

then it is possible that the dynamin-like proteins have taken over the role of FtsZ in mitochondrial division in these organisms. However, MsFtsZ-mt seems to be able to affect mitochondrial morphology even in an organism such as yeast that normally relies on Dnm1 for organelle division. It seems likely that there will be major mechanistic differences in mitochondrial fission catalyzed by dynamins or FtsZs: one protein working from the outside of the organelle and the other from the inside.

We conclude that *MsFtsZ-mt* is likely to have been acquired from an endosymbiotic  $\alpha$ -proteobacterium that was the ancestor of the present-day mitochondrion. We suggest that in the course of evolution, the gene was transferred from mitochondrion to nucleus (16) and that the nuclear-encoded protein is now targeted back to the mitochondrion to play a role in the division of the organelle. This is the first identification of a eukaryotic *ftsZ* whose protein seems to be specifically targeted to the mitochondrion, and which may thus be related to the earliest mitochondrial division genes. The *Mallomonas* mitochondrial FtsZ will be a useful key for identifying other components critical to moulding the shape and facilitating the intergenerational transmission of mitochondria.

*Note added in proof:* A possible mitochondrial FtsZ from a red alga has recently been reported. The predicted protein of *CmftsZ1* (GenBank accession number AB032071) is 53% identical to MsFtsZ-mt and clusters with MsFtsZ-mt and the proteobacteria in a phylogenetic analysis like that in Fig. 1.

References and Notes

1. J. Bereiter-Hahn and M. Vöth, *Microsc. Res. Tech.* **27**, 198 (1994); K. P. Gaffal, *Endocyt. Cell Res.* **4**, 41 (1987).
2. K. W. Osteryoung and E. Vierling, *Nature* **376**, 473 (1995); K. W. Osteryoung, K. D. Stokes, S. M. Ruthenford, A. L. Percival, W. Y. Lee, *Plant Cell* **10**, 1991 (1998); R. Strepp, S. Scholz, S. Kruse, V. Speth, R. Reski, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4368 (1998).
3. E. Bi and J. Lutkenhaus, *Nature* **354**, 161 (1991); D. Bramhill, *Annu. Rev. Cell Dev. Biol.* **13**, 395 (1997); H. P. Erickson, *Trends Cell Biol.* **7**, 362 (1997); J. Lutkenhaus and S. G. Addinall, *Annu. Rev. Biochem.* **66**, 93 (1997); J. Löwe and L. A. Amos, *Nature* **391**, 203 (1998); W. Margolin, *Trends Microbiol.* **6**, 233 (1998).
4. Polyadenylated RNA was extracted with phenol and chloroform from an asynchronously dividing culture of *M. splendens* [P. L. Beech and R. Wetherbee, *J. Phycol.* **26**, 90 (1990)] and used to construct a cDNA library with the help of a ZAP-cDNA Gigapack III Gold cloning kit (Stratagene).
5. Nucleotides 60 through 1265 of *AtFtsZ1-1* from *A. thaliana* (GenBank accession number U39877) were amplified by reverse transcription polymerase chain reaction from RNA isolated from leaves of *A. thaliana* (Columbia).
6. This second probe was a 535-base pair (bp) Nsi I-Mlu I fragment of the *ftsZ1* gene from *S. meliloti* (Genbank accession number M94386) [W. Margolin, J. Corbo, S. R. Long, *J. Bacteriol.* **173**, 5822 (1991)].
7. S. G. E. Andersson et al., *Nature* **396**, 133 (1998); M. W. Gray, G. Burger, B. F. Lang, *Science* **283**, 1476 (1999).
8. The sequence of MsFtsZ-mt was e-mailed for analysis

- with the PSORT algorithm (<http://psort.ims.u-tokyo.ac.jp>). A mitochondrial targeting probability of 0.744 was predicted.
9. J. Nunnari et al., *Mol. Biol. Cell* **8**, 1233 (1997).
10. This behavior was seen in unpublished observations by P. L. Beech on live cells of *M. splendens* labeled with Mitotracker Green FM or CMXRos (Molecular Probes) according to the manufacturer's instructions.
11. An Eco RI-Spe I fragment (representing amino acids 8 through 342) of the *MsFtsZ-mt* cDNA was cloned into the ProEX HTb expression vector (Gibco). Six histidine fusion proteins were expressed in *Escherichia coli*, purified against Talon resin (Clontech), and electrophoretically separated before injection of gel slices into rabbits for antibody production. Antibodies likely to be specific to MsFtsZ-mt (anti-MsFtsZ-mt) were selected by affinity-purifying rabbit sera against the first 76 amino acids of MsFtsZ-mt [the corresponding cDNA fragment from *MsFtsZ-mt* was cloned into pGEX-4T-1 (Pharmacia) and expressed in *E. coli* as a glutathione S-transferase fusion protein] immobilized on nitrocellulose.
12. Q. Sun and W. Margolin, *J. Bacteriol.* **180**, 2050 (1998).
13. W. Bleazard et al., *Nature Cell Biol.* **1**, 298 (1999); H. Sesaki and R. E. Jensen, *J. Cell Biol.* **147**, 699 (1999); A. M. Labrousse, M. D. Zappaterra, D. A. Rube, A. M. van der Blik, *Mol. Cell* **4**, 815 (1999).
14. E. Smirnova, D.-L. Shurland, S. N. Ryazantsev, A. M. van der Blik, *J. Cell Biol.* **143**, 351 (1998); K. A. Shepard and M. P. Yaffe, *J. Cell Biol.* **144**, 711 (1999); M. P. Yaffe, *Science* **283**, 1493 (1999).
15. Nucleotides -34 through 1157 of the 1203-bp *MsFtsZ-mt* cDNA open reading frame were inserted upstream of the GFP coding region on the centromeric plasmid p416MET25 that carried the GFP (Ser<sup>55</sup> → Thr<sup>65</sup>) isoform. *S. cerevisiae* cells were trans-

- formed with the MsFtsZ-mt/GFP fusion as described [R. George et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2296 (1998)]. Mitochondria were isolated and sub-fractionated as previously described [B. S. Glick et al., *Cell* **69**, 809 (1992)].
16. J. D. Palmer, *Science* **275**, 790 (1997).
17. Cells were examined on a Leica TCS 4D laser scanning confocal microscope under a  $\times 100$  1.3 numerical aperture objective. Signals from green (GFP or fluorescein isothiocyanate) and red (Mitotracker) channels were collected simultaneously and merged with Adobe Photoshop.
18. For immunofluorescence, a nonsynchronously dividing culture of *M. splendens* was labeled with 100 nM MitoTracker CMX-Ros (Molecular Probes) for 2 hours; cells were washed in growth medium, then fixed in 2% paraformaldehyde in methanol at  $-20^{\circ}\text{C}$  for 10 min, washed in phosphate-buffered saline (PBS), and labeled with primary antibody diluted in blocking buffer (PBS, 0.05% Tween-20, and 1% bovine serum albumin) overnight at  $4^{\circ}\text{C}$ . Primary antibodies were detected with FITC-conjugated goat anti-rabbit IgG (Selenius, Hawthorn, Victoria, Australia), and cells were observed as described (17). Seemingly identical labeling with anti-MsFtsZ-mt was achieved both with and without MitoTracker labeling. Controls for antibody labeling were performed with preimmune sera.
19. We thank W. Margolin for the *S. meliloti* FtsZ probe, C. Cobbett for *Arabidopsis* RNA, J. Pickett-Heaps and R. Wetherbee for materials and encouragement, and T. Spurck for help with confocal microscopy. P.L.B. is funded by the Australian Research Council and is a Queen Elizabeth II fellow.

9 November 1999; accepted 11 January 2000

## Convergent Solutions to Binding at a Protein-Protein Interface

Warren L. DeLano,<sup>1</sup> Mark H. Ultsch,<sup>2</sup> Abraham M. de Vos,<sup>2</sup> James A. Wells<sup>1\*</sup>

The hinge region on the Fc fragment of human immunoglobulin G interacts with at least four different natural protein scaffolds that bind at a common site between the C<sub>H2</sub> and C<sub>H3</sub> domains. This "consensus" site was also dominant for binding of random peptides selected in vitro for high affinity (dissociation constant, about 25 nanomolar) by bacteriophage display. Thus, this site appears to be preferred owing to its intrinsic physicochemical properties, and not for biological function alone. A 2.7 angstrom crystal structure of a selected 13-amino acid peptide in complex with Fc demonstrated that the peptide adopts a compact structure radically different from that of the other Fc binding proteins. Nevertheless, the specific Fc binding interactions of the peptide strongly mimic those of the other proteins. Juxtaposition of the available Fc-complex crystal structures showed that the convergent binding surface is highly accessible, adaptive, and hydrophobic and contains relatively few sites for polar interactions. These are all properties that may promote cross-reactive binding, which is common to protein-protein interactions and especially hormone-receptor complexes.

Protein-protein interactions are central to the control of many biological functions, but we do not yet understand what general features

(if any) of a protein surface are most important for binding. Nature often provides convergent solutions to biological problems, and study of this "consensus" information can provide insight into the essential requirements for function. There are many examples of hormones that bind multiple receptors, or receptors that bind multiple hormones (1). Structural analysis of these complexes shows that the common protein will use virtually the same set of contact residues for binding many

<sup>1</sup>Graduate Group in Biophysics, University of California, San Francisco, CA 94143, USA and Sunesis Pharmaceuticals, 3696 Haven Avenue, Suite C, Redwood City, CA 94063, USA. <sup>2</sup>Department of Protein Engineering, Genentech, One DNA Way, South San Francisco, CA 94080, USA.

\*To whom correspondence should be addressed. E-mail: jaw@sunesis-pharma.com

## REPORTS

partners (2). One of the most striking and well-characterized examples of this principle is a consensus binding site found on the constant fragment (Fc) of immunoglobulin G (IgG), which interacts with four different proteins, each having radically different folds (Fig. 1) (3). We wondered if this common site was selected because its location is optimal for biological function or because the intrinsic physical properties of the site make it optimal for binding many different proteins.

To address this question, we performed an *in vitro* selection (4) to isolate peptides that bound Fc without the constraint that the peptides function *in vivo*. A library of cyclic peptides was constructed that consisted of  $4 \times 10^9$  different peptides of the form  $X_iCX_jCX_k$  (where C is cysteine, X is a random amino acid, and  $i + j + k = 18$ ) (5). Peptides from this library were expressed polyvalently on the surface of M13 bacteriophage as  $\text{NH}_2$ -terminal fusions to the gene VIII protein and selected for binding to immobilized Fc (6).

In principle, peptides could have been selected to bind to potentially any region of the Fc because of the unbiased nature of the library. However, after several rounds of selection, the library became dominated by a single peptide, Fc-I (ETQRCTWHMGELVWCEREHN) (7). Repetition of the selection experiment again gave Fc-I and also a related peptide, Fc-II (KEASCSYWLGLVWCVA-GVE). The Fc-II peptide shared the cysteine spacing and the internal GELVW sequence seen in Fc-I. Apparently, these two peptides bound Fc with an affinity high enough to be selected over any other Fc binding peptides present in the starting pool. Both peptides were synthesized and found to compete with Protein A (Z-domain) (8) for binding to Fc with inhibition constants ( $K_i$ 's) of about 5

$\mu\text{M}$  (9, 10), implying that these peptides bind to an overlapping site on Fc coinciding with the Protein A binding site.

The gene sequence for Fc-II was transferred to gene III of M13 bacteriophage (11) and improved by monovalent phage display. Five residue blocks were randomly mutated in six separate libraries to exhaustively cover the non-cysteine positions in the peptide sequence (12). Preferred residues from selection of these libraries for binding to Fc were then recombined to give three more libraries spanning the peptide sequence (13). Selection patterns from these libraries suggested a 13-residue core Fc binding sequence (DCAWHLGELVWCT). The corresponding peptide (Fc-III) was synthesized and found to inhibit binding of Protein A (Z-domain) to Fc with a  $K_i$  of 25 nM (14). Thus, although Fc-III is seven residues shorter than Fc-II, it binds 200 times more tightly. Despite its smaller size, the binding affinity of Fc-III to Fc was only about twofold weaker than that of the domains from Protein A and Protein G, which are each about four times larger and bind with  $K_d$ 's of around 10 nM (15).

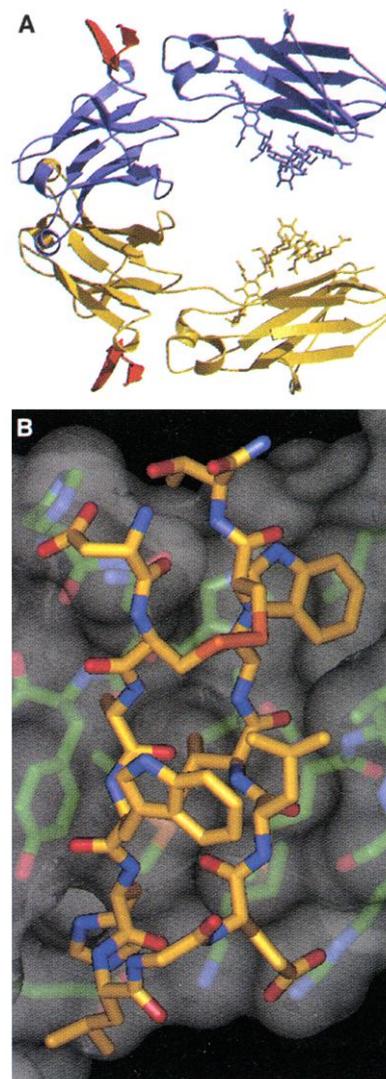
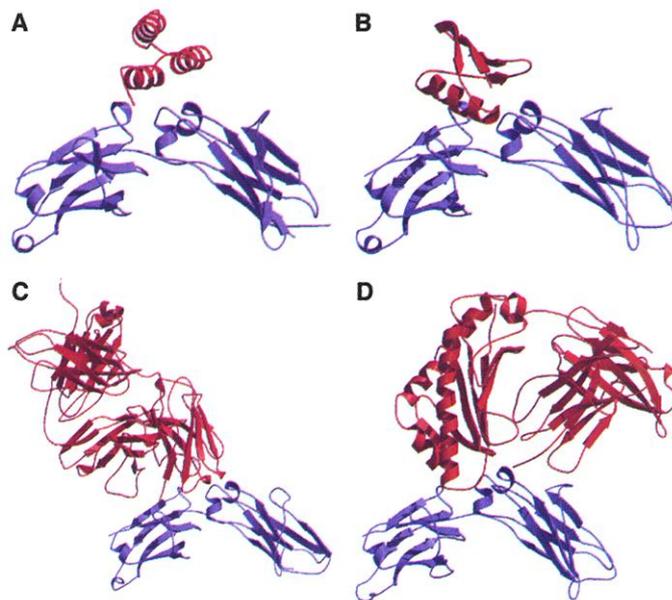
To understand whether binding of Fc-III occurs through interactions similar to those of the natural Fc-binding domains, we determined the x-ray crystal structure of the Fc-III peptide in complex with Fc at 2.7 Å resolution (Table 1) (16). We found that the peptide adopts a  $\beta$ -hairpin conformation unrelated to any known Fc binding motif (Fig. 2). A  $\beta$ -bulge conformation at Leu-9 in Fc-III accommodates the noneven number of residues separating the two cysteines, and 8 of the 12 amino acid side chains make extensive interactions with Fc. The remaining side chains (the disulfide bridged cysteines, a tryptophan, and a leucine) create a small hydrophobic core on the back side of the peptide that is facilitated by the  $\beta$  bulge and type II  $\beta$  turn.

Although the resolution of the structure is not sufficient to conclusively identify hydrogen bonding interactions, it appears that eight hydrogen bonds are formed between the peptide and Fc, including those involved in intermolecular salt bridges.

The Fc-III peptide targets the consensus binding site of the natural Fc binding proteins (Fig. 3). Although much smaller in size, the Fc-III peptide covers almost as much area ( $650 \text{ \AA}^2$ ) on the surface of Fc as do the fourfold larger IgG-Fc binding domains from Protein A and Protein G and the 15-fold larger rheumatoid factor (each cover about  $740 \text{ \AA}^2$ ) (17).

Despite the lack of structural similarity between the peptide and the natural Fc binding proteins, detailed inspection of the interactions formed by all of these molecules with

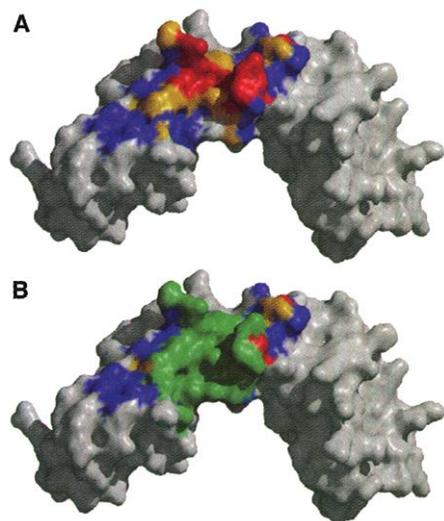
**Fig. 1.** Ribbon diagrams of an IgG-Fc subunit (blue) in complex with (A) domain B1 of Protein A, (B) domain C2 of Protein G, (C) rheumatoid factor, and (D) neonatal Fc-receptor (3). All four proteins bind to an overlapping region at the  $\text{C}_{\text{H}2}/\text{C}_{\text{H}3}$  domain interface.



**Fig. 2.** Crystal structure of Fc-III (DCAWHLGELVWCT-NH<sub>2</sub>), in complex with IgG-Fc. (A) Ribbon diagrams of two Fc-III peptides in complex with the Fc dimer; (B) close-up view of the peptide interacting with the surface of IgG-Fc. This structure has been deposited in the PDB under accession number 1DN2.

## REPORTS

Fc reveals a number of common features in the consensus binding region (Fig. 4). For example, the phenyl ring of Phe-14 (18) in Protein A (Fig. 4E, 3) occupies the same



**Fig. 3.** Molecular surface representation of the consensus binding site on IgG-Fc [coordinates from Deisenhofer (3)]. (A) Fc with superimposed binding interfaces of Protein A, Protein G, and rheumatoid factor. (The 4.5 Å crystal structure of the Fc-receptor/Fc complex was excluded because of its low resolution.) Atoms are colored blue, yellow, or red depending on whether they are involved in one, two, or three of the interfaces, respectively (17). (B) As in (A) with the interface of the Fc binding peptide (Fc-III) superimposed in green. Fc-III interacts with many of the atoms that are found in the interfaces of the other three Fc binding proteins.

position as the indole ring of Trp-11 of Fc-III (Fig. 4B, 3). The indole nitrogen of Trp-43 in Protein G (Fig. 4C, 2) makes the same hydrogen bond with Asn-433 on Fc as does the main-chain amide from Thr-13 of Fc-III (Fig. 4B, 2). Tyr-98H and Asp-31H (Fig. 4D, 6 and 9) from the rheumatoid factor heavy chain make identical hydrophobic and polar interactions, respectively, as do Val-10 and Glu-8 on Fc-III (Fig. 4B, 6 and 9), and Lys-28 from Protein G (Fig. 4C, 8) makes the same salt-bridge that His-5 does from the Fc-III (Fig. 4B, 8). In a striking example of repeated convergent evolution at the atomic level, the backbone amide of Val-10 in Fc-III (Fig. 4B, 5), Glu-27 in Protein G (Fig. 4C, 5), the backbone amide of Tyr-98H in rheumatoid factor (Fig. 4D, 5), and Gln-11 in Protein A (Fig. 4E, 5) all make the same buried hydrogen bond with the backbone amide proton of Ile-253 on Fc.

With its slightly smaller contact surface, the peptide mimics two to six polar interactions and many analogous nonpolar contacts formed by each of the other Fc binding domains. Charge-charge interactions are distributed at the top and bottom of the binding site, hydrophobic contacts line both sides, and a strip of hydrogen bonding interactions runs directly through the center (Fig. 4). Interactions that are present in all of the binding interfaces are mediated by a shared set of contacts with atoms on six amino acid side chains—Met-252, Ile-253, Ser-254, Asn-434, His-435, and Tyr-436—as well as shared contacts with adjacent atoms on the peptide backbone. These atoms form a contiguous

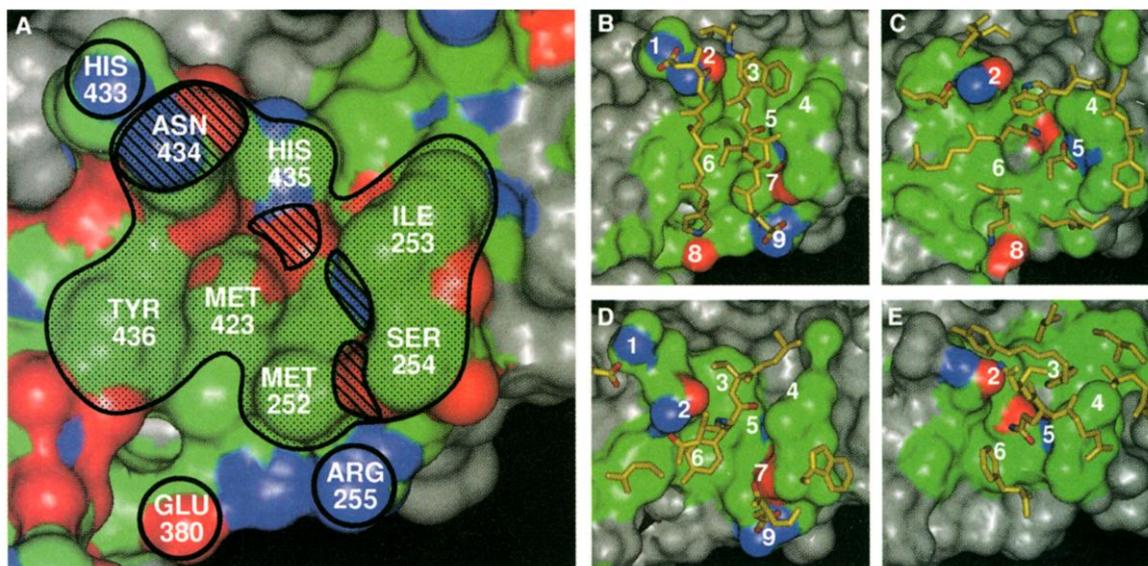
525 Å<sup>2</sup> patch of solvent-accessible surface area on Fc.

To assess whether the consensus region on the surface of Fc has unusual properties, we generated a comparison set of several million random surface patches of similar

**Table 1.** Data collection and refinement statistics.

Data collection	
Source	SSRL, Beam 7.1, λ = 0.908
Space group	P2 <sub>1</sub>
Unit cell (Å, °)	a = 67.54, b = 60.83, c = 68.17, β = 103.87
Data processing	
Resolution (Å)	25 to 2.7
R <sub>merge</sub> (%)	8.3
Completeness (%)	99.9
Total reflections	60,434
Unique reflections	14,847
Redundancy	4.1
Average I/σ <sub>i</sub>	11.3
Model refinement	
Resolution (Å)	20 to 2.7
Unique reflections (F > 0.1σ)	14,266
Bulk solvent correction (e, Å <sup>2</sup> )	0.26, 10.0
rmsd in bonds (Å)	0.007
rmsd in angles (°)	1.64
Average B factor (Å <sup>2</sup> )	26.6
rmsd in B factor for bonded atoms (Å <sup>2</sup> )	2.1
R (%)	19.4
R <sub>free</sub> (%)	25.2
Number of atoms	4,863
Number of water molecules	55

**Fig. 4.** Topology of the consensus binding site on Fc. (A) Conserved interaction sites. The predominantly hydrophobic consensus region is shaded. Hydrogen bonding sites are shown with diagonal lines and salt bridging locations are denoted by open circles. Nitrogen and oxygen are colored blue or red, respectively, and carbon and sulfur atoms are colored green. Hydrogens are not shown. (B to E) Comparison of the Fc binding interactions of (B) the selected peptide, Fc-III (DCAWHLGELVWCT-NH<sub>2</sub>), (C) domain C2 from Protein G, (D) rheumatoid factor, and (E) domain B1 of Protein A. Numbers indicate the following conserved interactions: (1) salt-bridges with His-433, (2) hydrogen bonding to Asn-434, (3) hydrophobic packing onto His-435, (4) burial of the hydrophobic "knob" formed by Ile-253 and Ser-254, (5) hydrogen bonding to main chain (N-H) of Ile-253, (6) hydrophobic packing onto Met-252 and Tyr-436, (7) hydrogen bonding to Ser-254, (8) salt-bridges with Glu-380,



and (9) salt-bridges with Arg-255. For clarity, only interfacial atoms are shown, and only nitrogen and oxygen atoms involved in conserved polar interactions are colored blue or red, respectively. The remaining contact atoms are colored yellow and green. The dynamic adaptability of this site can be viewed in a movie on Science Online at [www.sciencemag.org/feature/data/1044724.shl](http://www.sciencemag.org/feature/data/1044724.shl)

and (9) salt-bridges with Arg-255. For clarity, only interfacial atoms are shown, and only nitrogen and oxygen atoms involved in conserved polar interactions are colored blue or red, respectively. The remaining contact atoms are colored yellow and green. The dynamic adaptability of this site can be viewed in a movie on Science Online at [www.sciencemag.org/feature/data/1044724.shl](http://www.sciencemag.org/feature/data/1044724.shl)

size on five Fc crystal structures (3, 19). These hypothetical binding sites were then evaluated according to multiple criteria and compared with the consensus site (Fig. 5, A and B). The consensus binding region was distinguished by a high degree of solvent accessibility and a predominantly nonpolar character, suggesting that burial of exposed hydrophobic surface area is an important driving force behind binding at this site. This result is consistent with statistical surveys of distinct protein interactions, which suggest an important energetic role for hydrophobic burial in protein association (20). In addition, the low hydrogen bonding ability of the surface in this region (19% of the surface is capable of hydrogen bonding compared with 37% on average) indicates that this site places proportionately fewer specific geometric constraints on binding partners, because fewer complementary polar interactions are required for binding (Fig. 5B).

However, the consensus binding site is not the only exposed, less-polar region on the surface of the Fc dimer. We used patch analysis to search for regions with high solvent accessibility and below-average polarity (21) and found that the consensus binding site is part of a larger exposed hydrophobic region that extends about halfway across the C<sub>H2</sub> domain and includes residues near the 309-helix and the 280's loop (Fig. 5C). Residues on the tip of the C<sub>H2</sub> domain that would contact C<sub>H1</sub> in an intact IgG molecule were also identified. Given that other equally accessible and nonpolar sites exist on Fc, other properties, such as shape or adaptability, must also contribute to making the consen-

sus site the preferred locus for binding.

The consensus region on Fc undergoes a series of conformational changes in order to complement the distinct surface of each binding domain. Much of this adaptability is manifested in positional adjustments of Met-252 and its immediate neighbor Met-423. These residues adapt to form a pocket for Val-10 on Fc-III (Fig. 4B) and for Lys-31 on Protein G (Fig. 4C), or to present a much flatter surface for interaction with Tyr-98H in rheumatoid factor (Fig. 4D) and Phe-6 in Protein A (Fig. 4E). Similarly, Ile-253, His-433, and Asn-434 adopt different rotameric conformations depending on which protein is bound (Fig. 4, B to E).

Additional adaptability arises from changes in the relative orientation of the C<sub>H2</sub> and C<sub>H3</sub> domains. The location of the consensus binding site on Fc at a domain hinge is reminiscent of many cytokine receptor binding sites, which also show substantial structural adaptability and bind multiple ligands. The intrinsic adaptability of hinge region binding sites and the flexibility of the loop structures that typically constitute these sites may be important factors in facilitating promiscuity. Thus, the location of the consensus site at a domain hinge probably contributes to making this particular exposed nonpolar region the preferred locus for binding.

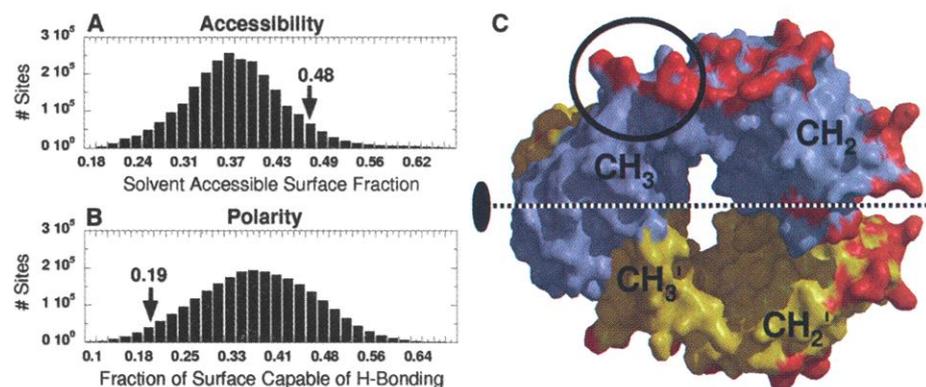
Mutagenesis experiments confirm an energetic role for the consensus patch in binding. Alanine replacement of Ile-253 or His-435 on Fc significantly disrupts binding ( $\Delta\Delta G > 2.5$  kcal/mol, where  $G$  is the Gibbs energy) of Protein A (Z-domain) (22, 23), and replacement of Asn-434, His-435, or Tyr-436 disrupts binding ( $\Delta\Delta G > 1.5$  kcal/

mol) of the selected peptide. On the complementary side of these interfaces, alanine substitutions of Gln-11, Phe-14, or Ile-32 on the Z-domain (23) and Val-10 or Trp-11 on the selected peptide (24) all result in  $\Delta\Delta G > 2.0$  kcal/mol reductions in affinity.

The consensus binding site on Fc is thus an adaptive, exposed, nonpolar, and energetically important region on the surface of Fc that is primed for interaction with a variety of distinct molecules. In recent years, numerous peptides selected to bind other protein receptors have been found to target and block the binding sites of natural ligands (25), suggesting that most receptors contain preferred binding sites on their surfaces. These sites have evolved to bind natural ligands but are competent for binding other ligands as well. Our results show that a peptide can target such a site by mimicking the specific interactions of the natural binding domains while presenting the interacting groups from an entirely different structural scaffold. Further investigation of the properties that characterize attractive sites on protein surfaces should improve our understanding of binding and assist in the development of therapeutic ligands.

References and Notes

1. J. A. Wells and A. M. de Vos, *Annu. Rev. Biochem.* **65**, 609 (1996).
2. B. C. Cunningham and J. A. Wells, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3407 (1991).
3. Fc/Protein A [Protein Data Bank (PDB) accession code 1FC2, 2.9 Å]; J. Deisenhofer, *Biochemistry* **20**, 2361 (1981); Fc/Protein G (1FC3, 3.5 Å): A. E. Sauer-Eriksson, G. J. Kelywegt, M. Uhlen, T. A. Jones, *Structure* **3**, 265 (1995); Fc/rheumatoid factor (1ADQ, 3.15 Å): A. L. Corper *et al.*, *Nature Struct. Biol.* **4**, 374 (1997); Fc/neonatal Fc receptor (1FRT, 4.5 Å): W. P. Burmeister, A. H. Huber, P. J. Bjorkman, *Nature* **372**, 379 (1994).
4. N. C. Wrighton *et al.*, *Science* **273**, 458 (1996); O. Livnah *et al.*, *Science* **273**, 464 (1996).
5. H. B. Lowman *et al.*, *Biochemistry* **37**, 8870 (1998). The library construct contained an STII secretion signal peptide, the peptide library of 20 amino acids (X<sub>i</sub>CX<sub>j</sub>CX<sub>k</sub>, where X is a random amino acid from an NNS codon,  $i + j + k = 18$ , and  $j = 4$  to 10), a GGSGGG linker, and the M13 gene VIII starting at the first residue of the mature protein.
6. Selections were done as described (5) with the following modifications: Microtiter wells were coated with IgG-Fc (5 μg/ml); Casein Blocker Buffer (Pierce) was used in place of 0.1% bovine serum albumin to better prevent nonspecific binding; elution of phage was effected with either 75 mM dithiothreitol or 0.2 mM glycine (pH 2.0) with equivalent results. IgG-Fc was obtained by papain cleavage of CD4-IgG, immunoadhesin protein [D. J. Capon *et al.*, *Nature* **337**, 525 (1989)]. Cleaved material was purified over Protein A-Sepharose followed by Superdex-75 (Pharmacia) and then quantified by absorbance at 280 nm.
7. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
8. B. Nilsson *et al.*, *Protein Eng.* **1**, 107 (1987).
9. Peptides were synthesized on solid phase with standard 9-fluorenylmethoxycarbonyl protocols and purified by reversed-phase chromatography. Masses were confirmed by electrospray mass spectrometry, and peptides were quantified by ultraviolet absorbance at 280 nm.
10. Competition binding assays were performed in a



**Fig. 5.** The consensus binding site compared with other IgG-Fc surface patches. Atoms in the consensus region are more accessible (A) and less polar (B) than atoms in most hypothetical binding sites of similar size. Arrows indicate the surface properties of the consensus region averaged over four complexes (excluding Fc-receptor) and in uncomplexed IgG-Fc (19). The comparison set consisted of 2.5 million 525 Å<sup>2</sup> surface patches of random globular shape distributed over the entire Fc dimer in five crystal structures. The solvent-accessible surface fraction in (A) is defined as the fraction of the maximum potential solvent-accessible surface area on protein atoms not buried by secondary or tertiary packing interactions. Because patches are of uniform size, those with high accessibility will have their accessible surface area concentrated onto fewer atoms than will patches with lower accessibility. In (B), atoms capable of acting as hydrogen bond donors or acceptors were classified as polar, and all other atoms were deemed nonpolar. (C) Fc dimer with symmetry related subunits shown in blue and yellow. Atoms found in patches with the highest average accessibility and lowest average polarity are colored red (21). The open circle identifies the location of the consensus binding site on one of the subunits. The consensus site on the other subunit is obscured on the back side of the dimer.

- manner similar to the method described in (5). Briefly, Z-domain was immobilized on microtiter wells at a concentration of 5 µg/ml, blocked, and washed as described. A matrix of mixtures of biotin-IgG-Fc (312 to 0.3 nM) and peptide (215 µM to 0.8 nM) were prepared. These mixtures were incubated with immobilized Z-domain for 1 hour. Plates were then washed and developed as described with avidin-horseradish peroxidase conjugate. Inhibition curves were then computed for each concentration of biotin-IgG-Fc, and the curve of half-maximal inhibition was extrapolated to zero biotin-IgG-Fc concentration to obtain a  $K_i$ .
11. The DNA sequence of the peptide was moved to a monovalent phage display format by cassette mutagenesis to give a construct with the STII signal sequence, the peptide KEASCSYWLGLVWCVAGVE, a GGGPGGG linker, and the M13 gene III protein starting at residue 253.
  12. A series of second-generation monovalent phage display libraries were constructed based on the sequence KEASCSYWLGLVWCVAGVE, in which five sequential residues were randomized by using NNS codons in each library starting at positions 1, 4, 7, 10, 12, and 16, excluding the two cysteines. Each library had a diversity of  $\sim 1 \times 10^8$ . These libraries were independently screened for binding to IgG-Fc for six rounds and then sequenced.
  13. Three additional libraries were constructed by using the degeneracy of the genetic code to recombine the preferred amino acids at each position into one peptide. The DNA sequences for these libraries contained the following mixtures of bases (IUPAC codes): DRG GWA GMA RRC TGC KCT TRS CAC MTG GGC GAG CTG GTC TGG TGC RVC RVM BKC GAS KDW, DRS VWG SVG RRC TGC KCC TRS YRS MTG GGC GAG CTG GTC TGG TGC RNC VVS NBS GWS KDM, and DNS NNS NNS VNS TGC BVG TDS HRS MDS GGC GAG STC KKG WRG TGC RNM NNS NNS NNS NNM. These libraries were also sorted against IgG-Fc for six rounds and then sequenced.
  14. Inhibition assays were performed as described (10) at pH 7.2 and at pH 6.0. The peptide was found to inhibit fourfold more tightly at the lower pH. Kinetic and steady-state binding to immobilized IgG<sub>1</sub> was also measured directly by BIAcore (Pharmacia), giving  $K_{on} = 1.6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ,  $K_{off} = 2.5 \times 10^{-2} \text{ s}^{-1}$ , and  $K_d = 16 \text{ nM}$  in 25 mM MES (pH 6.0), 0.05% Tween-20.
  15. S. R. Fahnestock *et al.*, in *Bacterial Immunoglobulin-Binding Proteins* (Academic Press, New York, 1990), vol. 1, chap. 11; R. Karlsson, L. Jendeborg, B. Nilsson, J. Nilsson, P. Nygren, *J. Immunol. Methods* **183**, 43 (1995).
  16. Crystals were grown in 100 mM NaOAc (pH 6.0), 20% polyethylene glycol 4000, and 20% isopropanol by vapor diffusion from 4-µl drops containing 100 µM IgG-Fc, up to 150 µM peptide, and a 50% dilution of reservoir solution. Data were collected to 2.6 Å at the Stanford Synchrotron Radiation Laboratory (SSRL) and were reduced with DENZO [W. Minor and Z. Otwinowski, *Methods Enzymol.* **176**, 307 (1997)]. Phasing was accomplished by molecular replacement with AmoRE [J. Navaza, *Acta Crystallogr.* **A50**, 157 (1994)], with an IgG-Fc subunit derived from Deisenhofer *et al.* (3) as a search model. The crystal contained one Fc dimer and two peptide molecules per asymmetric unit. The structure was refined with X-PLOR 3.1 [A. T. Brünger *et al.*, *Science* **245**, 458 (1987)], with noncrystallographic restraints on the Fc dimer over regions  $>10 \text{ Å}$  away from nonequivalent crystal contacts. The final dimeric Fc model consisted of IgG<sub>1</sub> residues 237 to 443 with eight sugars per monomer.
  17. Surface area and geometric measurements were made with the Crystallography and NMR System (CNS) [A. T. Brünger *et al.*, *Acta Crystallogr. D.* **54**, 905 (1998)]. A solvent probe radius of 1.4 Å was used, and surface area changes were computed by subtracting complexed from uncomplexed solvent-accessible surface areas. Contact regions were defined as the set of atoms that lie within 5.0 Å of any nonhydrogen atom on the opposing molecule.
  18. Protein A domain numbering is from H. Gouda *et al.*, *Biochemistry* **31**, 9665 (1992).
  19. The computer program SITEFINDER (WLD) was used to generate 2.5 million patches of contiguous surface atoms having solvent-accessible surface areas of 525 Å<sup>2</sup>. Patches were randomly distributed across all of the available structures (PDB codes: 1FC1, 1FC2, 1FCC, 1ADQ, and 1DN2) and were of a random globular shape. To ensure even sampling, probabilities were weighted so that each solvent-exposed atom was included in an equal number of surface patches ( $\sim 10,000$  patches per atom). The properties of each site were computed and then compared with those of the consensus binding patch on Fc.
  20. L. Young, R. L. Jernigan, D. G. Covell, *Protein Sci.* **3**, 717 (1994); S. Jones and J. M. Thornton, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13 (1996); C. Tsai, S. L. Lin, H. J. Wolfson, R. Nussinov, *Protein Sci.* **6**, 53 (1997); L. L. Conte, C. Chothia, J. Janin, *J. Mol. Biol.* **285**, 2177 (1999).
  21. Patches from (19) were ranked separately by polarity and solvent-accessible surface fraction. For each atom in the Fc dimer, the average rank of all patches involving the atom was then computed. The average atomic ranks for polarity and accessibility were then combined linearly ((accessibility) - (polarity)) to give a composite score incorporating both properties.
  22. C. Medesan, D. Matesoi, C. Radu, V. Ghetie, E. S. Ward, *J. Immunol.* **158**, 2211 (1997).
  23. W. L. DeLano, in preparation.
  24. Single alanine mutants of the peptide and Protein A Z-domain were displayed monovalently on gene III, assayed by enzyme-linked immunosorbent assay, and directly compared with unmutated control peptide as described [B. C. Cunningham, D. G. Lowe, B. Li, B. D. Bennett, J. A. Wells, *EMBO J.* **13**, 2508 (1994)]. Assays were performed under conditions where  $EC_{50}(\text{wt})/EC_{50}(\text{mut})$  will approximate  $K_d(\text{wt})/K_d(\text{mut})$  ( $EC_{50}$ , median effective concentration). Additional mutagenesis data are available in L. Jendeborg *et al.*, *J. Mol. Recog.* **8**, 270 (1995).
  25. B. K. Kay, A. V. Kurakin, R. Hyde-DeRuyscher, *Drug Discovery Today* **8**, 370 (1998).
  26. We thank B. C. Cunningham, J. K. Tong, and M. Dennis for assistance in the initial selection experiments against Fc; A. Braisted for training in solid phase peptide synthesis; C. Wiesmann for help with crystallographic refinement; and the SSRL for use of their facility in data collection.

19 August 1999; accepted 27 December 2000

## Evidence for a High Frequency of Simultaneous Double-Nucleotide Substitutions

Michalis Averof,<sup>1\*</sup> Antonis Rokas,<sup>2</sup> Kenneth H. Wolfe,<sup>3</sup> Paul M. Sharp<sup>4\*</sup>

Point mutations are generally assumed to involve changes of single nucleotides. Nevertheless, the nature and known mechanisms of mutation do not exclude the possibility that several adjacent nucleotides may change simultaneously in a single mutational event. Two independent approaches are used here to estimate the frequency of simultaneous double-nucleotide substitutions. The first examines switches between TCN and AGY (where N is any nucleotide and Y is a pyrimidine) codons encoding absolutely conserved serine residues in a number of proteins from diverse organisms. The second reveals double-nucleotide substitutions in primate noncoding sequences. These two complementary approaches provide similar high estimates for the rate of doublet substitutions, on the order of 0.1 per site per billion years.

Mutational events can be studied either by direct observation of mutations in the laboratory or by comparing sequences that have been accumulating mutations naturally, during evolution. Studies of the first kind have suggested that some mutations can involve multiple nucleotide changes (1, 2), and indeed, mechanisms that affect neighboring nucleotides are known. Examples include template-directed mutations occurring during DNA repair and

replication (1) or dipyrimidine lesions induced by ultraviolet light (2, 3). Some evolutionary comparisons have also suggested that simultaneous double-nucleotide substitutions occur at neighboring sites (4), but the significance and generality of these observations have been questioned (5). Thus, changes in neighboring nucleotides are usually attributed to coincidence of independent mutations.

We used two independent and complementary approaches based on sequence comparisons to study double-nucleotide substitutions and to obtain estimates of their frequency. The first approach examined changes that have occurred over long evolutionary time scales, between two particular dinucleotides, TC and AG. Serine is unique among amino acids in that it is encoded by two groups of codons, TCN and AGY, which cannot be interconverted by a single-nucleotide mutation. Switches between these groups of codons could occur indirectly, by two separate single-nucleotide mutations

<sup>1</sup>Institute of Molecular Biology and Biotechnology (IMBB)-FORTH, Vassilika Vouton, 711 10 Iraklio, Crete, Greece. <sup>2</sup>Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK. <sup>3</sup>Department of Genetics, Trinity College, University of Dublin, Dublin 2, Ireland. <sup>4</sup>Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK.

\*To whom correspondence should be addressed. E-mail: averof@imbb.forth.gr (M.A.) or paul@evol.nott.ac.uk (P.M.S.)