Principles for computational design of binding antibodies

Dror Baran[a,1,2], M. Gabriele Pszolla[a,1], Gideon D. Lapidoth[a], Christoffer Norn[a,3], Orly Dym[b], Tamar Unger[b], Shira Albeck[b], Michael D. Tyka[c], and Sarel J. Fleishman[a,4]

[a]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel; [b]The Israeli Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel; and [c]Google, Inc., Mountain View, CA 94043

Edited by William F. DeGrado, School of Pharmacy, University of California, San Francisco, CA, and approved August 25, 2017 (received for review May 1, 2017)

Natural proteins must both fold into a stable conformation and exert their molecular function. To date, computational design has successfully produced stable and atomically accurate proteins by using so-called "ideal" folds rich in regular secondary structures and almost devoid of loops and destabilizing elements, such as cavities. Molecular function, such as binding and catalysis, however, often demands nonideal features, including large and irregular loops and buried polar interaction networks, which have remained challenging for fold design. Through five design/experiment cycles, we learned principles for designing stable and functional antibody variable fragments (Fvs). Specifically, we (*i*) used sequence-design constraints derived from antibody multiple-sequence alignments, and (*ii*) during backbone design, maintained stabilizing interactions observed in natural antibodies between the framework and loops of complementarity-determining regions (CDRs) 1 and 2. Designed Fvs bound their ligands with midnanomolar affinities and were as stable as natural antibodies, despite having >30 mutations from mammalian antibody germlines. Furthermore, crystallographic analysis demonstrated atomic accuracy throughout the framework and in four of six CDRs in one design and atomic accuracy in the entire Fv in another. The principles we learned are general, and can be implemented to design other nonideal folds, generating stable, specific, and precise antibodies and enzymes.

Rosetta | AbDesign | V(D)J recombination | stability | expressibility

Due to their versatility, dozens of antibodies are in routine clinical use to diagnose and treat the most intransigent diseases and thousands more are used as research reagents. These antibodies were all isolated either by animal immunization or from synthetic repertoires that mimic the diversity of vertebrate immune systems. Notwithstanding these successes, however, natural repertoires have limitations, including biases and redundancy in representing the vast potential sequence and conformation space available to antibodies, and many antibodies exhibit polyspecificity and low expressibility, failing to meet the stringent requirements of research or clinical use (1–3). It has therefore been a long-standing goal of protein engineering to "build antibodies from first principles" (4).

Computational protein design has mostly targeted so-called "ideal" proteins with high secondary-structure content, where polar backbone atoms form regular, short-range hydrogen bonds (5–9); irregularities, such as those seen in long loop regions, were almost absent from these designs. By contrast, the functional surfaces of most natural proteins, including antibodies, contain nonideal features, such as unpaired polar groups, buried charges, and long loops that are essential for function (10, 11). It has therefore been postulated that computational design of "nonideal" backbone and sequence features is of fundamental importance for understanding protein structure, stability, and function, and may open the way to the application of computational-design methodology to difficult problems in design of function (7, 8, 11–13).

The antibody variable fragment (Fv) served us as an exemplary target for design of nonideal backbones since it comprises six loop segments in the antigen-binding surface [complementarity-determining regions (CDRs) L1–L3 in the light chain and CDRs H1–H3 in the heavy chain]; many other protein folds, including TIM-barrel enzymes and β-propellers, similarly use loops in active sites (7, 11, 12). Furthermore, the antibody Fv comprises two chains, light and heavy, adding a layer of complexity so far absent from fold-design studies. Finally, three decades of protein-engineering experience and the availability of >1,000 antibody molecular structures suggested that antibody Fv design would present an excellent opportunity to learn principles of design of function.

Results

Design Algorithm. Antibody CDR backbones are stabilized by irregular interactions of backbone and amino acid side chains comprising both short- and long-range contacts, including buried polar networks. To overcome the challenges in designing such nonideal features, we developed an algorithm called AbDesign (14), which operates in three stages (Movie S1): First, natural antibody Fv backbones are segmented into constituent parts, and new backbones are designed by recombining segments from

Significance

Antibodies are the most versatile class of binding molecule known, and have numerous applications in biomedicine. Computational design of antibodies, however, poses unusual difficulties relative to previously designed proteins, as antibodies comprise multiple nonideal features, such as long and unstructured loops and buried charges and polar interaction networks. We developed an algorithm that uses information on backbone conformations and sequence-conservation patterns observed in natural antibodies to design new antibody binders. Designed antibodies were very different in sequence from natural ones, but had similarly desirable properties of affinity and stability, and molecular structures showed high accuracy relative to the design models. The design principles we implemented can be used to design other functional folds, including many enzyme classes.

Author contributions: D.B., M.G.P., and S.J.F. designed research; D.B., M.G.P., O.D., T.U., S.A., and M.D.T. performed research; G.D.L. and C.N. contributed new reagents/analytic tools; D.B., M.G.P., O.D., S.A., and S.J.F. analyzed data; and M.G.P. and S.J.F. wrote the paper.

Conflict of interest statement: S.J.F. is a consultant for IgC Bio Ltd. The Weizmann Institute of Science has filed a patent on antibody design.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors for the unbound antigen-binding fragments 5ins16_ev and 5ins14 have been deposited in the RCSB Protein Data Bank with accession codes 5NB5 and 5NBI, respectively.

[1]D.B. and M.G.P. contributed equally to this work.

[2]Present address: IgC Bio Ltd., Ness Ziona 7403622, Israel.

[3]Present address: Department of Biochemistry and Structural Biology, Center for Molecular Protein Science, Lund University, SE-221 00 Lund, Sweden.

[4]To whom correspondence should be addressed. Email: sarel@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707171114/-/DCSupplemental.

different natural antibodies; second, these newly designed backbones are docked against a target antigenic surface; and, third, for each backbone segment in the designed antibody, different conformations from natural antibodies are sampled and the sequence is optimized by Rosetta design calculations. To solve the problem of simultaneously designing a protein fold and its binding activity, the last step optimizes both antibody stability and binding energy jointly (15); previous computational-design algorithms, by contrast, concentrated on only one feature, either stability or binding, depending on the application (5, 6, 8, 9, 16, 17).

AbDesign was developed through five consecutive design/experiment cycles, in which a total of 193 designed antibodies were experimentally evaluated (Table S1 and Dataset S1). We chose two target antigens, human insulin and the *Mycobacterium tuberculosis* acyl-carrier protein (ACP) 2, proteins that are rigid and stable, and for which no antibody-bound molecular structures were available. Each design was formatted as a single-chain variable fragment (scFv) (18), and its expressibility and binding to the designated antigen were assayed using yeast cell surface display (19). Yeast display showed that antibodies from the first design cycle had uniformly low expression levels compared with many natural antibodies (Fig. 1). Since expression levels are often correlated with protein stability (20, 21), we reasoned that low expression indicated design flaws, and through the five design/experiment cycles, yeast display expression levels served us as an invaluable metric to gauge progress as we improved the design algorithm. With the benefit of hindsight, there were two necessary criteria for designing stable and expressible antibody binders: (*i*) preservation of amino acid identities crucial for configuring the Fv backbone, including buried polar networks, and (*ii*) identification of appropriate backbone-segmentation points in the Fv framework.

A "Learning Loop" of Antibody Design Principles. The most significant flaw in design cycles 1–3 was unpaired charges and cavities in the Fv core (Fig. 1). Fold-design studies of the past did not yield such flaws, because all backbone polar groups in ideal folds form hydrogen bonds that stabilize secondary structures (5, 16), whereas antibody CDRs have many buried polar groups that require sequence-specific interactions to configure correctly. To address

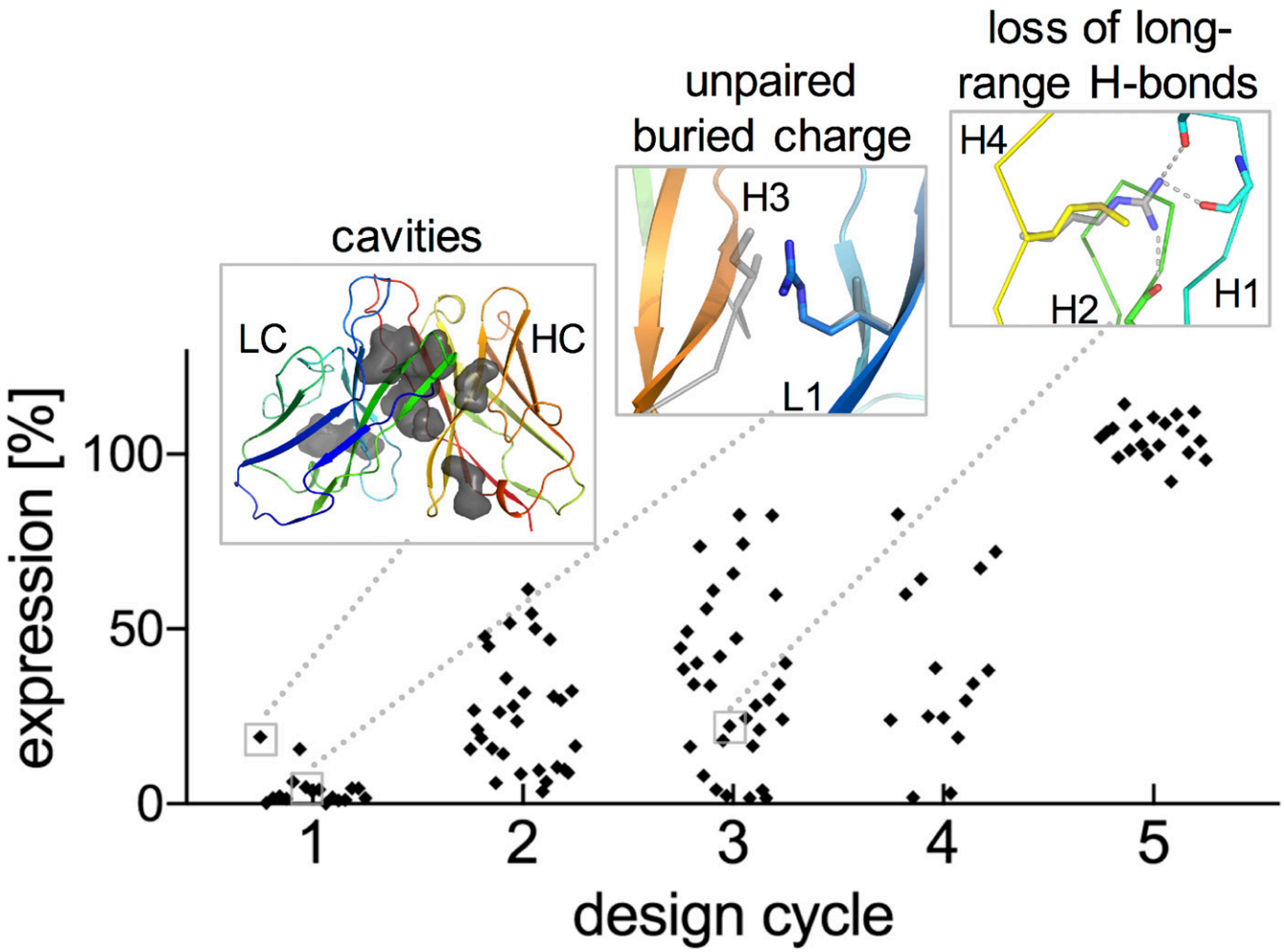

Fig. 1. Improved antibody expressibility through five design/experiment cycles. The 114 insulin-targeting designs were formatted as scFvs and their yeast surface expression levels were evaluated in five successive cycles of algorithm development (expression levels are normalized to those of the high-expression antibody 4m5.3 tested under identical conditions). Molecular representations show flaws observed in early design cycles: cavities (gray) in the protein core (1ins01) (*Left*), a buried but unpaired arginine (1ins10) (*Center*), and failure to maintain a buried hydrogen-bonding network between segments distant in sequence (3ins17) (*Right*). Starting in design cycle 4, conformation-dependent sequence constraints were used to guide Rosetta design choices. In this cycle, the entire Fv (framework and CDRs) was subjected to Rosetta design. Additionally, in cycle 5, the Fv backbone was segmented in two parts in each chain: one comprising the framework and CDRs 1 and 2 and another comprising CDR 3. Side chains in gray show identities typical of natural antibodies in the relevant positions. The backbone is rainbow-colored from the amino terminus (blue) to carboxy terminus (red). HC, heavy chain; LC, light chain.

these sequence-design flaws, starting in cycle 4, we implemented a conformation-dependent sequence-constraint strategy. Specifically, all natural Fv backbone conformations were clustered according to backbone similarity, and for each cluster, a position-specific scoring matrix (PSSM) was computed (21). As AbDesign assembled Fv backbones from segments of natural antibodies, it also assembled a corresponding PSSM to constrain sequence optimization to identities frequently observed in Fv multiple-sequence alignments. Since the use of PSSM constraints dramatically reduces sequence space open to design (14), we opened the entire Fv, comprising >200 amino acid residues, to sequence design in cycle 4, thereby optimizing the framework, including the interactions between light and heavy chains, and the framework's interactions with the CDRs. The sequence constraints notwithstanding, designed antibodies still showed substantial differences from mammalian germlines (>30 mutations; Dataset S1), demonstrating that the sequence space open to design was still vast. With the implementation of conformation-dependent sequence constraints, we observed modest to high expression levels for a majority of designs (Fig. 1). Such sequence constraints can be applied, in principle, to any natural protein to improve stability and expressibility (21, 22).

In design cycles 1–4, AbDesign segmented the Fv backbone into seven parts, comprising a single framework and six CDRs, following the conventional segmentation used by the majority of antibody-modeling and -engineering studies of the past three decades (23, 24) (Fig. 2). In the resulting designs, however, we observed structural defects, including cavities between CDRs 1 and 2 and the framework, as well as buried and unpaired polar groups. These observations led us to test whether the backbone differences among alternative Fv frameworks, albeit small, have a role in configuring the CDRs. In cycle 5, we therefore segmented each chain into two parts, one encompassing CDRs 1 and 2 and their supporting framework and the other encompassing CDR 3 (Fig. 2 and Movie S1). As a boundary between the two parts within each chain, we chose the disulfide-linked cysteine proximal to CDR 3, since this disulfide is conserved in all antibodies both in sequence and structure.

The segmentation of design cycle 5 is similar to the V(D)J partitioning of all vertebrate antibodies (25), and we saw that designs using this segmentation retained the intricate hydrogen bonding observed in natural antibody structures. The designs also had higher and more realistic core-packing densities compared with cycle 1 designs (Fig. S1). As a result, cycle 5 designs finally showed uniformly high expression levels (Fig. 1). That success in designing antibodies with high expression only came by segmenting the Fv into parts that retained contacts between the framework and CDR loops suggests a potentially general principle for computational design: that loop conformation depends on the scaffold for support, and is sensitive even to small structural perturbations in the scaffold. Furthermore, several common antibody-engineering techniques, including CDR grafting from one antibody Fv onto a different framework (23), do not retain the contacts between the framework and the CDRs that are seen here to be critical for expressibility; not surprisingly, in many cases, these engineering approaches require subsequent optimization steps to correct structural flaws (26).

Experimental Characterization. Although, on average, first- and second-cycle designs showed many defects and low expression, following extensive manual design, including 32 mutations, one of the second-cycle ACP-targeting designs, 2acp12, showed high expression and specifically bound its target ACP (Fig. S2, Table S1, and Dataset S1). The algorithm of cycle 5 consistently produced high-expression antibodies, and two additional binders, 5ins16 and 5acp14, which bound insulin and ACP, respectively, were isolated. As a further indication of improvements in the design algorithm, cycle 5 designs required substantially fewer (approximately five) manually introduced mutations than designs from cycles 1–3 (Table S1).

To increase the binders' affinity for their designated targets, we used error-prone PCR to introduce one to four random mutations in the scFv-coding genes of the three designs, followed by yeast display and fluorescence-activated cell sorting of higher affinity variants (19). Affinity in isolated clones increased by approximately an order of magnitude, from K_d = 900 to 50 nM for 2acp12 and from K_d = 300 to 30 nM for 5ins16 (Fig. S3*A*); due to its low initial affinity, no estimate was made for the affinity of 5acp14. Mapped on the design models, none of the mutations isolated by screening occurred at positions that were in direct contact with the target proteins, supporting the designed binding mode; instead, several mutations to positive charges were introduced at the periphery of the 5ins16 binding surface, likely improving long-range electrostatic attraction to the negative charge on insulin's surface (Fig. 3 and Dataset S1). We additionally isolated two mutations in the framework of 5ins16 near H3. Since mutations were only observed in the framework and away from the designed contact surface with insulin, we concluded that the binding surfaces largely formed as designed but that the H3 backbone was not optimally configured

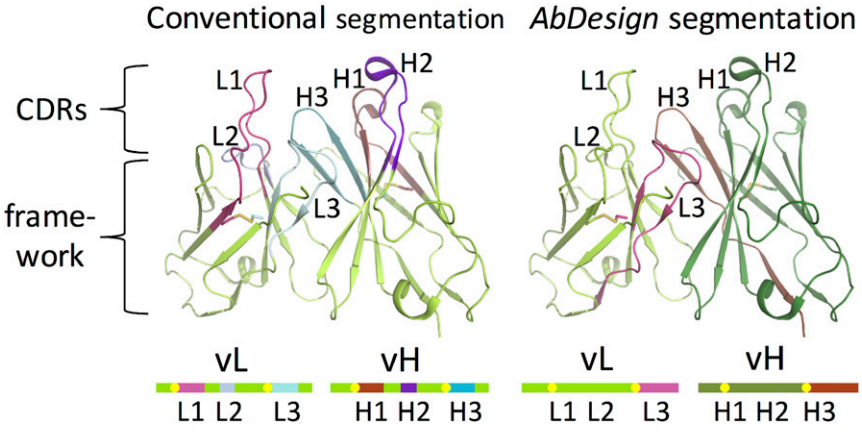

Fig. 2. Backbone segmentation of the antibody Fv. (*Left*) Conventional antibody design and engineering studies segmented the Fv into seven parts (a framework and six CDRs) and generated antibodies by combining segments from various antibodies. (*Right*) Segmentation used by AbDesign in design cycle 5, by contrast, uses four parts: two comprising CDRs 1 and 2 and the framework and two comprising CDR 3. The latter segmentation maintains the structural interactions between CDRs 1 and 2 and the framework, resulting in improved core packing. The conserved disulfides are shown as sticks in structural representations and as yellow dots on the primary-sequence representation. vH, heavy-chain variable domain; vL, light-chain variable domain.

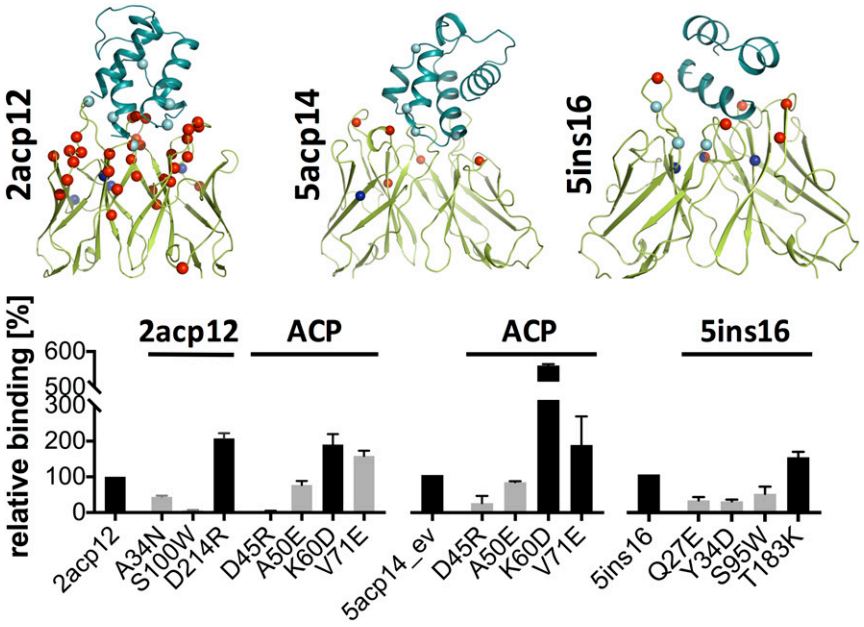

Fig. 3. Mutation analysis of designed binding modes. (*Top*) Design models were visually inspected, and mutations were introduced manually to improve packing and/or solvation (red spheres). For assessing the binding mode, three to four single-point mutations (cyan spheres) were introduced in designed antibodies 2acp12 and 5ins16 (5acp14 was not subjected to mutagenesis because of low affinity). (*Bottom*) Relative binding signal of variants containing single-point mutations in ligand-binding sites (gray bars) or outside the ligand-binding sites (black bars). Additional manual mutations were introduced on the ligand ACP (cyan spheres) in the intended binding surface (gray bars) or outside the binding surface (black bars) and tested for binding of the affinity-matured antibodies. Finally, mutations were introduced in the designed antibodies also through random mutagenesis and in vitro selection of improved binders (blue spheres). Mutations at the binding surface reduced affinity, except for Val71Glu on ACP, which enhanced binding of the affinity-matured 2acp12_ev. Mutations away from the interface, including those isolated from in vitro selection, increased binding affinity. Binding signals were tested in yeast surface display with the antibodies expressed as scFv and the ligand at 50 nM concentration for 2acp12_ev and at 1 μM concentration for 2acp12, 5acp14_ev, and 5ins16. SEs from the mean for two independent experiments are indicated.

by neighboring framework residues. We similarly isolated four mutations in 2acp12 and one mutation in 5acp14, all located in the framework at a distance >8 Å from the designed antigen-binding surface.

We sought to further verify the binding mode by introducing point mutations in the designed antibody CDRs. All mutations introduced in the core of the binding surface on 2acp12 and 5ins16 decreased binding (Fig. 3 and Dataset S1). Conversely, mutations to charged amino acids >8 Å from the designed binding surface increased affinity, likely due to electrostatic attraction. We next assessed the binding mode from the ligand's side of the interface by introducing mutations to ACP, both inside and outside the surfaces through which the antibodies were designed to interact with it. In agreement with the design models, mutations on ACP >8 Å from the binding surface did not perturb its binding to 2acp12 or 5acp14, whereas four of five mutations at the designed binding interface of 2acp12 and ACP interfered to some extent with binding (on 2acp12: Ala34Asn and Ser100Trp; on ACP: Asp45Arg and Ala50Glu); the fifth interface-core mutation (Val71Glu) improved affinity to 2acp12. The observation that most but not all core mutations in the 2acp12 interface decrease binding indicates that the binding mode of 2acp12 and ACP overlaps with the designed mode, but perhaps not in its entirety. Collectively, the mutational analysis suggested that the antibodies bound through at least parts of the designed surfaces, but similar to germline antibodies, remained flexible, and therefore susceptible to affinity-enhancing framework mutations (27).

By this point, we had tested the designed antibodies as scFv constructs, where the light and heavy chains were fused and attached to the yeast cell surface. We next tested the antibodies as soluble two-chain antigen-binding fragments (Fabs). The three antibodies showed apparent melting temperatures (T_m) in the range of 57–79 °C (Fig. S3*B*), whereas natural affinity-matured Fabs often show T_ms around 70 °C (28). Furthermore, measured by surface plasmon resonance, the Fabs had affinities in the range of 50–100 nM for their targets, despite the fact that the designs contained dozens of mutations compared with mammalian antibody germlines (Fig. S3*C*).

Finally, we determined the X-ray crystallographic structure of two designed Fabs, 5ins14 and the affinity-matured 5ins16_ev, both without ligand. Electron density was observed in all regions of the designed antibodies, allowing unambiguous assignment of the structure (Fig. S4*A* and Table S2). Furthermore, a comparison with the designed models revealed atomic accuracy throughout most of the Fv, with total root-mean-square deviations (rmsds) of 0.4 Å and 0.7 Å, respectively (Fig. 4), despite the fact that in design cycle 5, the entire Fv was subjected to backbone and sequence design. Detailed packing in the cores of the light chains, the light–heavy chain heterodimer interfaces, and hydrogen-bonding networks fastening the CDRs to the framework were also atomically accurate. The largest difference between the design model and experimental structure was in the interaction between H1 and H3 of 5ins16, where residues at the stem of H3 were packed differently, leading to a tilt in the H3 experimental structure and to a conformational change in H1 (1.8-Å and 2.0-Å rmsds for H3 and H1, respectively; Fig. 4*A*). We also noticed that H1 and H3 were stabilized in the altered conformation by packing interactions among crystallographic-symmetry neighbors (Fig. S4*B*). Nevertheless, these changes did not propagate to other parts of the Fv, highlighting its structural modularity. We were unable to obtain a crystal structure of 5ins16 bound to insulin, but we noted in our design model that the antibody engaged insulin primarily through H2, L1, and L3, where design accuracy was high (rmsd < 0.6 Å; Fig. 4*A*). To verify that the crystal structure was indeed compatible with insulin binding, we computationally aligned the crystal structure with the 5ins16 design model, producing a model of

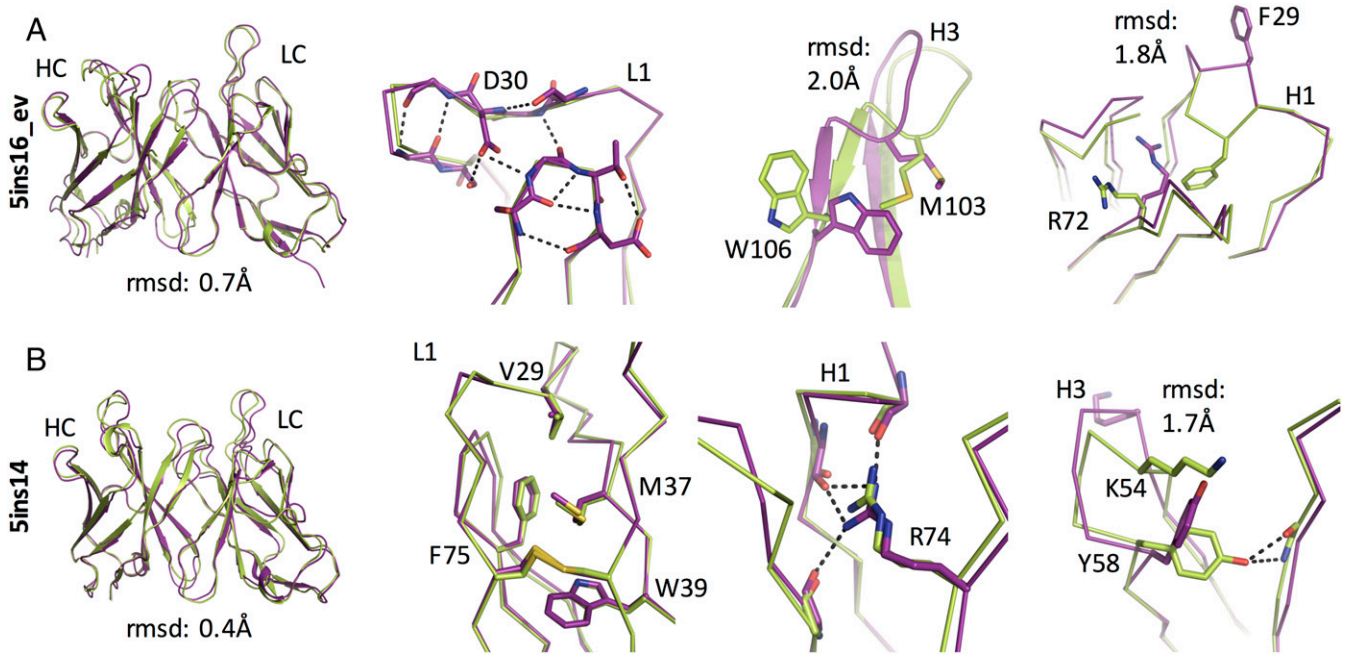

Fig. 4. Comparison of design models and experimental structures of 5ins16_ev and 5ins14. In both designed antibodies, the light chain (LC), the backbone conformation of the framework, and the LC–heavy chain (HC) heterodimer interfaces are atomically accurate. (*A*) 5ins16_ev: Backbone and side-chain packing deviations occur in H1 and H3, but other regions, including buried hydrogen-bonding networks (dashed lines) involving L1, are atomically accurate. (*B*) 5ins14: Core packing of hydrophobic residues on the framework specifying the L1 conformation, as well as a buried polar network specifying H1, are atomically accurate. Model and experimental structures are colored in lime and purple, respectively. Antibodies were expressed and crystalized as Fabs, and only the Fvs are shown.

the crystal structure bound to insulin through the designed binding mode. We then performed local all-atom docking simulations using RosettaDock to identify low-energy conformations (29); only limited side-chain conformation changes were observed between the 5ins16 design and this minimized conformation, suggesting that the experimental structure is compatible with binding to insulin through the designed binding mode (Fig. S4*C*). Furthermore, 5ins14 showed high accuracy throughout the entire Fv, including in regions with long loops and buried charges (Fig. 4*B*). We therefore concluded that the algorithm generated antibodies that were substantially different from natural antibody germline sequences (Table S1), and yet showed similar stability and binding properties; the designed antibodies also showed atomic accuracy in the framework and majority of the CDRs, including in regions containing nonideal features, such as long loops and buried polar networks, that have so far defied binder design.

Discussion

The antibody Fv is a larger and more complex structure than the proteins that have previously been the subject of fold design (5, 8, 12, 16), yielding three principal complications for computational design. First, as the Fv is a heterodimer, accurate design of not only one but two domains, as well as their interaction, is required. Second, each of these domains comprises long loops in its CDRs, which are stabilized by irregular, sequence-specific interactions within the CDRs and long-range interactions with the framework. Third, the Fv contains buried polar networks that are, on the one hand, crucial for configuring the CDRs but, on the other hand, may be disfavored from the standpoint of native-state energy, which favors completely hydrophobic cores. Indeed, buried polar networks in fold design have only been implemented in symmetric homo-oligomers, and have required specialized search heuristics (30). To overcome the multiplex challenge of maintaining the architecture of the Fv while incorporating the required nonideal features, the AbDesign algorithm used a combination of sequence and structure constraints derived from a bioinformatics analysis of natural Fvs. Despite the use of constraints, the sequence and backbone-conformation space was still vast, and allowed us to design binders for two arbitrarily chosen small proteins. Designed structures furthermore showed atomic accuracy in the framework, heterodimer interface, and in one of the designs in all CDRs.

Due to the size and complexity of the Fv structure, we found that two types of modeling constraints were required to make stable and expressible antibodies: (*i*) conformation-specific sequence constraints, and (*ii*) use of large backbone fragments that included CDRs 1 and 2 and their supporting framework; these constraints reduce and simplify sequence and conformation space, respectively. Both constraints are inspired by protein-engineering strategies. The use of sequence constraints in AbDesign has parallels in "consensus" design, which has been used to improve the stability of antibodies, enzymes, and repeat proteins (17, 31, 32). However, instead of using the most likely amino acid identity at each position, as in consensus design, AbDesign chooses the energetically most favorable one out of all identities that are likely at each position; these sequence constraints thereby allow room to optimize the sequence for the requirements of antigen binding, particularly in the hypervariable CDRs (21). Furthermore, the backbone-conformation fragments used in AbDesign are similar in spirit to the large gene fragments used in the schema structure-based genetic recombination of homologs, which has been successfully implemented in enzyme engineering (33); however, AbDesign optimizes the sequence of the combined fragments to increase antibody stability, rather than using the natural sequences as they are. Hence, the AbDesign algorithm combines evolutionary-based protein-engineering strategies with atomistic design, and the results indicate that this combination yields high stability and expressibility in designed antibodies.

Antibodies are ubiquitous in biomedical research and clinical practice. AbDesign therefore holds promise for the future generation of specific and stable research, diagnostic, and therapeutic tools. To achieve these goals, efforts should now be directed to targeting surfaces that have been traditionally difficult for antibody engineering, such as conserved sites in large human proteins, and to achieve consistently high design accuracy in all CDRs, including the flexible H3. Finally, the evolutionary-based constraints we adopted, and indeed the general iterative-learning approach through algorithm development and experimental testing, are applicable to any protein family of modular fold, including other large proteins that serve as binders or enzymes (12, 33). Future binders and enzymes may therefore be built by modular assembly of backbones from natural homologs, and then optimized through computational design for stability and molecular function.

Methods

Computational Methods. Computational design, modeling, and bioinformatics analysis were performed as described in *SI Methods*. RosettaScripts and commandline instructions are available in Dataset S2.

Experimental Procedures. Cloning, expression, purification, yeast display analysis, thermal stability, and structure determination were performed as described in *SI Methods*.

ACKNOWLEDGMENTS. We thank Gideon Schreiber and Olga Khersonsky for help in surface plasmon resonance analysis, and Ingemar André for critical reading. The research was funded by a European Research Council Starter's Grant (335439), the Israel Science Foundation through its Center for Research Excellence in Structural Cell Biology (Grant 1775/12), and a charitable donation from Sam Switzer. C.N. was supported by a PhD fellowship from the Boehringer Ingelheim Fonds. The work was also supported by a donation of 500 million core processing unit hours from Google, Inc. through the Exacycle eScience program and by a donation of 300 custom synthesized genes from Gen9 Bio, Inc.

1. Xu Y, et al. (2013) Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: A FACS-based, high-throughput selection and analytical tool. *Protein Eng Des Sel* 26:663–670.
2. Bradbury A, Plückthun A (2015) Reproducibility: Standardize antibodies used in research. *Nature* 518:27–29.
3. Jain T, et al. (2017) Biophysical properties of the clinical-stage antibody landscape. *Proc Natl Acad Sci USA* 114:944–949.
4. Winter G, Milstein C (1991) Man-made antibodies. *Nature* 349:293–299.
5. Koga N, et al. (2012) Principles for designing ideal protein structures. *Nature* 491: 222–227.
6. Fleishman SJ, et al. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332:816–821.
7. Stranges PB, Kuhlman B (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci* 22: 74–82.
8. Jacobs TM, et al. (2016) Design of structurally distinct proteins using strategies inspired by evolution. *Science* 352:687–690.
9. Kuhlman B, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–1368.
10. Ferreiro DU, Komives EA, Wolynes PG (2014) Frustration in biomolecules. *Q Rev Biophys* 47:285–363.
11. Fleishman SJ, Baker D (2012) Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 149:262–273.
12. Khersonsky O, Fleishman SJ (2016) Why reinvent the wheel? Building new proteins based on ready-made parts. *Protein Sci* 25:1179–1187.
13. Netzer R, Fleishman SJ (2016) Inspired by nature: Designed proteins have structural features resembling those of natural active sites. *Science* 352:657–658.
14. Lapidoth GD, et al. (2015) AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins* 83:1385–1406.
15. Warszawski S, Netzer R, Tawfik DS, Fleishman SJ (2014) A "fuzzy"-logic language for encoding multiple physical traits in biomolecules. *J Mol Biol* 426:4125–4138.
16. Huang P-S, Boyken SE, Baker D (2016) The coming of age of de novo protein design. *Nature* 537:320–327.
17. Röthlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453:190–195.
18. Skerra A, Plückthun A (1988) Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science* 240:1038–1041.
19. Chao G, et al. (2006) Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* 1:755–768.
20. Shusta EV, Kieke MC, Parke E, Kranz DM, Wittrup KD (1999) Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J Mol Biol* 292:949–956.
21. Goldenzweig A, et al. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol Cell* 63:337–346.
22. Campeotto I, et al. (2017) One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proc Natl Acad Sci USA* 114:998–1002.
23. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 321:522–525.
24. Al-Lazikani B, Lesk AM, Chothia C (1997) Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 273:927–948.
25. Janeway CA, Travers P, Walport M, Shlomchik MJ (2005) *Immunobiology: The Immune System in Health and Disease* (Garland Science, New York), 6th Ed.
26. Kettleborough CA, Saldanha J, Heath VJ, Morrison CJ, Bendig MM (1991) Humanization of a mouse monoclonal antibody by CDR-grafting: The importance of framework residues on loop conformation. *Protein Eng* 4:773–783.
27. Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC (1997) Structural insights into the evolution of an antibody combining site. *Science* 276:1665–1669, and erratum (1997) 277:1423.
28. Wang F, et al. (2013) Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. *Proc Natl Acad Sci USA* 110:4261–4266.
29. Gray JJ, et al. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 331:281–299.
30. Boyken SE, et al. (2016) De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* 352:680–687.
31. Steipe B, Schiller B, Plückthun A, Steinbacher S (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol* 240:188–192.
32. Kohl A, et al. (2003) Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proc Natl Acad Sci USA* 100:1700–1705.
33. Hiraga K, Arnold FH (2003) General method for sequence-independent site-directed chimeragenesis. *J Mol Biol* 330:287–296.
34. Dunbar J, et al. (2014) SAbDab: The structural antibody database. *Nucleic Acids Res* 42:D1140–D1146.
35. Norn CH, Lapidoth G, Fleishman SJ (2017) High-accuracy modeling of antibody structures by a search for minimum-energy recombination of backbone fragments. *Proteins* 85:30–38.
36. Wu TT, Kabat EA (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* 132:211–250.
37. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res* 33:W363–W367.
38. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* 234:946–950.
39. Cooper S, et al. (2010) Predicting protein structures with a multiplayer online game. *Nature* 466:756–760.
40. Fleishman SJ, et al. (2011) RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* 6:e20161.
41. Das R, Baker D (2008) Macromolecular modeling with Rosetta. *Annu Rev Biochem* 77: 363–382.
42. Hoover DM, Lubkowski J (2002) DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* 30:e43.
43. Benatuil L, Perez JM, Belk J, Hsieh C-M (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* 23:155–159.
44. Read RJ, Sussman JL, eds (2007) *Evolving Methods for Macromolecular Crystallography* (Springer, Dordrecht, The Netherlands).
45. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62:72–82.
46. French S, Wilson K (1978) On the treatment of negative intensity observations. *Acta Crystallogr A* 34:517–525.
47. McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63:32–41.
48. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
49. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
50. Vedadi M, Arrowsmith CH, Allali-Hassani A, Senisterra G, Wasney GA (2010) Biophysical characterization of recombinant proteins: A key to higher structural genomics success. *J Struct Biol* 172:107–119.