilation of these data has resulted in the paradigm that binding energy is generally focused in a limited region of the interface, termed the "hot spot" of binding energy. However, while alanine scanning provides information about which residues are important, it does not provide insights into the functional consequences of substitution by the full spectrum of chemical and conformational diversity within the 20 natural amino acids. This question is fundamental to appreciating the possible basis for evolutionary conservation among binding site residues, as well as for understanding the underlying biophysical parameters governing the overall energetics of the interaction.

While the isolated chemical properties of each of the 20 amino acids can be described by well established physicochemical criteria, it is understood that these properties are highly context dependent and will certainly be influenced by the local dielectric and packing environments within protein interfaces. This context dependence greatly limits the ability to predict how a particular substitution will affect binding, even in cases of supposedly conservative changes. Unfortunately, the current mutagenesis data bases provide a very sparse matrix of structure-function effects, which is inadequate for even a qualitative description of the binding energy landscape.

In the case of hGH site 1, which encompasses 35 positions, the compilation of a comprehensive functional matrix by conventional methods is a Herculean task involving the construction and analysis of close to 700 individual protein variants. To overcome the technical barriers inherent to such an exhaustive analysis, we developed a novel saturation mutagenesis strategy that is rapid yet quantitative. The QS scanning method builds a complete mutagenesis data base through a simple process that is not time consuming and uses standard molecular biology techniques. The importance of completeness cannot be over emphasized, as this data set allows for not only the identifica-

tion of many overall trends but also reveals subtle effects that have been missed by classical approaches. Importantly, analysis of trends within the data suggest that it is not feasible to identify a smaller "representative" subset of chemically and structurally diverse amino acid types that by themselves could adequately describe the overall character of the binding surface. In other words, shortcuts are not possible.

Substitution of Hydrophilic Residues—The site 1 surface of hGH is solvent exposed and approximately two-thirds of the side chains are hydrophilic. Surprisingly, the data indicate that hydrophobic groups are preferred at almost every position, including positions occupied by wt hydrophilic side chains involved in hydrogen bonding. Of the nine hydrophilic positions that are evolutionarily conserved across species, most can be substituted by hydrophobic side chains. Conversely, there are no hydrophilic side chains that productively substitute for hydrophobes, even at positions where such side chains are observed in the sequences of related species. Overall, hydrophilic groups can generally be replaced by small hydrophobic groups but are resistant to replacement by larger ones.

Conservative Substitutions Do Not Exist—The data suggest that the accepted definitions of homologous pairs and clusters for hydrophilic side chains (*e.g.* Asn substituted by Gln or Asp) are not operable in this protein-protein interface. Likewise, there is no systematic connection between physicochemically related hydrophobic pairs (*e.g.* Val/Ile or Ile/Leu), although interchangeability among hydrophobic groups is better accommodated. A point of clarification should be made about the apparent bias toward hydrophobic substitutions in the particular case of hGH. In a number of instances, hydrophilic groups (even some that are involved in hydrogen bonding) interact with the receptor through the hydrophobic portions of their side chains. Thus, it appears that hydrophobic substitutions tend to accentuate the role of the van der Waals contacts without apparently paying a large penalty for the loss of hydrogen bonds.

The SI—The SI value provides an unbiased assessment of the number of residue types that could functionally contribute to binding at each position in the interface, and thus, it is a powerful metric to gauge not only the specificity requirements but also the plasticity of the interface. The SI identifies positions where only a single residue type is acceptable, as in the case of Pro^{61} where there are specific conformational requirements and also positions where specificity has more to do with size rather than chemical character (*e.g.* Thr^{67} and Thr^{175}). In the case of position 183, Arg and Lys are highly preferred but Ala and Gly are the next best substitutions, suggesting that this position has an "all or nothing" character. In contrast, Arg is highly preferred at position 178, but Lys is strongly selected against. This is an interesting finding, since a Lys residue is found at this position in growth hormones from a number of mammalian species, and this is one of several positions that influence the cross species specificity characteristics of these hormones.

In general, the hot spot residues defined by alanine scanning have the highest SI values. This indicates that these positions require the highest degree of specificity, and furthermore, the wild-type residue type is usually preferred. Large negative SI values are rare and indicate positions where there is strong expression bias for residue types that are poor for binding. This situation occurs at position 174 where there is a strong expression bias for Trp or Phe, which perhaps increases protein stability, but these side chains are too bulky to fit in the interface without large disruptions in the structure.

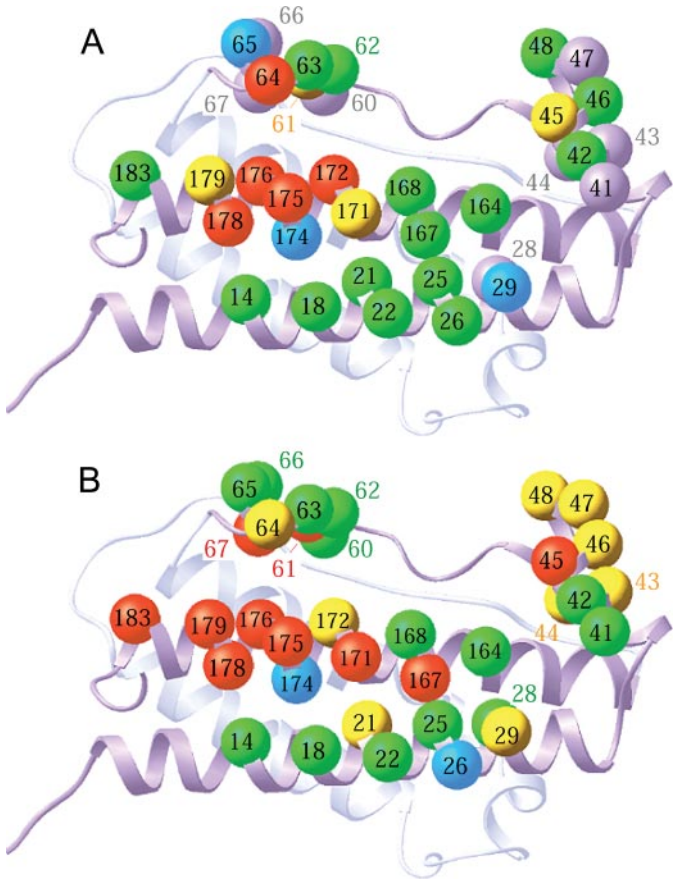


FIGURE 3. **Comparison of alanine scanning and QS scanning.** The hGH structure is shown as a schematic with the main chain depicted as a ribbon and scanned residues depicted as spheres. *A*, the results of alanine scanning mapped onto hGH site 1 (12, 18). The residues are colored according to the $\Delta\Delta G_{Ala\text{-}wt}$ values (in kcal/mol), as follows: *cyan* < -0.4; $-0.4 \leq$ *green* < 0.4; $0.4 \leq$ *yellow* < 1.0; *red* ≥ 1.0; *gray*, unscanned. *B*, the results of QS scanning mapped onto hGH site 1. The residues are colored according to the SI values, as follows: *cyan* < -2; $-2 <$ *green* < 3; $3 \leq$ *yellow* < 6; *red* ≥ 6. Structures were derived from the coordinates of the 2:1 hGHR-hGH complex (11) and were rendered in Ribbons (24).

The SI is an extremely robust probe of the energy surface, and it is significantly more powerful than alanine scanning for assessing the functional adaptability of the binding site. Fig. 3 compares the functional epitope derived by alanine scanning to that based on the SI values. While there is general correspondence, as the high SI values superimpose on the alanine scanning hot spot residues, the epitope defined by SI values is somewhat more expansive. This is because, at several "high specificity" positions, the preferred residue type is not the wt amino acid, a characteristic evident at position 167 where Asn and Gln are highly preferred over the wt Arg.

Characteristics of the Binding Energy Surface—Analysis of a complete QS scan data base demonstrates just how rough the binding energy surface is, and it reveals the inadequacies of sparse structure-function data bases for predicting the func-

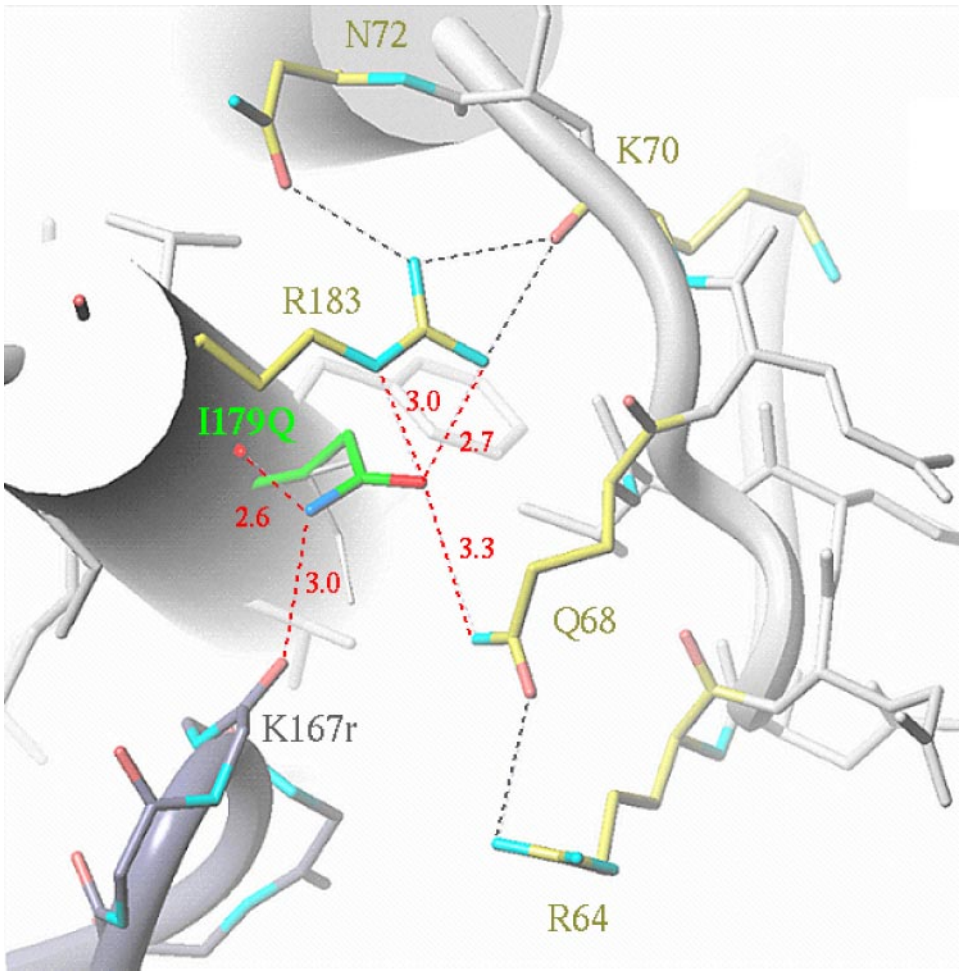


FIGURE 4. **Modeled hydrogen-bonding interactions between the side chain of Gln^{179} and Gln^{68} and Arg^{183} of hGH and Lys^{167} of the hGHR (labeled as *K167r*).** The side chain of Gln^{179} is modeled in a low energy conformation and is also positioned to form a hydrogen bond to its own main chain carbonyl group. The side chain of the wt Ile^{179} is shown in *white*. The hydrogen-bonding network involving the side chains of Arg^{64}, Gln^{68}, and Arg^{183} is important in stabilizing the 60's loop, which plays a role in the binding of hGH site 1 to the hGHR. Hydrogen bonds are shown as *dashed lines* with the interatomic distances shown in angstroms. Hydrogen bonds involving Gln^{179} are colored *red*, and others are colored *black*. The model was built using the coordinates of a 2:1 hGHR-hGH complex (Protein Data Bank entry 3HHR).

tional effects of small differences between homologous side chain types. At one end of the spectrum, there are a substantial number of positions that can tolerate many amino acid types, while at the other end there are positions that display significant specificity requirements. It is at these positions of broad and narrow specificity where the character of the energy surface is fairly well defined, at least in a qualitative sense.

Between these two extremes are areas where the energetics are very sensitive to small perturbations that are not easily explained by relationships between side chain characteristics and structural environment. For example, the wt Ile is close to optimal at position 179, but Val is even better. In fact, among the growth hormones of different species, Val is found in the majority of cases. Surprisingly, the close homologue Leu is a poor substitute, whereas Gln is almost as good as Ile, but Asn is not tolerated at all. Taken together, these data suggest that two distinct packing scenarios can exist in this region of the interface. In the first, the volume of the hydrophobic hole occupied by the residue at position 179 is just large enough to accommodate an Ile group, but the structure can relax in an energetically positive way to accommodate a Val side chain. Ala is too small and Leu is too big, and each pays a van der Waals penalty. The hydrophilic Gln side chain clearly cannot pack into the hydrophobic hole, but it can be modeled to form a good hydrogen-bonding interaction with Gln^{68} (Fig. 4). The side chain of Gln^{68} forms a hydrogen bond with Arg^{64} and plays an important role in stabilizing the so-called 60's loop in a conformation suitable for binding to the hGHR. Presumably, the interaction mediated by Gln179 further stabilizes this loop. The Asn side chain is likely too short to do this, and consequently, it is not tolerated.

There are numerous examples of other seemingly subtle changes that affect significant changes in function. In particular, as predicted from the QS scan, the mutation R167N results in a dramatic increase in affinity. Interestingly, the R167N mutation is also found in a phage-derived hGH variant (hGH$_v$) that contains a total of 15 mutations and binds to the hGHR ~400-fold tighter than wt hGH (10). It had been observed previously that Arg^{167} in wt hGH plays an unusually small role in binding, considering that it forms a salt bridge with Glu^{127} of the hGHR (12). The crystal structure of hGH$_v$ in complex with the hGHR reveals that the substitution of Arg^{167} by Asn eliminates the salt bridge with Glu^{127}, but the Asn side chain does not make new contacts with the receptor (7). These observations suggest that Arg^{167} has a null effect on receptor binding and that Asn^{167} provides some positive influence in an apparently indirect way. It is also worth noting that hGH binds to not only the hGHR but also to the prolactin receptor (hPRLR); in contrast to the interaction with the hGHR, the interaction with the hPRLR is critically dependent on Arg^{167}, as substitution by Ala reduces binding affinity by ~700-fold (12).

In the QS scan data set, position 171 is biased in favor of Ser over the wt Asp. The D171S mutation is also found in the high affinity hGH$_v$ (10), where the Ser^{171} side chain forms a hydrogen bond with the Trp^{104} side chain of the hGHR (7). It is noteworthy that, in most species, position 171 is occupied by His, a residue that is not significantly represented in the QS scan. This apparent discrepancy can be explained on evolutionary grounds, since it has been determined that position 171 was a site for coevolution between growth hormones and their receptors during the transition between old world and new world monkeys (1). In lower species, there is a pairing of His at

position 171 in the hormones and a Leu at position 43 in the receptors. In primates, however, position 171 is occupied by an Asp that forms a salt bridge with an Arg at position 43 of the cognate receptors (11). Interestingly, primate growth hormones can bind receptors containing a Leu at position 43, but the growth hormones of lower species containing His^{171} cannot bind primate receptors containing Arg^{43}. Thus, it has been speculated that coevolution occurred first by mutation of position 171 in the hormone (H171D) followed by mutation of position 43 in the receptor (L43R) (1).

The picture that is developed from the QS scan data is that of an energy landscape composed of three distinctly different regions. The first is the high specificity region that correlates reasonably well with the binding hot spot paradigm but is better described by the SI epitope. The second is the low specificity region where many residue types can be inserted without large deleterious effects. The third consists of positions that are sensitive to subtle changes and are not optimized for binding function. It is within this last region that the detailed nature of the binding energy surface is poorly characterized and not well understood.

We believe that the type of quantitative data produced from the comprehensive QS scan begins to fill the gaps in the structure-function matrix. While this represents a significant advance, many similar studies will be required to statistically average relationships between side chain character and structural context. The next major challenge will be to synthesize these types of data into a computational framework that can predict the important subtleties that are inherent properties in molecular recognition.

CONCLUSIONS

Among the systems that have been extensively studied by mutagenesis and structural approaches, hGH has proven to be an extremely robust model for identifying structure-function relationships involved in the formation and maintenance of protein-protein interactions. Therefore, we believe that the trends seen in this study are likely to be applicable to protein-protein interfaces in general. Below, we summarize the conclusions derived from this study.

Even though the interaction between hGH and the hGHR is arguably the most exhaustively studied protein-protein interaction, the previously existing mutational data set represents too sparse a sampling of structure-function space to have general predictive power.

The QS scan reveals that the hGH binding interface is extremely tolerant to mutations. In many cases, even substitutions expected to introduce steric and chemical features incompatible with the wild-type residue are tolerated, and these outliers provide critical data points for characterizing the binding energy surface.

The SI metric identifies functionally sensitive residues with higher resolution than alanine scanning. The QS scan matrix assesses the effect of every amino acid type at every position within a binding site in a quantitative manner, and it can predict 2-fold effects without requiring structural knowledge.

There are a number of examples where the wild-type residue leads to poor protein expression and cases where mutations that greatly increase expression reduce function, implying that expression levels and function are often uncoupled.

Hydrophobic effects dominate the function of hGH site 1, and consequently, the hGH-hGHR interaction can be maintained under mutational pressure, because hydrophobic residues are more interchangeable than hydrophilic residues.

The QS scan matrix challenges generally accepted views of conservative mutations in molecular recognition, especially whether they exist at all in the case of hydrophilic side chains. For hydrophilic groups, there are no trends among the acid and amide side chains and only weak trends between Arg and Lys or Ser and Thr. Surprisingly, there are only weak functional connections between Tyr and Phe, and Val, Ile, and Leu show only limited homology in functional space.

An important observation is that sequence conservation across species is a poor predictor of the chemical character of tolerated substitutions in a protein-protein interface. The concept that evolution is a dynamic process that is ever fine-tuning an interaction is probably incorrect. Conservation across species does not necessarily mean that a particular residue is important for structure or function but rather may reflect other constraints imposed by the requirements of the complex biological system.

The implications of this last point, when viewed in the context of the other conclusions, is that the design of a functional hGH molecule based strictly on biophysical principles would be very different from that of the natural hGH based on evolutionary pressures.

REFERENCES

1. Yi, S., Bernat, B., Pal, G., Kossiakoff, A., and Li, W. H. (2002) *Mol. Biol. Evol.* **19,** 1083–1092
2. Pal, G., Fong, S. Y., Kossiakoff, A. A., and Sidhu, S. S. (2005) *Protein Sci.* **14,** 2405–2413
3. Kouadio, J. L., Horn, J. R., Pal, G., and Kossiakoff, A. A. (2005) *J. Biol. Chem.* **280,** 25524–25532
4. Kossiakoff, A. A. (2004) *Adv. Protein Chem.* **68,** 147–169
5. James, L. C., and Tawfik, D. S. (2003) *Trends Biochem. Sci.* **28,** 361–368
6. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A., and Sidhu, S. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8950–8954
7. Schiffer, C., Ultsch, M., Walsh, S., Somers, W., de Vos, A. M., and Kossiakoff, A. (2002) *J. Mol. Biol.* **316,** 277–289
8. Pal, G., Ultsch, M. H., Clark, K. P., Currell, B., Kossiakoff, A. A., and Sidhu, S. S. (2005) *J. Mol. Biol.* **347,** 489–494
9. Pal, G., Kossiakoff, A. A., and Sidhu, S. S. (2003) *J. Mol. Biol.* **332,** 195–204
10. Lowman, H. B., and Wells, J. A. (1993) *J. Mol. Biol.* **234,** 564–578
11. de Vos, A. M., Ultsch, M., and Kossiakoff, A. A. (1992) *Science* **255,** 306–312
12. Cunningham, B. C., and Wells, J. A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3407–3411
13. Cunningham, B. C., and Wells, J. A. (1989) *Science* **244,** 1081–1085
14. Cunningham, B. C., and Wells, J. A. (1993) *J. Mol. Biol.* **234,** 554–563
15. Cunningham, B. C., Ultsch, M., De Vos, A. M., Mulkerrin, M. G., Clauser, K. R., and Wells, J. A. (1991) *Science* **254,** 821–825
16. Cunningham, B. C., Jhurani, P., Ng, P., and Wells, J. A. (1989) *Science* **243,** 1330–1336
17. Clackson, T., and Wells, J. A. (1995) *Science* **267,** 383–386
18. Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000) *Methods Enzymol.* **328,** 333–363
19. Kunkel, T. A., Bebenek, K., and McClary, J. (1991) *Methods Enzymol.* **204,** 125–139
20. Jin, L., Fendly, B. M., and Wells, J. A. (1992) *J. Mol. Biol.* **226,** 851–865
21. Bond, C. J., Wiesmann, C., Marsters, J. C., Jr., and Sidhu, S. S. (2005) *J. Mol.*

Biol. **348,** 699–709
22. Shannon, C. E. (1948) *Bell System Technical J.* **27,** 379–423, 623–656
23. Hill, M. O. (1973) *Ecology* **54,** 427–432
24. Carson, M. (1987) *J. Mol. Graph.* **5,** 103–106
25. Bernat, B., Pal, G., Sun, M., and Kossiakoff, A. A. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 952–957
26. Valdar, W. S. (2002) *Proteins* **48,** 227–241
27. Stewart, J. J., Lee, C. Y., Ibrahim, S., Watts, P., Shlomchik, M., Weigert, M., and Litwin, S. (1997) *Mol. Immunol.* **34,** 1067–1082
28. Shenkin, P. S., Erman, B., and Mastrandrea, L. D. (1991) *Proteins* **11,** 297–313
29. Schwehm, J. M., Kristyanne, E. S., Biggers, C. C., and Stites, W. E. (1998) *Biochemistry* **37,** 6939–6948
30. Zhang, Z., and Palzkill, T. (2004) *J. Biol. Chem.* **279,** 42860–42866
31. Zhang, Z., and Palzkill, T. (2003) *J. Biol. Chem.* **278,** 45706–45712
32. Yang, J., Swaminathan, C. P., Huang, Y., Guan, R., Cho, S., Kieke, M. C., Kranz, D. M., Mariuzza, R. A., and Sundberg, E. J. (2003) *J. Biol. Chem.* **278,** 50412–50421
33. Vajdos, F. F., Adams, C. W., Breece, T. N., Presta, L. G., de Vos, A. M., and Sidhu, S. S. (2002) *J. Mol. Biol.* **320,** 415–428
34. Thorn, K. S., and Bogan, A. A. (2001) *Bioinformatics* **17,** 284–285
35. Skelton, N. J., Koehler, M. F., Zobel, K., Wong, W. L., Yeh, S., Pisabarro, M. T., Yin, J. P., Lasky, L. A., and Sidhu, S. S. (2003) *J. Biol. Chem.* **278,** 7645–7654
36. DeLano, W. L. (2002) *Curr. Opin. Struct. Biol.* **12,** 14–20
37. Bernat, B., Sun, M., Dwyer, M., Feldkamp, M., and Kossiakoff, A. A. (2004) *Biochemistry* **43,** 6076–6084
38. Baker, B. M., and Murphy, K. P. (1997) *J. Mol. Biol.* **268,** 557–569