Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition

Frederic A. Fellouse*, Christian Wiesmann*, and Sachdev S. Sidhu*†

*Department of Protein Engineering, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080

Edited by James A. Wells, Sunesis Pharmaceuticals, Inc., South San Francisco, CA, and approved July 15, 2004 (received for review March 13, 2004)

Antigen-binding fragments (Fabs) with synthetic antigen-binding sites were isolated from phage-displayed libraries with restricted complementarity-determining region (CDR) diversity. Libraries were constructed such that solvent-accessible CDR positions were randomized with a degenerate codon that encoded for only four amino acids (tyrosine, alanine, aspartate, and serine). Nonetheless, high-affinity Fabs (K_d = 2–10 nM) were isolated against human vascular endothelial growth factor (hVEGF), and the crystal structures were determined for two distinct Fab-hVEGF complexes. The structures revealed that antigen recognition was mediated primarily by tyrosine side chains, which accounted for 71% of the Fab surface area that became buried upon binding to hVEGF. In contrast, aspartate residues within the CDRs were almost entirely excluded from the binding interface. Alanine and serine residues did not make many direct contacts with antigen, but they allowed for space and conformational flexibility and thus played an auxiliary role in facilitating productive contacts between tyrosine and antigen. Tyrosine side chains were capable of mediating most of the contacts necessary for high-affinity antigen recognition, and, thus, it seems likely that the overabundance of tyrosine in natural antigen-binding sites is a consequence of the side chain being particularly well suited for making productive contacts with antigen. The findings shed light on the basic principles governing the evolution of natural immune repertoires and should also aid the development of improved synthetic antibody libraries.

The antigen-binding fragment (Fab) contains six hypervariable complementarity-determining regions (CDRs) (1) that present a large contiguous surface for antigen recognition (2, 3). The immune system contains a highly diverse population of antibodies, and each is distinguished by a unique set of CDRs that confer antigen specificity (4). The database of natural antibody sequences has revealed that, whereas the compiled CDR sequences are highly diverse, there are clear biases for particular amino acids (5–11). Furthermore, the structural database reveals that these biases are even greater when one considers residues that mediate antigen recognition through direct contacts (12–14). In particular, tyrosine is highly abundant in antigen-binding sites because it accounts for ≈10% of the total CDR composition and for ≈25% of the antigen contacts (12).

The overabundance of tyrosine can be explained by at least two possible scenarios. Amino acid bias may be a coincidental consequence of biases that exist in the base composition of antibody genes (10, 15–18). Alternatively, and more intriguingly, it may be that tyrosine is particularly well suited for molecular recognition (10, 12, 19), and, thus, selective pressure exists for the enrichment of tyrosine in functional antigen-binding sites. Insights into this question would have profound implications for our understanding of how the immune repertoire has evolved and, also, would greatly facilitate efforts aimed at the *in vitro* evolution of synthetic antibodies (20–22). Unfortunately, it is not at all obvious how the basic principles underlying molecular recognition can be elucidated from the complexities of the natural immune system.

Herein, we used phage-displayed antibody libraries with precisely defined and highly restricted diversities to address directly the question of whether tyrosine is a favored amino acid for molecular recognition. Structural analysis of two synthetic antibodies bound to antigen revealed that tyrosine side chains dominate the antigen-binding sites. Our results demonstrate that the tyrosine side chain is well suited for mediating molecular recognition at protein–protein interfaces, and, as a consequence, the natural antibody repertoire has likely evolved under selective pressure for the enrichment of tyrosine in antigen-binding sites.

Materials and Methods

Construction of Libraries. For the construction of naive heavy-chain libraries, we modified a previously described phagemid designed to display bivalent Fab4D5 (Fab′-zip) on the surface of M13 bacteriophage. The gene coding for the Fab′-zip was fused to the C-terminal domain of the M13 gene-3 minor coat protein and expressed under the control of the phoA promoter (23). The phagemid was modified by a single mutation in the light chain (R66G) and by the introduction of TAA stop codons into all three heavy-chain CDRs. For each library construction, the resulting phagemid (pV-0116c) was used as the "stop template" in a mutagenesis reaction with oligonucleotides designed to repair simultaneously the stop codons and introduce designed mutations at the desired sites, as described (22, 24).

For the construction of each light-chain library for affinity maturation, a phagemid selected for the display of a heavy-chain sequence was modified by the introduction of TAA stop codons into all three light-chain CDRs. The resulting phagemid was used as the "stop template" in a mutagenesis reaction that repaired the stop codons and introduced desired mutations, as described above.

Library Sorting and Binding Assays. Phage from the naive heavy-chain libraries were cycled through rounds of binding selection with antigen immobilized on 96-well Maxisorp immunoplates (NUNC) as the capture target, as described (22, 24). Bound phage were eluted with 0.1 M HCl for 10 min, and the eluant was neutralized with 1.0-M Tris base. Phage were propogated in *Escherichia coli* XL1-blue (Stratagene) with the addition of M13-KO7 helper phage (New England Biolabs).

Phage from the light-chain libraries were incubated for 2 h at room temperature in PBS, 0.05% Tween 20 (Sigma), and 0.5% Superblock (Pierce) with 100 nM hVEGF biotinylated with Sulfo-NHS-LC-Biotin reagent (Pierce). Biotinylated human vascular endothelial growth factor (hVEGF) and bound phage were captured for 5 min with neutravidin (Pierce) immobilized on Maxisorp immunoplates. The plates were washed with PBS and 0.05% Tween 20, and the bound phage were eluted and propagated for additional rounds of selection, as described above.

After selection, individual clones were grown in a 96-well

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CDR, complementarity-determining region; Fab, antigen-binding fragment; Flt-1_{D2}, domain 2 of Flt-1; hVEGF, human vascular endothelial growth factor; mVEGF, murine vascular endothelial growth factor; rmsd, rms deviation.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 1TZH and 1TZI).

†To whom correspondence should be addressed. E-mail: sidhu@gene.com.

Table 1. Fabs isolated from libraries with tetranomial diversity

Library*	Diversity†	Unique binding clones‡			
		hVEGF	mIgG	hGHbp	hIGF-1
KMT	Y/A/D/S	11	1		1
WMT	Y/T/N/S			5	
YMT	Y/P/H/S	1			2
RSA	T/R/G/A		3		6
SRC	H/R/D/G	1	2		1
RST	G/A/T/S		2		
MRT	H/N/R/S				
RMC	T/N/D/A				
RMG	T/K/E/A				
RRC	N/S/D/G				
MKC	L/R/I/S				

*The library name corresponds to the degenerate codon used; equimolar DNA degeneracies are represented by IUB code (K = G/T, M = A/C, R = A/G, S = G/C, W = A/T, and Y = C/T).

†The amino acid diversity encoded by the degenerate codon is shown in the single-letter code.

‡The number of unique binding clones selected from each library against each antigen was determined by DNA sequencing. Specific binding clones were defined as those that exhibited at least 10-fold greater phage ELISA signals on plates coated with antigen compared with plates coated with BSA. The following antigens were tested: hVEGF, murine immunoglobulin G (mIgG), human growth hormone-binding protein (hGHbp), and human insulin-like growth factor (hIGF-1).

format in 500 μl of 2YT broth supplemented with carbenicillin and M13-KO7. The culture supernatants were used in phage ELISAs to detect positive clones that bound to antigen-coated plates but not to BSA-coated plates (22). Positive clones were subjected to DNA sequence analysis.

Table 2. Kinetic analysis of Fabs binding to immobilized VEGF

Fab	k_{on}, $s^{-1}\cdot M^{-1}$		k_{off}, s^{-1}		K_d, nM	
	hVEGF	mVEGF	hVEGF	mVEGF	hVEGF	mVEGF
YADS 1	3×10^5	ND*	5×10^{-4}	ND*	1.8 ± 0.3	>1,000
YADS 2	1×10^6	8×10^5	1×10^{-2}	4×10^{-3}	10 ± 2	5.0 ± 0.8
YADS 3	1×10^6	2×10^6	3×10^{-3}	5×10^{-3}	2.0 ± 0.4	3.7 ± 0.6

*Values could not be determined due to the low affinity of the interaction.

Protein Purification and Affinity Analysis. Fab proteins were purified from *E. coli* shake-flask cultures, as described (25). Affinities were determined by surface plasmon resonance by using a BIAcore-3000 (Biacore, Piscataway, NJ) with hVEGF immobilized on CM5 chips at ≈100 response units, as described (26). Serial dilutions of Fab proteins (3–500 nM) were injected, and binding responses on the hVEGF flow cell were corrected by subtraction of responses on a blank flow cell. For kinetic analysis, a 1:1 Languir model with separate fittings of k_{on} and k_{off} was used. The K_d values were estimated from the ratios of k_{on} and k_{off}.

Crystallization, Structure Determination, and Refinement. For large-scale preparation of Fab protein, whole-cell broth was obtained from a 10-liter *E. coli* fermentation. The cells were lysed with a Manton–Gaulin homogenizer. The suspension was centrifuged, the supernatant was loaded on a protein A-Sepharose column (Genentech), and the column was eluted with 0.1 M acetic acid.

A

Fab	CDR-H1						CDR-H2										CDR-H3											
	28	29	30	31	32	33	50	51	52	52a	53	54	55	56	57	58	94	95	96	97	98	99	100	100a			100b	
YADS1	D	I	Y	D	D	D	Y	I	A	P	S	Y	G	Y	T	D	R	S	S	D	A	S	Y	S	Y	S	A	
YADS2	A	I	Y	D	Y	D	D	I	A	P	Y	A	G	A	T	A	R	S	S	Y	A	Y	Y	A			A	
YADS3	S	I	S	D	Y	D	A	I	A	P	Y	S	G	Y	T	Y	R	S	S	Y	A	Y	Y	S			A	
4	D	I	Y	D	Y	A	S	I	A	P	Y	A	G	S	T	Y	R	D	S	Y	A	Y	Y	A			A	
5	S	I	A	Y	Y	D	S	I	A	P	Y	A	G	A	T	D	R	S	S	Y	A	Y	A	S			A	
6	A	I	A	Y	D	D	S	I	A	P	Y	A	G	A	T	A	R	A	A	A	Y	A	A	Y			A	
7	A	I	D	Y	Y	D	S	I	A	P	Y	A	G	S	T	Y	R	S	A	A	Y	Y	A	Y			A	
8	A	I	A	A	Y	D	S	I	S	P	Y	A	G	A	T	S	R	A	A	A	Y	Y	Y	D			A	
9	D	I	S	A	Y	A	Y	I	S	P	S	S	G	A	T	A	R	S	A	Y	Y	S	Y	A			A	
10	D	I	Y	D	Y	S	Y	I	S	P	Y	S	G	A	T	A	R	S	S	Y	A	Y	Y	D			A	
11	D	I	S	A	S	D	S	I	S	P	Y	A	G	Y	T	A	R	S	A	Y	Y	A	A	S			A	
12	D	I	A	Y	Y	D	S	I	A	P	Y	A	G	S	T	A	R	A	A	A	A	Y	A	Y			A	
13	S	I	D	Y	Y	A	S	I	A	P	Y	A	G	Y	T	Y	R	A	A	A	Y	Y	Y	D			A	
14	D	I	S	Y	Y	D	S	I	A	P	Y	A	G	A	T	Y	R	A	A	D	Y	A	Y	D			A	
15	A	I	A	A	Y	A	D	I	Y	P	A	A	G	D	T	A	R	S	Y	Y	Y	S	Y	S	S	Y	A	

B

Fab	CDR-L1					CDR-L2				CDR-L3						
	28	29	30	31	32	50	51	52	53	90	91	92	93	94	95	96
YADS1	A	S	Y	S	S	A	A	S	Y		S	S	A	S	P	A
YADS2	S	Y	A	Y	A	D	A	S	Y	Q	A	Y	S	S	P	D
YADS3	A	S	Y	Y	D	A	A	S	Y	Q	Y	Y	Y	A	P	A

Fig. 1. CDR sequences of anti-VEGF Fabs. Residues in gray were not randomized in the libraries. The numbering is shown for the parental Fab4D5 according to the nomenclature of Kabat *et al.* (1). (*A*) Heavy-chain sequences selected from a naive library in which solvent-accessible CDR residues were randomized with a KMT degenerate codon. In CDR-H3, loops of variable length were inserted between residues 94 and 100b. (*B*) Light-chain sequences of high-affinity Fabs obtained by recombining selected heavy chains with a light chain CDR library. Gln-90 in the light chain of YADS1 was deleted by spontaneous mutation.

Table 3. Data collection and refinement statistics for YADS1 and YADS2 hVEGF complexes

Unit cell	YADS 1	YADS 2
Space group	$P2_1$	$C222_1$
a, Å	83.3	96.5
b, Å	76.3	149.6
c, Å	112.5	117.4
β, °	105.8	
Diffraction data		
Resolution, Å	50–2.6 (2.7–2.6)*	50–2.8 (2.9–2.8)*
No. of reflections	156,868	111,731
No. of unique reflections	41,828	20,861
R_{merge}†	0.078 (0.423)*	0.076 (0.399)*
Completeness (%)	99.9 (99.1)*	97.3 (83.2)*
Refinement		
R_{work}‡, R_{free}‡	0.212, 0.271	0.218, 0.254
No. of non-H atoms	8,104	4,077
No. of waters	110	0
rmsd bond length, Å	0.011	0.011
rmsd angles, °	1.2	1.3

*Values for the outer resolution shell are given in parentheses.

†$R_{merge} = \Sigma_{hkl} (|I_{hkl} - |I_{hkl}|)/\Sigma_{hkl} \langle I_{hkl} \rangle$, where I_{hkl} is the intensity of reflection hkl, and $\langle I_{hkl} \rangle$ is the average intensity of multiple observations.

‡$R_{work} = \Sigma \mid F_o - F_c \mid / \Sigma F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is the R factor for a randomly selected 5% of reflections that were not used in the refinement.

The pH was adjusted to 4.0 with 1.0 M Tris (pH 8.0), and the eluant was loaded on an SP-Sepharose column (Pharmacia). The column was washed with equilibration buffer (20 mM Mes, pH 5.5), and Fab protein was eluted with an NaCl gradient in equilibration buffer.

The complex between each Fab and the receptor-binding fragment of hVEGF was formed and purified, as described (25). The complex (in PBS, 25 mM EDTA) was concentrated to an optical density of $A_{280} = 10$. Hanging-drop experiments were performed by using the vapor-diffusion method with 10-μl drops consisting of a 1:1 ratio of protein solution and reservoir solution. The reservoir solution for the YADS1 complex was 0.2-M ammonium sulfate, 25% polyethylene glycol (PEG) 3350 (wt/vol), 0.1 M Hepes (pH 7.5). The reservoir solution for the YADS2 complex was 1.0 M lithium chloride, 10% PEG 6000 (wt/vol), and 0.1 M Mes (pH 6.0). After 1–2 weeks at 19°C, plate- or spindle-shaped crystals grew for the YADS1 or YADS2 complex, respectively.

Crystals were incubated in reservoir solution supplemented with 25% glycerol before flash freezing. A data set was collected from a single frozen crystal at the beam line 5.0.2 of the Advanced Light Source (Berkeley) for YADS1 and at the beam line 9.2 of the Stanford Synchrotron Radiation Laboratory (Stanford University) for YADS2. The data were processed by using the programs DENZO and SCALEPACK (27). The structures were solved by molecular replacement by using the program AMORE (28) and the coordinates of a solved Fab–hVEGF complex (PDB entry 1BJ1). The structure was refined by using the programs REFMAC (28). The models were manually adjusted by using the program O (29).

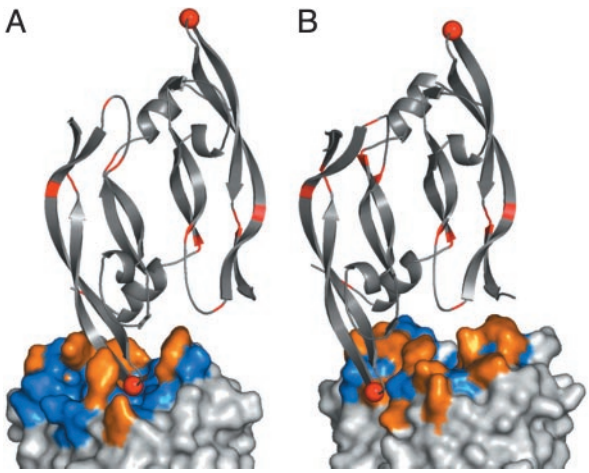

Fig. 2. The complex of hVEGF with YADS1 (*A*) and YADS2 (*B*). The main chain of the hVEGF homodimer is depicted as a gray ribbon, and residues that differ in comparison with mVEGF are colored red. Gly-88 is shown as a red sphere. The hVEGF homodimer binds to two symmetry-related Fabs, but only one Fab molecule is shown and is depicted as a molecular surface. Tyrosines at randomized positions are colored orange, whereas all other amino acid types at randomized positions are colored blue. Residues at positions that were not randomized in the libraries are colored white. These and other structural figures were derived from the crystal structure coordinates and were generated by using PYMOL (DeLano Scientific, San Carlos, CA).

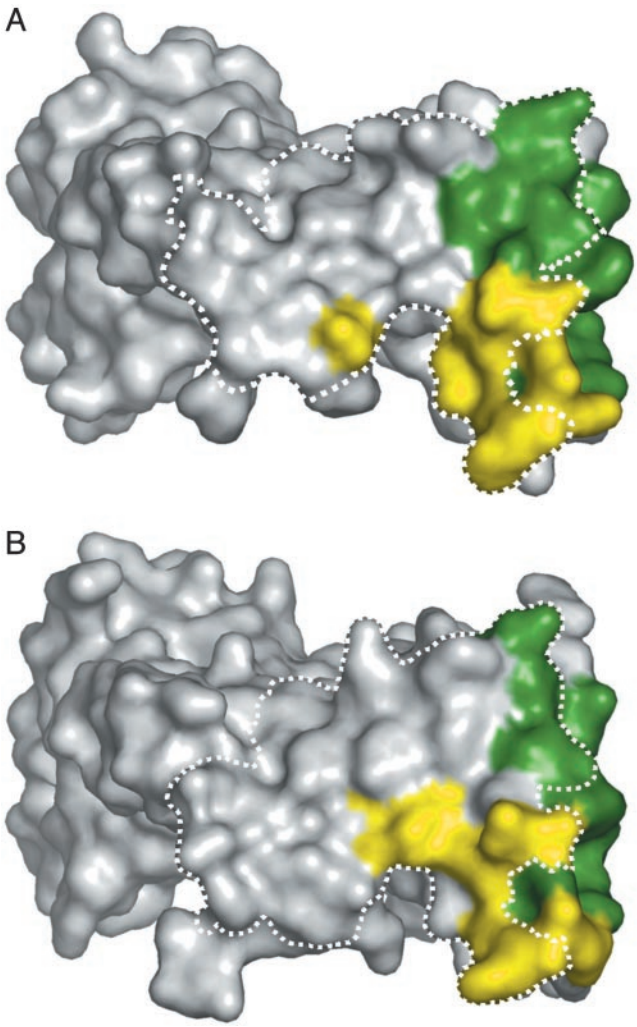

Fig. 3. The structural epitope for binding to YADS1 (*A*) or YADS2 (*B*) mapped on the molecular surface of hVEGF. The structural epitope consists of hVEGF residues that make contact with one or more residues of the Fab, with "contact" defined as a distance <4.1 Å. Residues that contact the heavy or light chain are colored green or yellow, respectively. The dashed line outlines the structural epitope for binding to Flt-1_{D2}, as determined from a previously described x-ray structure (PDB code 1FLT) (34).

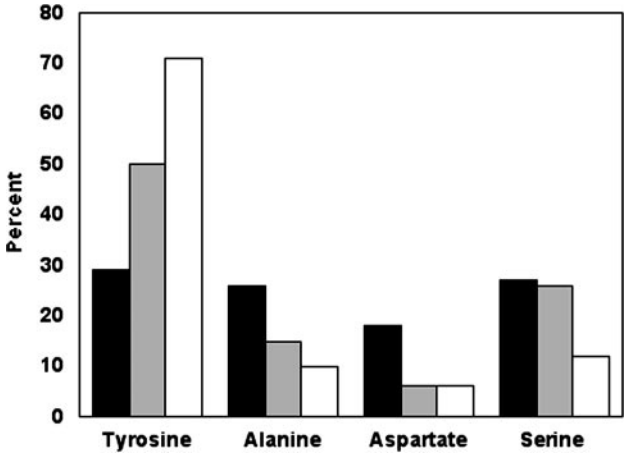

Fig. 4. Composition of the antigen-binding sites. The CDRs of YADS1 and YADS2 contained a total of 66 aa located at positions that were randomized in the libraries. Among these 66 residues, the graph shows the percentage (*y* axis) that each of the amino acid types allowed in the library (*x* axis) represented in terms of (*i*) the overall abundance (black bars), (*ii*) the residues that contact hVEGF as defined in Fig. 3 (gray bars), and (*iii*) the surface area buried upon binding to hVEGF (white bars).

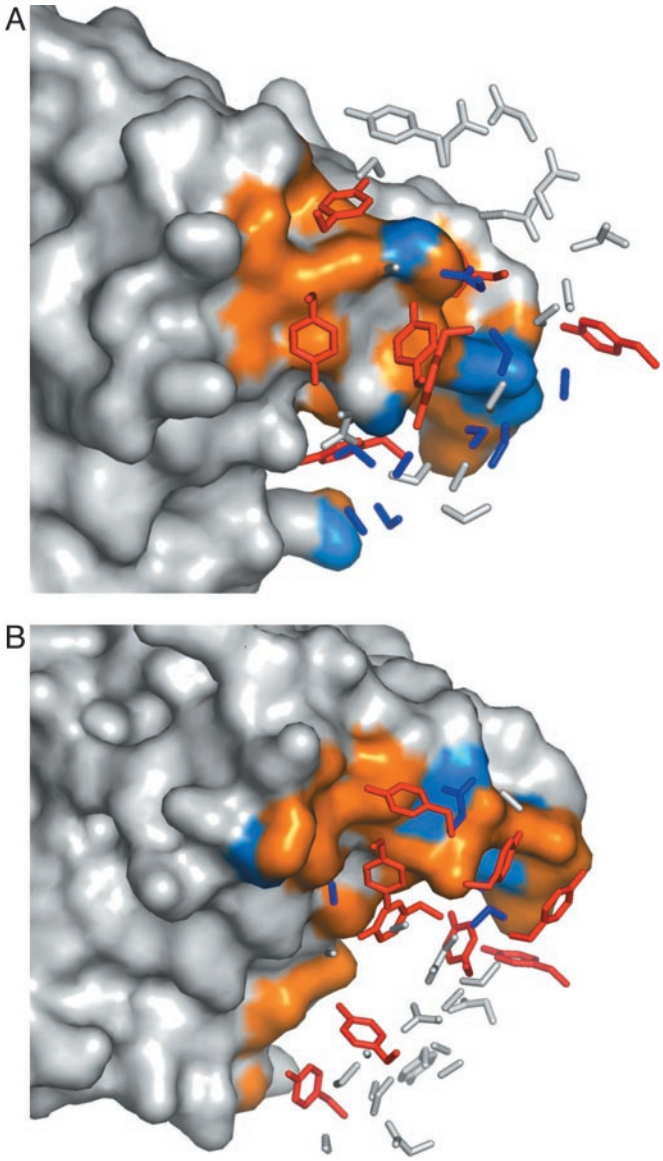

Fig. 5. The CDR side chains of YADS1 (*A*) and YADS2 (*B*) that contact hVEGF. The structural epitope for binding to the Fab (see Fig. 3) was mapped onto the molecular surface of hVEGF, and residues that made contacts with tyrosines or other residue types are colored orange or blue, respectively. The side chains at CDR positions that were randomized in the libraries are shown. Side chains that do not make contact with hVEGF are colored white. Tyrosine side chains that make contacts with hVEGF are colored red, whereas all other contacting side chains are colored blue. The hVEGF molecules are shown in the same orientation in both panels.

Results

Survey of Tetranomial Diversities. We first investigated whether small subsets of the natural amino acids could be used to generate antigen-binding surfaces. We constructed 11 phage-displayed libraries of antigen-binding fragments (Fabs) based on the humanized Fab4D5 (30), which recognizes the extracellular domain of the human receptor tyrosine kinase ErbB2 (31). In each library, solvent-accessible positions within the heavy-chain CDRs were replaced by a single type of degenerate codon that produced equal proportions of four amino acids (Table 1). The number of possible tetranomial combinations of the 20 natural amino acids is very large, and, thus, we chose combinations that fulfilled two criteria. First, we chose combinations that could be accessed with standard DNA synthesis methods. Second, we chose tetranomial sets that contained at least one small amino acid (glycine, serine, or alanine) because we reasoned that small residues would provide conformational flexibility and prevent steric crowding. A total of 18 positions were chosen for randomization: positions 28 and 30–33 in CDR-H1; positions 50, 52, 53, 54, 56, and 58 in CDR-H2; and positions 95–100a in CDR-H3. Each constructed library contained $\approx 10^{10}$ unique members, and, thus, the library diversities were only about one order of magnitude less than the maximum theoretical diversity ($4^{18} = 7 \times 10^{10}$).

Phage from the libraries were pooled together and cycled through rounds of binding selections against a panel of four protein antigens. After three rounds of selection, individual clones were assessed for antigen-specific binding with phage ELISAs (22). Approximately 100 clones were screened against each antigen, and specific binding clones were identified in each case. DNA sequencing revealed the number of unique clones isolated against each antigen (Fig. 7, which is published as supporting information on the PNAS web site), and also allowed us to determine the library of origin (Table 1). The 36 unique clones originated from 6 of the 11 libraries, and, notably, at least one tyrosine-containing library was successful against each antigen. In particular, Library-KMT was successful against three of the four antigens and generated 11 unique clones against human vascular endothelial growth factor (hVEGF).

Anti-VEGF Fabs Derived from a Naive Tetranomial Library. Having ascertained that Fab libraries with tetranomial diversity could generate specific binding clones, we next focused our attention on a detailed analysis of Library-KMT and hVEGF. We chose hVEGF because it is an angiogenic hormone of great biological importance (32, 33), and, also, the protein has proven very amenable to structural analyses (25, 34). We constructed new versions of Library-KMT in which the CDR-H1 and CDR-H2 diversities were the same as described above, but the diversity of CDR-H3 was increased by allowing for all possible length variations ranging from 3 to 15 residues inserted between residues 94 and 100b. All together, the pooled libraries contained a diversity of $\approx 10^{10}$ unique members that were cycled through selections for binding to hVEGF. Phage ELISA screens identified 93 hVEGF binders, and DNA sequencing revealed 15

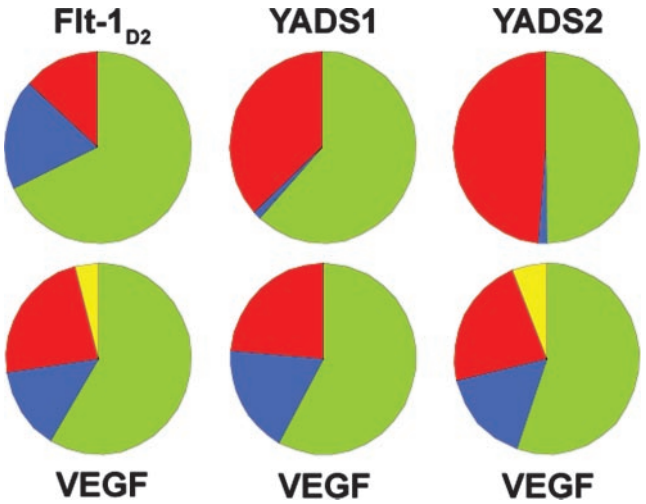

Fig. 6. The atomic composition of hVEGF–ligand interfaces. Each pair of circles represents the surface area buried upon complexation of ligand (*Upper*) with hVEGF (*Lower*). The colors indicate the proportion of the buried surface area composed of carbon (green), oxygen (red), nitrogen (blue), or sulfur (yellow).

unique sequences (Fig. 1*A*). Most of the clones contained CDR-H3 sequences with seven inserted residues, but we also identified two clones that contained longer insertions. The unique clones were subjected to competitive phage ELISAs (22) and exhibited estimated affinities in the 10-μM range (data not shown).

Affinity Maturation of Anti-VEGF Fabs. We next investigated whether the low-affinity anti-VEGF clones could be affinity matured to obtain Fabs with affinities comparable to those of natural antibodies. To this end, we recombined the 15 heavy chains (Fig. 1*A*) with a light-chain library in which 12 solvent-accessible positions were replaced with the same tetranomial KMT codon. Specifically, the following light-chain positions were randomized: positions 28–32 in CDR-L1, positions 50 and 53 in CDR-L2, and positions 91–94 and 96 in CDR-L3. The libraries contained $\approx 10^{10}$ unique members, which greatly exceeded the theoretical diversity of possible light chains ($4^{12} = 2 \times 10^7$). hVEGF in solution was used for a high-stringency selection. We sequenced 256 clones and identified 64 unique light chains combined with 9 of the 15 heavy chains (top 9 sequences in Fig. 1*A*). Competitive phage ELISAs (22) were used to estimate affinities, and the three best clones (YADS1, -2, and -3 in Fig. 1) were purified as free Fab proteins for detailed analysis.

Surface plasmon resonance was used to study the binding kinetics of the purified Fabs (Table 2). All three Fabs bound with high affinity to hVEGF, but only two exhibited appreciable affinity for the highly homologous murine VEGF (mVEGF, 90% amino acid identity). We reasoned that YADS2 and YADS3 recognized VEGF through a very similar mechanism because they exhibited high sequence homology in their CDRs and bound to both human and murine VEGF. In contrast, YADS1 likely represented a unique mode of antigen recognition because it contained very different CDR sequences and did not recognize mVEGF.

Crystal Structures of Fab–hVEGF Complexes. We wanted to study the structural basis for antigen recognition, and, therefore, the crystal structures of YADS1 and YADS2 in complex with hVEGF were solved and refined at 2.65- and 2.8-Å resolution, respectively (Table 3). The Fab frameworks were essentially unchanged in comparison with the structure of the parental Fab4D5; the C_α atoms of the YADS1 and YADS2 frameworks superimposed with Fab4D5 with rms deviations (rmsd) of 0.87 and 0.55 Å, respectively. The C_α atoms of the hVEGF molecules in the two structures superimpose well onto each other, with rmsd of 0.7 Å for 87 C_α positions. The largest deviation of 3.7 Å occurs at residue glutamic acid 64. The loop containing this residue has inherent flexibility, as shown by Muller *et al.* (25).

In both complexes, antigen recognition was entirely mediated by contacts with the CDR loops (Fig. 2). In terms of buried surface area, YADS1 used both the heavy (498 Å^2) and light (407 Å^2) chain, whereas YADS2 used mostly the heavy chain (543 Å^2) and a small contribution from the light chain (157 Å^2). Notably, residues at randomized positions accounted for essentially all of the buried surface area (98% and 100% for YADS1 and YADS2, respectively), and furthermore, the buried surface area involved almost entirely side-chain atoms (82% and 80% for YADS1 and YADS2, respectively). Thus, both Fabs bound to antigen through interactions that were almost entirely mediated by side chains located at positions that were randomized in the libraries.

On the hVEGF side, the structural epitopes for binding to YADS1 and YADS2 overlap with each other, and also, with the structural epitope for binding to domain 2 of the hVEGF receptor Flt-1 (Flt-1$_{D2}$, Fig. 3). Nonetheless, there are significant differences between the structural epitopes for the two Fabs (Fig. 2). In particular, of the 8 residues that differ between human and murine VEGF, only residue 88 is in contact with the Fabs, but the interactions involving this residue explain the differing affinities of YADS1 and YADS2 for mVEGF. In the YADS2 complex, Gly-88 is partially exposed to solvent, whereas, in the YADS1 complex, it is completely buried in the interface. Murine VEGF contains a larger serine residue at position 88. This substitution can be readily accommodated in the YADS2 complex, but, in the YADS1 model, the introduction of a serine side chain at the buried Gly-88 position would require major rearrangements for the complex to be preserved.

Composition of the Antigen-Binding Sites. As described above, both Fabs bind to antigen through contacts almost exclusively involving side chains at varied sites. In total, the CDRs of YADS1 and YADS2 contain 66 residues derived from randomized codons, and these residues are almost equally distributed among the four amino acid types allowed in the library design (Fig. 4). However, when we consider the subset of residues that make contact with antigen, there is a clear bias, in that 16 tyrosines account for 50% of the contact residues. Indeed, all but two of the tyrosines selected in the CDRs of YADS1 and YADS2 make contacts with antigen (Fig. 5), and, all told, tyrosines contribute 71% of the surface area buried upon complexation with hVEGF (Fig. 4). Thus, essentially every selected tyrosine side chain is involved in directly mediating antigen recognition, and the other selected amino acids apparently play auxiliary roles.

Despite the predominance of tyrosine in the synthetic antigen-binding sites, an examination of the heavy atom (non-hydrogen) content of buried surface areas reveals that the Fab–hVEGF interfaces are no more hydrophobic than the interface between hVEGF and Flt-1$_{D2}$ (Fig. 6). On the hVEGF side, the heavy atom composition of the buried surface area is very similar in all three cases, being composed predominantly of carbon but also containing significant proportions of nitrogen and oxygen. Within the buried surface areas of the Fabs, nitrogen atoms are almost entirely absent because the side chains allowed in the libraries were composed entirely of carbon, oxygen, and hydrogen. Nonetheless, both Fabs bury a large number of oxygen atoms upon binding to hVEGF, and, in both cases, the proportion of the buried surface area contributed by carbon is considerably less than that contributed by carbon to the buried surface area of Flt-1$_{D2}$ (Fig. 6). Thus, the predominance of tyrosine in

the synthetic CDRs does not produce highly hydrophobic Fab–antigen interfaces dominated by aromatic interactions. On the contrary, the tyrosine residues make specific contacts with a wide variety of residues on the hVEGF surface, and these interactions use both the side-chain hydroxyl groups and aromatic rings.

Discussion

We circumvented the complexity of the natural immune system by using precisely defined synthetic libraries, and, as a result, we were able to investigate the special role that tyrosine plays in antigen-binding sites. We generated libraries with restricted diversities and displayed the diverse surfaces on a fixed scaffold formed by the framework regions and buried CDR residues. Our results dramatically demonstrate that, in the context of a suitable scaffold, the tyrosine side chain is capable of mediating most of the contacts necessary for high-affinity antigen recognition. Thus, it seems very likely that the overabundance of tyrosine in natural antigen-binding sites is a consequence of the side chain being particularly well suited for making productive contacts with antigen.

This supposition is also consistent with the chemical nature of tyrosine. As noted previously, the tyrosine side chain is large enough to sweep out large volumes of space with only a few torsion angles, and it can form hydrogen bonds, hydrophobic interactions, and attractive electrostatic interactions with positively charged groups (10–12). In addition, the uncharged tyrosine side chain avoids electrostatic repulsion effects, and its midrange hydrophilicity allows it to adapt favorably to both hydrophilic and hydrophobic environments (10–12).

We also observed that, whereas alanine and serine residues did not make many direct contacts with antigen, they allowed for space and conformational flexibility that may be crucial for appropriate positioning of the large tyrosine side chains. Thus, these small residues may serve an auxiliary function in facilitating productive contacts between tyrosine and antigen. It is worth noting that, perhaps not coincidentally, serine is also highly abundant in natural antigen-binding sites (12). Finally, the paucity of antigen contacts mediated by aspartate suggests that it may be possible to further minimize the chemical diversity of these synthetic antigen-binding sites.

In summary, our results demonstrate that synthetic antibody libraries can be used to investigate the relative suitability of different amino acids for antigen recognition. These findings shed light on the basic principles governing the evolution of natural immune repertoires and also should aid the development of improved synthetic antibody libraries.

We thank the Genentech Fermentation and DNA Synthesis Groups, B. Currell for DNA sequencing, C. Eigenbrot and M. C. Franklin for assistance in x-ray data collection, and H. Darbon for his support. We also thank the staff at the Stanford Synchrotron Radiation Laboratory (SSRL), beam line 9.2, and Advanced Light Source (ALS), beam line 5.0.2. The SSRL is funded by the Department of Energy (Basic Energy Sciences and Biological and Environmental Research) and the National Institutes of Health (National Center for Research Resources and National Institute of General Medical Sciences). The ALS facility is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy.

1. Kabat, E. A., Wu, T. T., Redi-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (Natl. Inst. of Health, Bethesda).
2. Amit, A. G., Mariuzza, R. A., Phillips, S. E. & Poljak, R. J. (1986) *Science* **233,** 747–753.
3. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986) *Nature* **321,** 522–525.
4. Berek, C., Griffiths, G. M. & Milstein, C. (1985) *Nature* **316,** 412–418.
5. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1977) *J. Biol. Chem.* **252,** 6609–6616.
6. Padlan, E. A. (1990) *Proteins* **7,** 112–124.
7. Lea, S. & Stuart, D. (1995) *FASEB J.* **9,** 87–93.
8. Lo Conte, L., Chothia, C. & Janin, J. (1999) *J. Mol. Biol.* **285,** 2177–2198.
9. Collis, A. V., Brouwer, A. P. & Martin, A. C. (2003) *J. Mol. Biol.* **325,** 337–354.
10. Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J. A., Schroeder, H. W., Jr., & Kirkham, P. M. (2003) *J. Mol. Biol.* **334,** 733–749.
11. Ivanov, I., Link, J., Ippolito, G. C. & Schroeder, H. W., Jr. (2002) in *The Antibodies*, eds. Zanetti, M. & Capra, J. (Taylor & Francis, London), pp. 43–67.
12. Mian, I. S., Bradwell, A. R. & Olson, A. J. (1991) *J. Mol. Biol.* **217,** 133–151.
13. Padlan, E. A. (1994) *Mol. Immunol.* **31,** 169–217.
14. Davies, D. R. & Cohen, G. H. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7–12.
15. Tonegawa, S. (1983) *Nature* **302,** 575–581.
16. Padlan, E. A. (1997) *Mol. Immunol.* **34,** 765–770.
17. Wilson, P. C., Wilson, K., Liu, Y. J., Banchereau, J., Pascual, V. & Capra, J. D. (2000) *J. Exp. Med.* **191,** 1881–1894.
18. Bassing, C. H., Swat, W. & Alt, F. W. (2002) *Cell* **109,** Suppl., S45–S55.
19. Villar, H. O. & Kauvar, L. M. (1994) *FEBS Lett.* **349,** 125–130.
20. Barbas, C. F., 3rd, Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 4457–4461.
21. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A. & Virnekas, B. (2000) *J. Mol. Biol.* **296,** 57–86.
22. Sidhu, S. S., Li, B., Chen, Y., Fellouse, F. A., Eigenbrot, C. & Fuh, G. (2004) *J. Mol. Biol.* **338,** 299–310.
23. Lee, V., Sidhu, S. S. & Fuh, G. (2004) *J. Immunol. Methods* **284,** 119–132.
24. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. (2000) *Methods Enzymol.* **328,** 333–363.
25. Muller, Y. A., Chen, Y., Christinger, H. W., Li, B., Cunningham, B. C., Lowman, H. B. & de Vos, A. M. (1998) *Structure* **6,** 1153–1167.
26. Chen, Y., Wiesmann, C., Fuh, G., Li, B., Christinger, H. W., McKay, P., de Vos, A. M. & Lowman, H. B. (1999) *J. Mol. Biol.* **293,** 865–881.
27. Ötwinowski, Z. M., W. (1997) *Methods Enzymol.* **276,** 307–326.
28. Collaborative Computational Program Number 4 (CCP4) (1994) *Acta Crystallogr. D* **50,** 760–763.
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991) *Acta Crystallogr. A* **47,** 110–119.
30. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. (1993) *J. Mol. Biol.* **229,** 969–995.
31. Fendly, B. M., Winget, M., Hudziak, R. M., Lipari, M. T., Napier, M. A. & Ullrich, A. (1990) *Cancer Res.* **50,** 1550–1558.
32. Ferrara, N. (2001) *Am. J. Physiol.* **280,** C1358–C1366.
33. Folkman, J. (1995) *Nat. Med.* **1,** 27–31.
34. Wiesmann, C., Fuh, G., Christinger, H. W., Eigenbrot, C., Wells, J. A. & de Vos, A. M. (1997) *Cell* **91,** 695–704.