Biochemistry

pubs.acs.org/biochemistry

From the Bench

Modeling Immunity with Rosetta: Methods for Antibody and Antigen Design

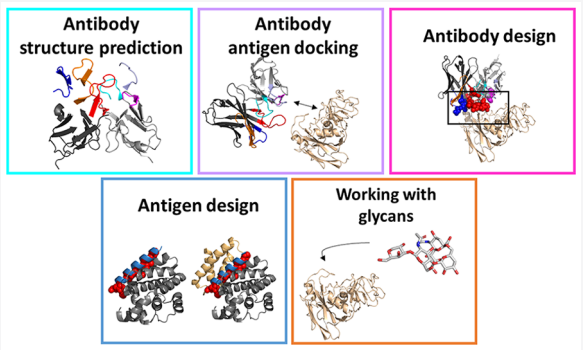
Clara T. Schoeder, Samuel Schmitz, Jared Adolf-Bryfogle, Alexander M. Sevy, Jessica A. Finn, Marion F. Sauer, Nina G. Bozhanova, Benjamin K. Mueller, Amandeep K. Sangha, Jaume Bonet, Jonathan H. Sheehan, Georg Kuenze, Brennica Marlow, Shannon T. Smith, Hope Woods, Brian J. Bender, Cristina E. Martina, Diego del Alamo, Pranav Kodali, Alican Gulsevin, William R. Schief, Bruno E. Correia, James E. Crowe, Jr., Jens Meiler,* and Rocco Moretti*

Cite This: *Biochemistry* 2021, 60, 825−846

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Structure-based antibody and antigen design has advanced greatly in recent years, due not only to the increasing availability of experimentally determined structures but also to improved computational methods for both prediction and design. Constant improvements in performance within the Rosetta software suite for biomolecular modeling have given rise to a greater breadth of structure prediction, including docking and design application cases for antibody and antigen modeling. Here, we present an overview of current protocols for antibody and antigen modeling using Rosetta and exemplify those by detailed tutorials originally developed for a Rosetta workshop at Vanderbilt University. These tutorials cover antibody structure prediction, docking, and design and antigen design strategies, including the addition of glycans in Rosetta. We expect that these materials will allow novice users to apply Rosetta in their own projects for modeling antibodies and antigens.



Antibodies are used in a variety of applications, ranging from therapeutics for life-threatening diseases to molecular probes and diagnostic molecular biology tools.[1−3] With the increasing number of available antibody structures from methods such as X-ray crystallography, NMR, and cryo-electron microscopy (cryo-EM), including many structures of antibody−antigen complexes, the molecular determinants of antibody specificity, affinity, and selectivity not only can be predicted but also can be engineered for the intended purposes. Computational methods have been successful in predicting antibody structure,[4] increasing antibody affinity and breadth,[5] and designing antibodies to bind new targets.[6−9] These computational methods can also be applied to create antigens from known epitopes, which can be used as molecular probes as well as immunogens to study antibody responses.[10−13] Such methods and protocols have been implemented in the Rosetta software suite and are readily accessible for additional application cases.

Rosetta contains packages and frameworks for protein structure prediction and design and is maintained and developed by the RosettaCommons, an international collaboration of currently more than 70 academic research groups.[14] Within its software framework, Rosetta comprises various applications for molecular modeling, including protein structure prediction, ligand docking, homology modeling, protein design, and RNA structure prediction.[9,15] Rosetta protocols can incorporate experimental data to guide computational predictions and can be easily modified via different interfaces, including Python and XML.[16,17] General information about how to use Rosetta is available on the RosettaCommons homepage (www.rosettacommons.org). Previous papers have also presented an overview of a number of general protocols and applications in Rosetta, accompanied by tutorials and examples to provide a general workflow for interested users.[4,9,15,18] These tutorials are a valuable resource for people who are new to Rosetta, and we encourage readers to look into these basic resources before starting with the more advanced protocols of this work.

As the field of structure-based antibody design has been evolving rapidly in the past few years, more and more antibody- and antigen-specific applications and protocols have become available in Rosetta. In this work, we summarize the protocols for antibody and antigen modeling, in particular those that were

Received: November 20, 2020
Revised: March 2, 2021
Published: March 11, 2021

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

Table 1. Overview of Available Protocols in Rosetta for Antibody and Antigen Modeling and Design

name	implementation	weblink	ref
		Antibody Structure Prediction	
RosettaAntibody	Application	https://www.rosettacommons.org/docs/latest/application_documentation/antibody/antibody-protocol	Weitzner et al.[4] Weitzner et al.[25] Sivasubramian et al.[26]
AbPredict	XML	http://abpredict.weizmann.ac.il https://www.rosettacommons.org/demos/latest/tutorials/AbPredict/AbPredict	Norn et al.[27]
RosettaCM	XML	https://www.rosettacommons.org/docs/latest/application_documentation/structure_prediction/RosettaCM	Song et al.[28]
		Antibody−Antigen Docking	
RosettaDock	XML	https://www.rosettacommons.org/docs/latest/application_documentation/docking/docking-protocol https://rosie.rosettacommons.org/docking/	Gray et al.[29] Chaudhury et al.[30] Chaudhury et al.[31] Marze et al.[32]
SnugDock	Application	https://www.rosettacommons.org/docs/latest/application_documentation/antibody/snugdock	Weitzner et al.[4] Sirkar et al.[33]
		Antibody Design	
RosettaAntibodyDesign	Application	https://www.rosettacommons.org/docs/latest/application_documentation/antibody/RosettaAntibodyDesign	Adolf-Bryfogle et al.[34]
AbDesign	XML		Lapidoth et al.[35]
RECON	XML	https://www.rosettacommons.org/docs/latest/application_documentation/RECON-multistate-design	Sevy et al.[5] Sevy et al.[6]
MSD with negative design states	XML	https://www.rosettacommons.org/demos/latest/protocol_capture/multistate_apl/README	Leaver-Fay et al.[36]
		Antigen Design	
side chain and backbone grafting	XML	https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/Movers/movers_pages/MotifGraftMover	Azoitei et al.[37] Correia et al.[38] Silva et al.[39]
FunFolDes	XML	https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/composite_protocols/fold_from_loops/FunFolDes	Bonet et al.[40] Correia et al.[11]
		Scaffold Design	
RosettaRemodel	Application	https://www.rosettacommons.org/docs/latest/application_documentation/design/rosettaremodel	Huang et al.[41]
SEWING	XML	https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/composite_protocols/sewing/SEWING	Jacobs et al.[42] Guffy et al.[43]
PROSS	Webserver	https://pross.weizmann.ac.il/step/pross-terms/	Goldenzweig et al.[44]
		Working with Glycans	
GlycanTreeModeler	XML	https://www.rosettacommons.org/docs/latest/application_documentation/carbohydrates/WorkingWithGlycans	Adolf-Bryfogle et al. (publication in preparation)
	PyRosetta		Labonte et al.[45]

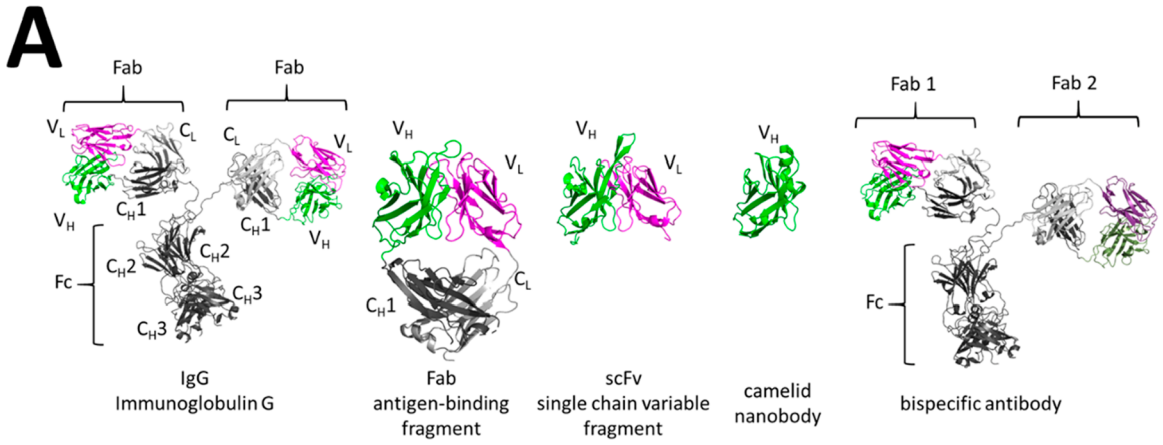
taught in the Rosetta workshop held at Vanderbilt University. State-of-the-art methods for antibody structure prediction, antibody−antigen docking, antibody affinity maturation, antibody design, antigen design, and glycan modeling will be discussed and illustrated by a tutorial section as part of the Supporting Information. Alongside these protocols, other methods and protocols in Rosetta will be compared to provide an overview of existing methods for modeling of proteins implicated in immunity.

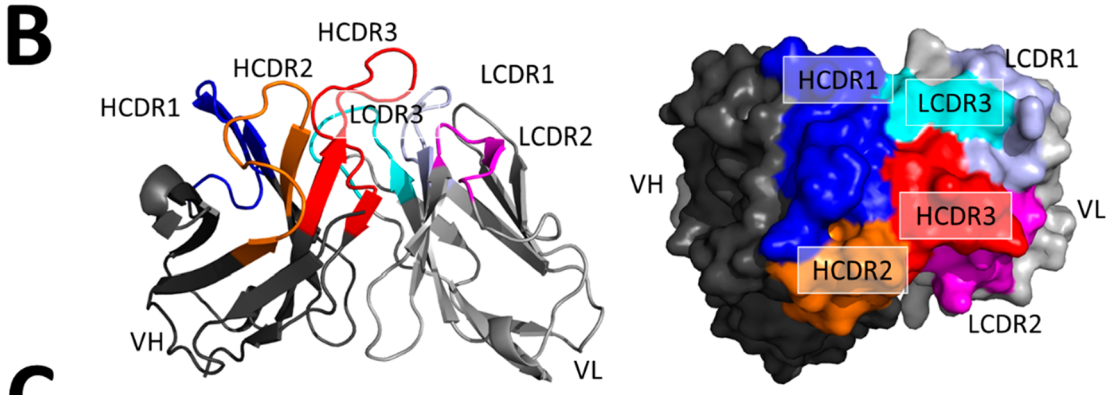
A BRIEF BACKGROUND ABOUT ROSETTA

Rosetta has emerged as one of the leading computational tools for biomolecular structure prediction and design. Most Rosetta protocols apply a Metropolis Monte Carlo sampling strategy to explore conformational space efficiently. Sampling is guided by the Rosetta energy function, which consists of a weighted sum of physics-based and statistically derived potentials.[19,20] This energy function is under continuous optimization and has been extended to provide scoring for a variety of different use cases, such as membrane proteins, small molecule ligand protein interactions, nucleic acids or noncanonical amino acids, etc.[21−24] The current default energy function ref2015, which will also be used in the provided tutorials for antibody and antigen modeling and design, has been reviewed extensively by Alford et al.[19]

Rosetta runs on most UNIX-like operating systems and comes as a suite of command line executable programs, based on the core Rosetta library. The RosettaScripts interface is an XML-based scripting language that allows for the development of custom protocols in a human readable text format.[16] For those who wish to have more control over Rosetta protocols, PyRosetta provides access to the Rosetta internal code structure through a Python scripting interface.[17] While most of the protocols presented here are in RosettaScripts XML format, a range of protocols in the RosettaAntibody framework have been implemented as stand-alone applications or are accessed easily through PyRosetta. Example command lines will be provided in the protocol section (Supporting Information). For more general information about the Rosetta protein design software suite, see Koehler-Lehman et al.,[9] Bender et al.,[15] and Leaver-Fay et al.[20] It should be noted that the tasks presented in the tutorials are examples for training purposes with limited

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846





C

LCDR1											
PyIgClassify	L1-11-1										
sequence	R	P	S	Q	N	I	R	S	F	L	N
Kabat	24	25	26	27	28	29	30	31	32	33	34
Chothia	24	25	26	27	28	29	30	31	32	33	34
IMGT				27	28	29	36	37	38		
AHo	24	25	26	27	28	29	38	39	40	41	42

HCDR1														
H1-15-1														
S	V	S	G	V	S	V	T	S	D	I	Y	Y	W	T
								31	32	33	34	35	35A	35B
				26	27	28	29	30	31	31A	31B	32		
				27	28	29	30	31	34	35	36	37	38	
24	25	26	27	28	29	30	31	36	37	38	39	40	41	42

LCDR2							
PyIgClassify	L2-8-1						
sequence	Y	A	A	S	N	L	Q	S
Kabat		50	51	52	53	54	55	56
Chothia			51	52	53	54	55	56
IMGT		56	65					
AHo	57	58	59	60	69	70	71	72

HCDR2														
H2-9-1														
Y	I	F	Y	N	G	D	T	N	Y	N	P	S	L	K	S
50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
		52	53	54	55	56									
	56	57	58	59	63	64	65								
57	58	59	60	61	66	67	68	69							

LCDR3									
PyIgClassify	L3-9-cis7								
sequence	Q	Q	S	Y	N	T	P	P	T
Kabat	89	90	91	92	93	94	95	96	97
Chothia	89	90	91	92	93	94	95	96	97
IMGT	105	106	107	108	109	114	115	116	117
AHo	107	108	109	110	111	135	136	137	138

HCDR3																		
H3-18-1																		
A	R	G	T	E	D	L	G	Y	C	S	S	G	S	C	P	N	H	
		95	96	97	98	99	100	100A	100B	100C	100D	100E	100F	100G	100H	101	102	
		95	96	97	98	99	100	100A	100B	100C	100D	100E	100F	100G	100H	101	102	
105	106	107	108	109	110	111	111.1	111.2	112.3	112.2	112.1	112	113	114	115	116	117	
107	108	109	110	111	112	113	114	115	130	131	132	133	134	135	136	137	138	

Figure 1. General overview of antibody structure and numbering schemes. (A) Structural overview of antibodies and antibody-derived structures. From left to right, an IgG (PDB entry 1IGT[48]) contains two heavy and two light chains, of which the antigen-binding fragment (Fab) is depicted in detail (PDB entry 6OBZ[49]). A single-chain variable fragment (scFv) is composed of only the variable region of the heavy and light chain (VL and VH, respectively); connected by a linker (PDB entry 5C2B[50]). A nanobody contains solely the VH domain (PDB entry 1F2X[51]). Lastly, a bispecific antibody carries two different variable domains. (B) Canonical structure of the antibody variable domain (PDB entry 6OBZ, namely FluA-20[49]), with color-coded the complementary-determining region shown as a cartoon (left) or a surface (right). (C) FluA-20 antibody loops numbered in the most common antibody numbering schemes (Chothia, Kabat, IMGT, and AHo) and their assignment to the canonical loop cluster as determined with PyIgClassify.[52]

numbers of calculations, which can be carried out on a personal laptop. As these tutorials are derived from in-person teaching with hands-on training, they are designed to be finished in a couple of hours. When these protocols are transferred to production runs with higher numbers of calculations, the usage of a computer cluster might become necessary, to achieve a reasonable amount of outputs (1000−10000) (Table 1).

METHODS FOR ANTIBODY STRUCTURE PREDICTION AND DESIGN

General Description of Antibody Structure and Antibody Numbering Schemes. Structure prediction and design of antibodies is facilitated by their common fold. Briefly, human antibodies consist of a heavy and a light chain, which share a well-conserved constant domain (F_c) and framework area (FR)

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

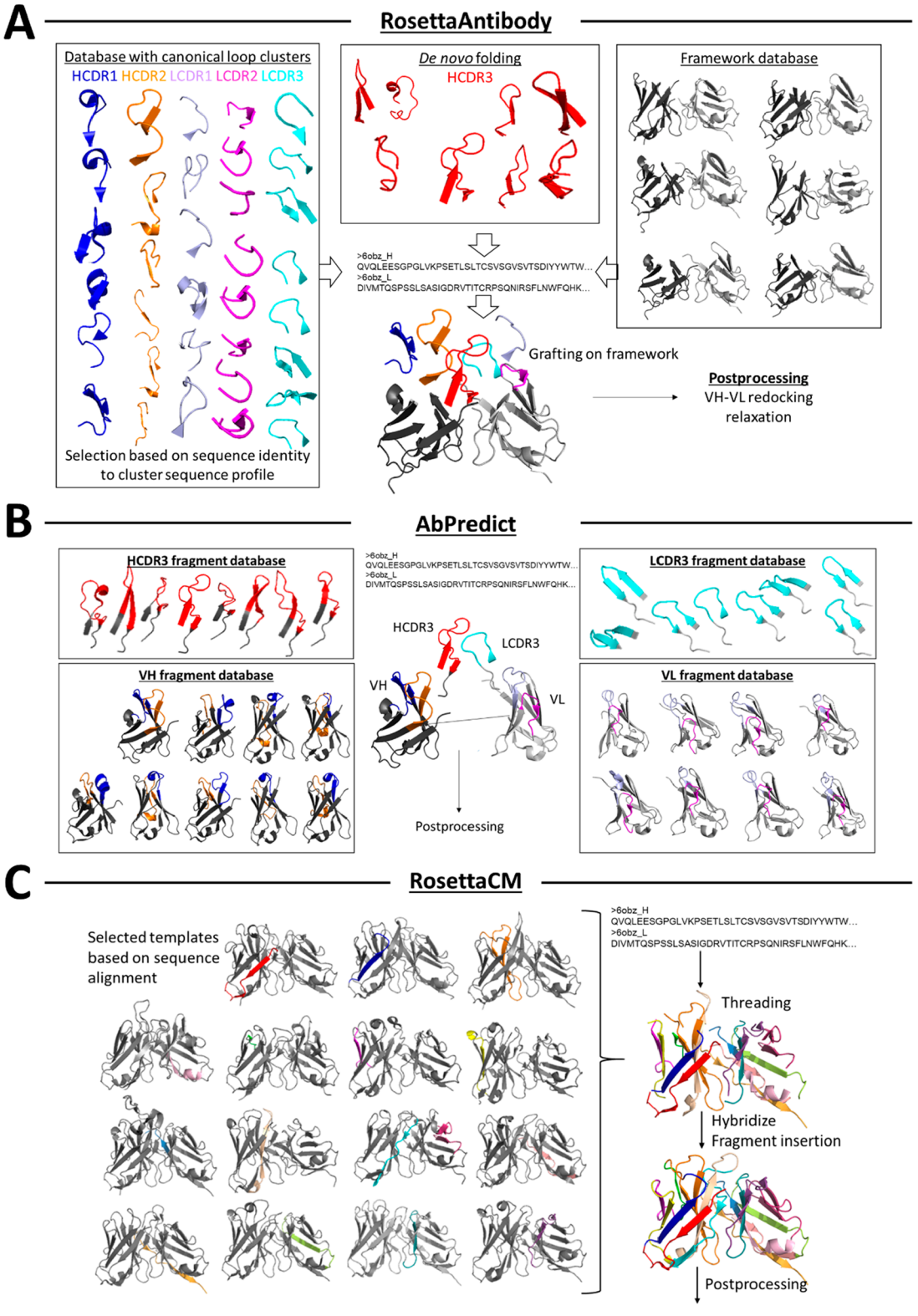


Figure 2. Methods in Rosetta for antibody structure prediction. (A) Schematic workflow of the RosettaAntibody application, in which HCDR1−2 and LCDR1−3 are modeled from templates in the loop database, and HCDR3 is *de novo* folded and grafted on a selected framework. (B) Schematic of the AbPredict protocol, which assembles an antibody from templates in four fragment databases, containing VL, LCDR3, VH, and HCDR3 templates. Antibody fragments displayed in panels A and B were taken from PDB entries 5ITB,[65] 5JRP,[66] 5CGY,[67] 5CHN,[67] 3QHZ,[68] 4GXV,[69] 6MEE,[70] and 5XRQ.[71] (C) Schematic overview of RosettaCM, which creates models by threading and hybridization of template structures based on user-provided sequence alignments (used PDB entries 4HT1,[72] 5UKP,[73] 5JRP,[66] 2EKS,[74] 4JPW,[75] 5TF1,[76] 5T4Z,[77] 6B0H,[78] 2R0K,[79] 4KMT,[80] 1WT5,[81] 5UMN,[82] and 4HS6[83]).

within the variable domain (F_v). Antibody variability is established through the process of recombination of the V, D, and J genes in the creation of the naïve B cell repertoire and by the subsequent somatic hypermutation of antibody variable

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

genes in the stimulated B cells during germinal center reactions. Sequence variation and structural variation of the antibody manifest in the complementary-determining regions (CDRs) as three highly variable loop regions in each heavy and light chain, which facilitates antigen recognition (compare Figure 1). The sequences of most antibodies are very similar in the constant and framework (FR) regions. The high variability in the CDR loop regions of the variable domain impedes accurate structure prediction and design of antibodies and has posed a significant challenge in modeling the native conformations of antibody−antigen structures.[46,47]

A number of studies have been undertaken to classify loop conformations of the antibodies' heavy and light chain variable domains (VH and VL) based on their experimentally determined sequences and structures, namely, HCDR1−3 and LCDR1−3, which determine the antibody−antigen interface.[53−55] Here, we use the abbreviations VH and VL to designate the entire variable domains of heavy and light chain proteins, respectively, comprising the Fv. V_H and V_L are often used in the literature to designate DNA gene segments found in the genome, which encode the FR1, CDR1, FR2, CDR2, and FR3 regions of the variable domain (but not CDR3 or FR4). Sometimes a HCDR4 and LCDR4, or DE loop, are also mentioned; however, these loops are not undergoing VDJ recombination but can be altered during somatic hypermutation. There have been cases reported in the literature noting the involvement of insertions in these regions to contribute to the stability and conformation of other loops.[56] Several numbering schemes have been introduced to identify the CDRs of a given antibody from sequence and to provide a consistent structure-based alignment system. RosettaAntibody applications offer the use of several prominent numbering schemes, including Chothia, Kabat, and AHo numbering schemes, to align CDRs spatially.[53,54,57] Another commonly used numbering scheme is IMGT numbering, which is derived from the gene assignment.[58,59] The requirements of an individual protocol should always be checked, in case there are special requirements for how the antibody numbering or chain designation needs to be annotated.

Antibody loops can be categorized by their structural similarity. Structure-based loop clustering for all but HCDR3 loops from available experimental structures has been reported by Chothia et al. and overhauled by North et al.[47,53] In the more recent update by North et al., canonical CDR loop conformations of distinct CDR clusters are identified using affinity propagation of structurally determined loops with similar amino acid lengths and differences in backbone dihedral angle geometries. This approach yielded 72 clusters for non-HCDR3 loops as compared to 25 clusters of non-HCDR3 loops in Chothia et al.[47,53] These clusters cover approximately 85% of all non-HCDR3 loop structures and can be used for antibody structure prediction and design.[47] For example, there are only two different length LCDR2 loops in the protein database (8 and 12 residues). Only four loop clusters exist for length 8, and two clusters for length 12, with one cluster containing 85% of all observed antibody structures. The structure of a LCDR2 loop therefore can be predicted with high accuracy given its cluster assignment, which can be performed by submitting a given antibody structure to the regularly updated PyIgClassify database and server (supported by the Dunbrack laboratory), which also provides a downloadable loop cluster database for CDR cluster assignment, as well as heavy chain V gene and light chain κ and λ gene assignment.[52] An exemplary overview of antibody numbering can be found in Figure 1.

Antibody Structure Prediction. In protein structure prediction, two major approaches are used: (1) *de novo* folding in the absence of a structural reference or template and (2) comparative modeling, which takes advantage of the availability of a structurally similar template to build a target model.[15] Given the large number of experimental antibody structures deposited in the PDB and the conserved immunoglobulin (Ig) fold, the large number of homology templates provides little to no need for *de novo* folding of the complete F_v domain. This makes antibodies ideal targets for comparative modeling approaches. However, the true challenge of antibody structure prediction lies in the correct orientation and fold of the CDRs, as all further scientific questions concerning antigen binding depend on the accurately modeled loop conformations. Excluding HCDR3, five of the six loops usually fall into canonical clusters as defined by North et al., which can greatly simplify structure prediction.[47,52] Here, we will review three available protocols for antibody structure prediction from sequence in Rosetta: RosettaAntibody, AbPredict, and RosettaCM.

The RosettaAntibody application uses a three-step protocol for modeling the variable domain from sequence (compare Figure 2A): (1) template selection for the framework and the five canonical loops, (2) grafting of selected templates into a preliminary model, and (3) HCDR3 *de novo* loop modeling while simultaneously optimizing the VH−VL interface orientation.[4,25,26] For template selection, a BLAST sequence search matches the parsed sequence to a modified copy of the PyIgClassify database provided as part of Rosetta to assign both the F_v template and CDR conformations. This assignment can be checked with the `identify_cdr_clusters` application in Rosetta such that any mismatches or other poor assignments within the template selection can be manually modified.[4,26] As a next step, the initial VH−VL orientation is diversified by sampling VH−VL orientations from the BLAST list based on light−heavy orientational coordinates (LHOC), a metric that combines the VL and VH opening angles, the packing angle between the VH and VL domains, and the interdomain distance.[60] Somatic hypermutation at the interface results in multiple angles between VL and VH even from sequences derived from the same germline genes such that a small difference in VH−VL distance and orientation may result in a drastic change in the CDR placement. This modulation of chain interface relationships has been investigated recently by Cisneros et al., who found VH−VL interface residues were reverted to the germline sequence, which resulted in significant loss of affinity, and indicated that the rigidification of the VH−VL interface, which will determine its orientation, is a major driver for affinity maturation.[61] RosettaAntibody selects 10 different framework matches as starting structures for loop grafting. The selected template loops are superimposed on the framework based on two overlapping residues and optimized through a cycle of minimizations, random torsional sampling and cyclic coordinate descent (CCD).[62,63] Subsequently, HCDR3 conformations are modeled with the next-generation kinematic loop closure (KIC) algorithm in a low-resolution step.[64] The full model is then refined in full atom mode, with the VH−VL orientation reoptimized with rigid-body docking,[29] and the model is subsequently refined with an additional high-resolution step of next-generation KIC, residue side chain packing, and minimization.[4,26]

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

Accurate modeling of the target antibody with RosettaAntibody relies on the availability of templates in the database that are highly similar in sequence to the antibody target. Most of the antibody structures determined so far are either human- or mouse-derived. Given the variability of the species-specific germline repertoire, such as the varying number of V genes or the different structural features represented, modeling of non-human or non-murine antibodies may be problematic due to the lack of appropriate templates. Therefore, when antibodies from other species are being modeled, it may be advisable to either curate a custom database or provide selected templates manually.

RosettaAntibody participated in both the 2011 and 2014 antibody modeling assessments (AMAs).[84,85] RosettaAntibody performed well overall on the basis of MolProbity scores and loop C_α RMSDs in AMA I.[84] In AMA II, RosettaAntibody was compared to six other software suites on a set of 11 unpublished antibody structures. It predicted 42 of 55 non-HCDR3 loops with an accuracy of better than 1 Å and generated the best HCDR3 model for 4 of 11 antibody structures from the other six competing methods.[25,85] Subsequent analysis of the AMA II results identified some areas in the protocol that had weakened its performance: the lack of good loop templates, the inaccurate modeling of the HCDR3 due to limitations in the loop modeling protocols, and the wrong orientation of the VH−VL interfaces.[25] All of these issues were addressed in the present RosettaAntibody protocol, which samples a variety of VH−VL starting structures[60] and incorporates next-generation KIC with HCDR3 conformational constraints.[4,86] The problem of missing starting structures, which prevents accurate sampling of rare CDR loop conformations, can be further improved only when more structural data are deposited in the PDB that are continuously integrated into PyIgClassify and the RosettaAntibody database.[4,52]

A protocol capture for antibody structural modeling with RosettaAntibody is described in the Supporting Information. For users who would like to have a graphical user interface, RosettaAntibody is further available through ROSIE [Rosetta Online Server that Includes Everyone (https://rosie.graylab.jhu.edu/antibody)], where the target sequence can be submitted via a web interface.[4,26,60,86−89]

A similar approach that combines antibody structural templates in another way has been implemented in the AbPredict protocol (compare Figure 2B).[27] AbPredict selects low-energy combinations of backbone fragments derived from experimentally determined structures of antibodies in the PDB.[27] The template antibodies are segmented into four parts: (1) heavy chain CDR3, (2) light chain CDR3, and (3 and 4) heavy and light chain V gene regions each containing CDR1 and CDR2 and the framework as defined by the conserved core disulfide in the variable region. Additionally, AbPredict considers the rigid-body orientation between VL and VH, which is represented by the spatial distance of the disulfide's cysteine residues to L88 and H92 (Kabat numbering). Briefly, a database of randomly recombined backbone fragments and rigid-body orientations with the target sequence length is created. After the target sequence has been threaded on a random starting conformation, a Monte Carlo search that samples backbone fragments from the curated database, repacks side chains, and minimizes the whole structure is executed, which is output as scFv.[27,35]

AbPredict has been benchmarked using the AMA II antibody set and compared to the methods presented therein. It performed in the upper third of all compared methods and showed beneficial performance in the prediction of the HCDR3 stem and the rigid-body orientation.[27]

Because AbPredict draws from an antibody template database provided as part of Rosetta, the representation of rare CDR loop length combinations is again a potential limitation, especially because AbPredict requires that target and template length match. A protocol capture is included within Rosetta.

Although antibody-tailored homology modeling protocols like RosettaAntibody can take advantage of knowledge-derived features of antibody structure, Rosetta's general multitemplate homology modeling protocol, RosettaCM can also be used (Figure 2C).[28] RosettaCM might be advantageous in specific cases, especially if the antibody structure shows noncanonical structure elements such as unusual loop lengths or conformations, which would not be available in the antibody template databases. Using the `DetailedControls` option, RosettaCM can be employed to model only specific ranges of peptide sequences within a protein, for example, just one CDR. A similar approach was used to model G protein-coupled receptor loop regions with great accuracy.[90]

Overall, for most antibody structure prediction tasks, a good starting point is to employ RosettaAntibody as described in the tutorial section. Depending on specific features of the target antibody such as template availability or unusual loop length, models may need further refinement. In this case, the user can consider using only selected templates or perform a partial remodeling with RosettaCM. It is advisable to run smaller test runs with only a few output models in the beginning and monitor the outcome for reasonable modeling performance by looking at the total_score, a metric for predicted protein stability, which should be negative. In production runs, up to 10000 models should be created, depending on the complexity of the modeling task and the specific requirements of the protocol. Using metrics such as the total_score and C_α RMSD, the performance of the modeling run and the quality of the models can be assessed. This can also be used to compare the modeling performance of different protocols.

HCDR3 Structure Prediction. Structure prediction of HCDR3 has been challenging to date due to its high length and conformational diversity. Although half of HCDR3 loops are shorter than 16 residues, HCDR3 has been described to adopt loop sizes far longer, up to 32 residues, and even longer outliers have been described (IMGT nomenclature).[47] The mean HCDR3 loop length has been determined to be 16 residues.[91] Ultralong HCDR3 loops (≥28 amino acids) have been described as necessary for the neutralization of disease states such as HIV or malaria,[92−94] making the accurate modeling of long HCDR3 loops increasingly important for the structure prediction of therapeutically relevant antibodies.

Canonical loop clustering fails in the case of HCDR3 due to its high degree of diversity. PyIgClassify lists HCDR3 up to lengths of 5−9 residues, which are more restrained in their structural diversity, but structural clusters are not defined for longer HCDR3 lengths.[52] However, the HCDR3 "torso" region, encompassing the first three (T1−T3) and the last four residues (T4−T7) of HCDR3 (based on the IMGT numbering scheme), can be classified into "kinked" ("bulged") or "extended" ("non-bulged").[55,95] The kinked conformation is predominant in antibodies, although structure prediction software rarely samples this conformation type.[46,86] In the past, sequence-based approaches have been employed to make a distinction between the kinked and extended conformation, relying on the

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

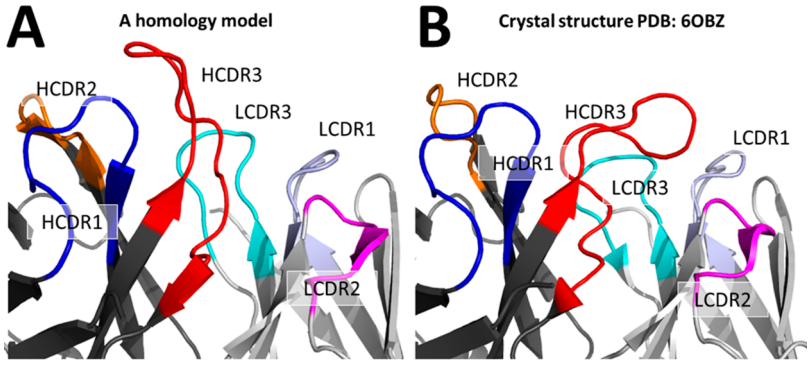


Figure 3. Incorrect long HCDR3 loop structure prediction (A) Model of FluA-20 created with RosettaCM. HCDR1 and HCDR2 are predicted very well; however, the HCDR3 loop has an incorrect conformation that will impair future studies using this model. (B) Experimental structure of FluA-20 (PDB entry 6OBZ[49]) for comparison.

presence of an arginine or lysine in the second position and an aspartic acid in the second to last position of the HCDR3 loop to classify an antibody as having a kinked conformation.[47,55] Although these amino acids are present in a large number of kinked conformations, they fail to cover the entirety of existing kinks.[46] Therefore, other metrics for describing the kink conformation have been introduced and used as penalties during loop modeling. In an independent protocol to more accurately model near-native HCDR3 loop conformations, Finn et al. described in greater detail the range of dihedral angles present in the torso region and identified a set of rules to guide kinked conformation sampling.[46] The dihedral angle restraints are defined by the ψ angle at the sixth torso residue (T6) and are added as a Rosetta constraint file using a circular harmonic scoring function that penalizes the incorrect torso residues. A tutorial using this approach can be found in the Supporting Information.

In RosettaAntibody, this limitation was overcome by integrating a structurally derived filter based on the kink definition by Shirai et al.[95] so that bulged conformations are enriched.[25] To refine the definition of a "kinked" HCDR3 loop, Weitzner et al. integrated the conformation bias constraint to increase the likelihood of sampling native-like geometries of the last two C-terminal dihedral angles of HCDR3 plus the following framework residue's dihedral angle (as defined by the Chothia numbering scheme).[86] Weitzner et al. hypothesized that the kink increases the degree of HCDR3 structural diversity by disrupting the propagation of β-strand pairing. Such a trend was also observed for proteins from other families where similar kinks occur in ligand recognition sites.[96]

Homology modeling of influenza hemagglutinin protein-specific human monoclonal antibody FluA-20 provides an illustrative example of challenging HCDR3 loop modeling (Figure 3). While the best scoring homology model created with a RosettaCM protocol had accurate HCDR1 and HCDR2 predictions, the HCDR3 tip is flipped compared to the crystal structure. Structure prediction methods have difficulty with FluA-20 due to its 18-residue HCDR3 loop. The rules and protocols that Finn et al. and Weitzner et al. provide are a good starting point to improve native-like HCDR3 placement despite its noncanonical conformation. Accurate prediction of all CDR loops, especially HCDR3, from an antibody modeling protocol is paramount in obtaining biologically relevant results in downstream protocols, such as antibody−antigen docking.

Camelid V_HH Nanobody Modeling. Camelid V_HH nanobodies are heavy chain only antibodies, which are interesting biotechnological tools and drug candidates, because of their thermodynamic stability, their compatibility with phage display techniques, and an equivalent breadth of epitope recognition ability as other antibody types. However, their lack of a light chain results in a number of differences from classical VH−VL domain antibodies. Sircar et al. analyzed structural variations in camelid V_HH nanobodies and implemented the observed rules in RosettaAntibody.[97] Because HCDR1 is more diverse in camelid V_HH nanobodies, it is subject to additional refinement during the modeling protocol. Additionally, due to this greater structural diversity, an ideal HCDR1 template from a canonical cluster is not always available; therefore, the template selection is done on the basis of a BLAST search alone. Also, a disulfide bond between HCDR1 and HCDR3 is regularly found,[98] which can be explicitly modeled. Sircar et al. further described the existence of three distinct structural motifs in camelid: neutral, twisted, and stretched−twisted. The last type creates a framework interaction in a position where the light chain normally interacts in a regular antibody. These structural motifs could be linked to sequence rules, enabling RosettaAntibody to select the corresponding constraints for the loop modeling process.[97] Furthermore, Sircar et al. presented a modified Chothia numbering system to enable RosettaAntibody to detect CDR regions with high reliability.[97] In recent years, the number of camelid V_HH nanobody structures deposited in the PDB has been growing so that more templates for camelid V_HH nanobody modeling have become available.

Antibody−Antigen Docking. The structural study of antibody−antigen complexes is crucial for the understanding of antibody−antigen interactions, guides optimization and design approaches of both docking partners, and ultimately helps develop new antibody-based therapies. Prediction of antibody−antigen complexes with computational protein−protein docking is of particular interest in investigating antibody function, as high-resolution experimental models of antibody−antigen complexes are rare due to the difficulty of co-crystallization. While more and more antibody−antigen complexes are now becoming available through the use of cryo-EM, the experimental data may not fully support atomic-level accuracy in all regions.

In Rosetta, a general protocol called RosettaDock can be employed for rigid-body docking with full backbone flexibility of two interacting proteins.[30,31,62] This protocol was reviewed previously by Bender et al.[15] and will be discussed only briefly here. A low-resolution docking step, where docking poses are identified by rigid-body movements about the surface of the

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

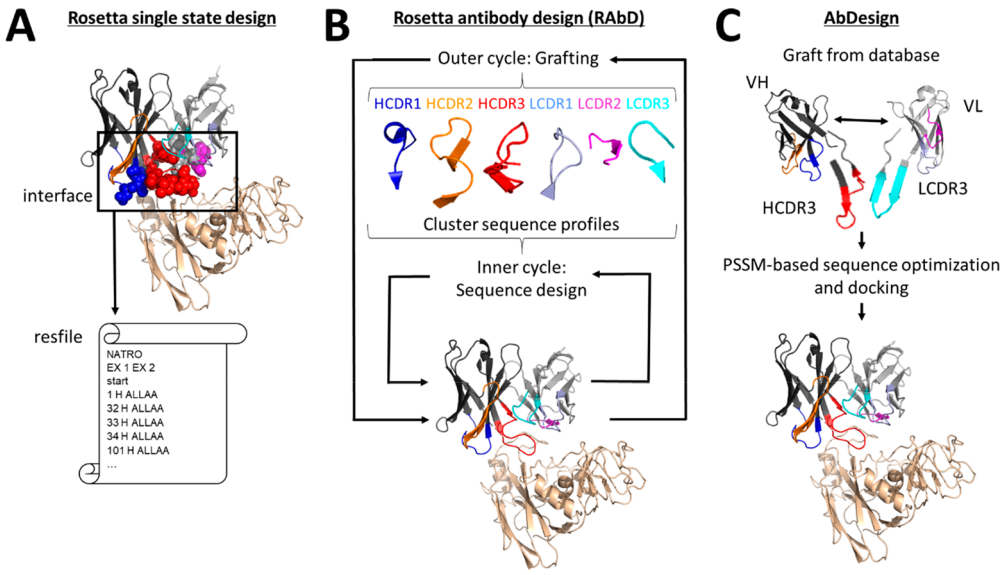


Figure 4. Comparison of Rosetta single-state design, RosettaAntibodyDesign (RAbD), and AbDesign. (A) In Rosetta single-state design, residues are redesigned in the interface according to a resfile. (B) Rosetta antibody design (RAbD) utilizes two cycles: the outer graft design and the inner sequence design cycle, with sequence design being based on the canonical loop clustering defined by PyIgClassify. (C) AbDesign recombines antibodies based on VH, VL, LCDR3, and HCDR3 segments and optimizes both the sequence and backbone conformation for antibody–antigen docking.

binding partner(s) (namely rotation and translation moves), is followed by a high-resolution step in full atom mode with fine-grained docking moves and side chain optimization stages.[29] RosettaDock requires as input a structure of both docking partners, optimally with a user-defined starting point. However, RosettaDock also can perform a global docking step to identify low-energy docking poses.[29,30] RosettaDock is implemented as a RosettaScripts protocol, which enables a high degree of customization,[20,30] and an example protocol for general antibody–antigen docking is presented in the Supporting Information.

SnugDock is an antibody- and antigen-specific extension of the RosettaDock protocol that is especially useful for docking homology modeling-derived antibody structures. SnugDock incorporates antibody-specific moves to overcome limitations of homology model-based inaccuracy in rigid-body docking that were observed in docking challenges.[4,33] Specifically, SnugDock adds a refinement step for HCDR2 and HCDR3 loops after low-resolution docking, allowing for greater loop backbone sampling with small, shear, and CCD moves. During the high-resolution phase, explicit sampling of the rigid-body VH–VL orientation and HCDR2 and HCDR3 conformations is achieved by CDR minimization, and loop backbone perturbation accompanied by additional small, shear, or CCD moves. SnugDock also can be combined with EnsembleDock, providing a database of input models for a higher diversity of starting structures.[4,26,33] SnugDock (together with EnsembleDock) has been benchmarked on a set of 11 antibody–antigen complex structures, resulting in four medium and seven acceptable ratings using the critical assessment of prediction of interactions (CAPRI) criteria.[33] SnugDock performed significantly better than did the standard RosettaDock protocol. However, SnugDock can also overfit, closing voids and constructing unnaturally tight interfaces.[33] A protocol capture for SnugDock has been published by Weitzner et al.[4]

Generally, a docking approach will greatly benefit from including experimentally obtained restraints, which can be used to limit the conformational space to relevant structures. Examples of such experimentally derived restraints are alanine or site-directed mutagenesis, hydrogen–deuterium exchange mass spectrometry (HDX) or also HDX-NMR, NMR chemical shift perturbations, low-resolution cryo-EM, and chemical cross-linking data.[99,100] In the presence of a low-resolution EM map, however, it can be very difficult to dock an antibody in the right orientation, and a combination of structural methods may be necessary to obtain a high-confidence antibody–antigen complex model.[100] Both SnugDock and RosettaDock are compatible with a wide variety of general constraints and filters in Rosetta. The general performance of a docking attempt can be assessed by calculating the interface energy for the created models, and also the C_α RMSD, for example, to the best scoring model. In many cases, some kind of experimental or knowledge-derived restraints are available that can also guide model selection, either manually or using filters in Rosetta. As docking normally has a high number of degrees of freedom, it is advisable to sample a high number of models when performing production runs for thoroughly sampling the conformational landscape (e.g., 10000, depending on the complexity of the problem).

Antibody Design. Where structure prediction seeks to identify the optimal three-dimensional protein fold for a particular one-dimensional amino acid sequence, protein design seeks to find potential amino acid sequences that can maintain at least one previously determined, stable three-dimensional protein structure. Therefore, in contrast to antibody structure prediction and docking, where an antibody of fixed sequence is considered, antibody design modifies the sequence of an antibody to improve antibody affinity, specificity, and breadth, guided by knowledge-based sampling strategies.

Single-State Design. Single-state design protocols focus on the optimization of the binding affinity of a single antibody to a specific antigen. Such an approach can be used either to improve an already existing interaction or to create a new interaction for a nonbinding antibody–antigen pair. This refinement of an antibody sequence can be seen as a computational analogy to the natural affinity maturation process.[101] Somatic hypermutation introduces changes in sequence in the highly variable

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825–846

CDR regions during clonal expansion, leading to a high adaption to the presented antigen and to the expression of the tightest binder in a plasma cell. Rosetta on the contrary samples random mutations, using its energy function and Monte Carlo sampling to differentiate between beneficial and destabilizing mutations. While such a design process can proceed naïvely, naturally occurring patterns can be used as knowledge-based restraints to restrict the sequence search space.

Sequence design in the presence of an antigen can be performed by a very basic design algorithm in Rosetta, focusing the design to amino acids within the antibody−antigen interface. An example for this procedure is given in Bender et al.[15] First, a Python script is used to identify residues that are within a distance of specified residues that define the antibody−antigen protein interface. Subsequently, these interface residues are listed in a so-called "resfile", or a space-delimited file that designates designable residues, labeled by their residue number and chain identification, and to what entities, e.g., amino acid side chains, each residue may be designed. In essence, the resfile controls which residue side chains can be mutated through design, repositioned through repacking, or kept rigid during design. Because interface design includes more than one protein, it is important to consider which side of the interface should be "mutated"; typically, it is desired to optimize the binding interface of the antibody through design while maintaining the antigen-binding interface. Therefore, it is most common to specify the residues within the antibody's interface as designable residues, while the antigen interface residues are limited to repacking to accommodate amino acid changes in the interface (Figure 4A). Example files and scripts are included in the Supporting Information.

The Rosetta design protocol optimizes the sequence on the basis of the overall energy of the complex, including the internal energy of the antibody and antigen, rather than the binding energy specifically. The resulting binding energy can be evaluated afterward by using InterfaceAnalyzer. Ideally, the binding affinity increases or decreases in value, while the overall energy (as a measurement for stability) does remain relatively constant. These criteria provide an initial filter to select models for further evaluation. More rigorous analysis, however, should evaluate each proposed mutation independently for its contribution to the total energy and binding energy in relation to the native model. A notable application using a similar protocol and analysis was the redesign of PG9, a human monoclonal antibody targeting the HIV envelope glycoprotein, where a RosettaDesign variant displayed increased potency and neutralization breadth.[102]

This method is generally applicable to protein−protein interactions, and as such, it does not use any information about the natural sequence profiles for antibodies. Furthermore, its ability to sample backbone conformations is limited, which in turn limits accurate prediction of residues critical for forming antibody−antigen interaction. To circumvent such a limitation, it may be advisable to run the protocol on an ensemble of pregenerated starting conformations, or to integrate a backrub step,[103] which will introduce greater backbone conformational flexibility.

RosettaAntibodyDesign (RAbD). RosettaAntibodyDesign (RAbD) is capable of both *de novo* antibody design from a nonbinding antibody and also affinity maturation of an already existing antibody. It classifies the antibody into regions, including framework, the five canonical loops, and the HCDR3 loop, similar to the methodology in RosettaAntibody. Additionally, it can also redesign the DE loop, or H/LCDR4, as reported by Lehmann et al. for anti-EGFR scFv antibodies.[104] RAbD starts from an assembled antibody−antigen complex and allows for both sequence and graft design based on the canonical clusters described by North et al.:[47] GraftDesign exchanges a whole CDR for another from the canonical cluster database, and SequenceDesign optimizes the sequence on the basis of the canonical cluster sequence profiles. The protocol is highly tunable by using a CDR instruction file, which allows users to include and exclude clusters, loop length, or PDB entries on the basis of the user's preferences. An example for this can be found in the tutorial section in the Supporting Information. Briefly, RAbD consists of an outer loop, which performs the graft design if enabled, and then passes the structure to an inner loop of sequence design, side chain repacking, CDR minimization, and optional integrated docking with epitope and paratope constraints. The structure is energy-minimized through the use of cluster-based CDR dihedral constraints and uses the Metropolis Monte Carlo criterion in the inner and outer loop for optimization. The default cycle number is set to 25 outer loops and one inner loop. The RAbD Metropolis Monte Carlo criterion can be set to the total energy (the protein stability score) or can be set to look specifically at the interface energy (corresponding to the computational binding affinity) using the integrated InterfaceAnalyzer methodology, as described above.[34]

RAbD therefore samples through all experimentally observed antibody conformations of different lengths and their corresponding sequence and structure space, allowing the design of loops with different lengths if desired. The protocol was benchmarked on a set of ∼60 antibody−antigen complex structures and tested in two experimental antibody design cases, where it improved binding affinities for both antibodies. A protocol example for RAbD is included in the Supporting Information.

AbDesign. AbDesign relies on backbone fragment recombination from experimental structures of antibodies deposited in the PDB, mimicking V(D)J recombination and allowing more native-like packing between the heavy and light chain than other antibody design protocols.[35,105] In short, AbDesign first predicts candidate apo structures of an antibody and, then following antibody docking, optimizes the antibody-binding interface against the target antigen. Like AbPredict, each heavy and light chain is segmented into one segment containing the CDR1, CDR2, and framework region (resembling the part of the protein encoded by the V gene) and another containing CDR3. Conformational representatives of each of the four segments are selected from precomputed Rosetta databases containing backbone segment torsion and sequence profiles. The selected segments are inserted, or grafted, onto the template scaffold by being subjected to CCD moves[63] using dihedral and coordinate constraints. Afterward, the individual antibody segments are scored against the original segment, and if the difference is <1 Å across all segments, the predicted antibody model is kept for design. In addition, the antibody sequence is optimized on the basis of conformation-dependent position-specific scoring matrices (PSSMs) for each segment cluster, thereby combining knowledge-based sequence space with backbone plasticity. Following sequence and backbone optimization, the pool of generated models is docked onto the target antigen using low-resolution docking. This is followed by a last design step. It is important to note that the sequence constraint is less strict for residues in the antigen interface, thereby encouraging a high

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

degree of sequence variability for the optimization of the binding energy upon design, whereas the more conserved framework regions have stricter constraints, to encourage selection of naturally occurring sequences to maintain stability. Using a many-valued fuzzy-logic approach[106] in the final selection, antibodies are chosen on the basis of stability (total energy), binding energy, buried surface area, packing between the heavy and light chain,[107] and shape complementarity[108] between the antibody and antigen.[35]

AbDesign was benchmarked on a set of nine antibody−antigen complexes and evaluated on sequence recapitulation and interface side chain rigidity.[35] Furthermore, AbDesign was used for two *de novo* designs of scFv in combination with yeast display and error-prone PCR in five consecutive cycles over which the protocol was adapted to its final version. Major modifications were necessary, however, because the first designs expressed poorly, which was attributed to cavities in designs, unpaired buried charges, and the loss of long-range hydrogen bonds.[105] Even with the two successfully predicted scFv models, crystallization of the models as Fabs (notably without antigen) revealed structural differences between the experimentally determined models and the AbDesign models, especially in HCDR3 and HCDR1.[105]

Design with Noncanonical Amino Acids. Rosetta not only can design with the 20 genetically encoded amino acids but also can incorporate noncanonical amino acids, such as L-3,4-dihydroxyphenylalanine (L-DOPA).[23] A study by Xu et al. tested Rosetta's ability to predict the placement of L-DOPA correctly for chemical cross-linking reactions between the antibody and antigen with nucleophilic amino acids, like lysine or histidine, using side chain packing on a fixed backbone and minimization. Optimization of the estimated cross-linking distance and solvent accessibility of L-DOPA, while maintaining the binding and total energy of the antibody−antigen complex, provided the best strategy for the successful design of antibody−antigen complexes that could be cross-linked.[109] Rosetta contains a number of noncanonical amino acids that can also be used for design cases, along with protocols to incorporate those noncanonical amino acids.[23,110]

Design of Supercharged Single-Chain Variable Fragments (scFv's). The rational design of antibody surfaces can enhance stability and improve physicochemical properties in solution. Miklos et al. introduced charged amino acids into the surface of the anti-MS2 scFv to enhance solubility and weaken its aggregation tendency,[111] problems to which scFvs are prone.[112] To achieve incorporation of charged residues, fixed backbone design was undertaken on all residues excluding any CDRs, by providing a resfile to designate the placement of charged residues. Placement of charged residue side chains was favored by slowly incrementing the reference energies for all charged amino acids. Through this process, a scFv design showed enhanced stability against heating, solubility, and increased binding affinities.[111]

Balancing between Sampling and Stability. The protocols presented above represent multiple options to design antibodies in Rosetta. The optimal choice of protocol depends on the design task. The more changes are made to the native antibody that initially is expressible and capable of being crystallized, the less likely are the designs to be expressed and stable.[5,105] Like in affinity maturation, however, it is often necessary during protein design to sample a broad sequence and conformation space to identify the optimal combination of antibody sequence and structure to achieve both high specificity and binding affinity for a target antigen. This can require sampling beyond energy barriers that confine the native antibody's sequence and structure space to a local energy minimum, and in such cases, protocols that provide a means for more extensive sampling may be superior to more conservative approaches that limit the sampling space to a local energy minimum. In general, if the goal is to improve the binding affinity of an antibody within an already determined antibody−antigen complex, it is generally advisable to begin with a more conservative approach. Otherwise, it is often a good idea to use more than one protocol and to compare results for convergence onto the same sampling space. Even after cross-checking multiple approaches, it may be necessary to alter the chosen protocol to account for problems like expressibility or solubility. However, to overcome energy maxima in the conformational landscape, it might be necessary to sample more thoroughly, and in these cases, protocols with more sampling can be superior compared to more conservative approaches. Upon comparison or establishment of protocols, a smaller size of models can be sampled and evaluated for chosen parameters, which could include the interface energy as metric for predicted binding affinity, but also sequence similarity, type of newly created interactions, or other knowledge-derived metrics depending on the complexity and the specific questions of the design task. The number of models that should be created for a design task can vary quite heavily depending on the number of positions to design and the protocol used. Generally, more output models will be needed for less conservative approaches.

Multistate Design (MSD). While single-state design considers just a single antibody or antibody−antigen structure, MSD protocols provide a wide platform for addressing several types of higher-complexity design problems. Most commonly, MSD encompasses the design of one antibody in the presence of more than one antigen. The goal can be to optimize the breadth of the antibody to bind multiple antigens, find an antibody that can bind to multiple conformations of a single antigen, or optimize the selectivity of the antibody through negative design against a subgroup of antigens.

For all three of these possibilities, protocols have been developed in Rosetta and used in the field of antibody design.

Broadly neutralizing antibodies (bnAbs) have proven to be a powerful therapeutic tool. A highly optimized antibody is at risk of losing its binding affinity when small changes in the antigen's amino acid composition occur, whereas bnAb maintains its ability to bind to antigens from multiple strains, subtypes, or even species. The bnAb therefore is more likely to provide protection for a longer period of time. Such breadth is normally mediated through limited but tight binding to conserved residues that are functionally less susceptible to antigenic drift.

One classical MSD task includes designing an antibody initially known to bind to a single antigen to optimize its sequence to form multiple novel binding interactions with one or more antigens. The Rosetta MSD design protocol using the REstrained CONvergence (RECON) algorithm was originally developed to perform such a task to increase antibody breadth by constraining the sampled sequence space to adopt multiple (binding) conformations.[5,6] Broad antigen recognition, or polyspecificity, may be linked to germline antibody sequences; it has been hypothesized that naïve germline antibodies exhibit greater conformational flexibility, which enables polyspecificity.[113] Interestingly, using RECON MSD to design the sequence space of a single antibody when in complex with a set of antigens reverted an antibody's sequence back toward its germline gene sequence (Figure 5A). Conversely, using single-state design-

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

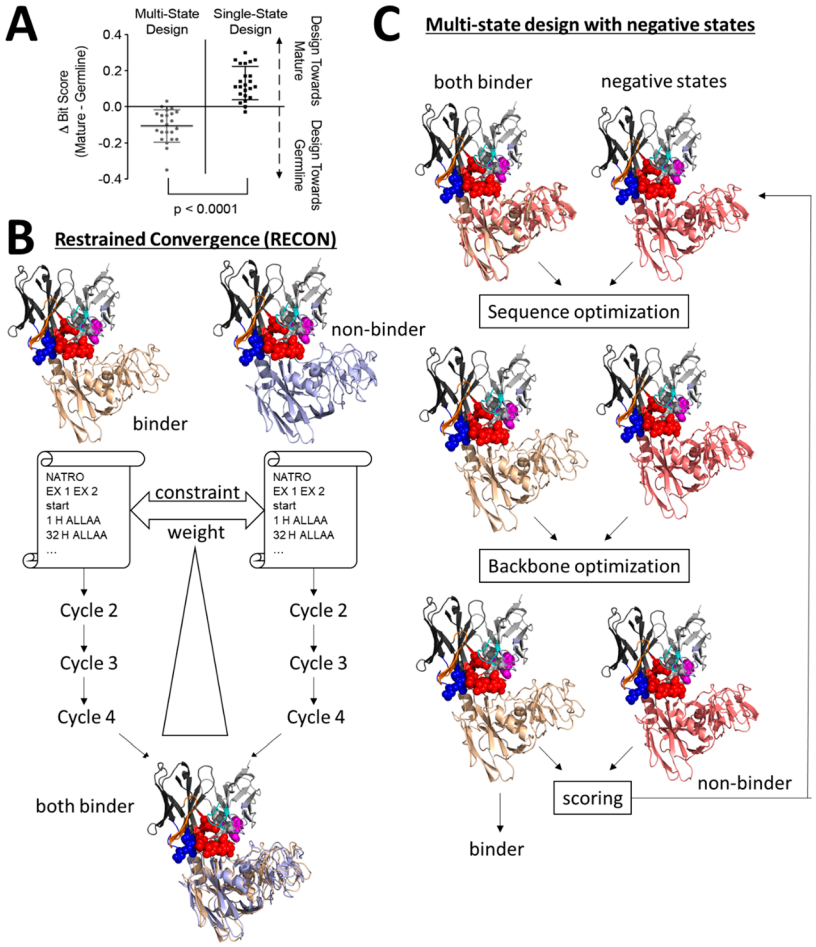


Figure 5. Overview of multistate design protocols in Rosetta. (A) Multistate design reverts the antibody sequence back to the germline sequence, while single-state design approximates affinity maturation (figure reproduced under CC-BY from ref 101). (B) The RECON protocol (REstrained CONvergance) expedites discovery of states that bind multiple targets faster than traditional MSD algorithms because of its independent search of sequence space. (C) Design with negative states performs selectivity design against binders.

introduced mutations will make the difference from the germline gene sequence greater.[101]

Design of polyspecificity requires that the antibody of interest be spatially aligned with all antigens for which a common binding motif should be found, which comprise the antibody's intended targets, and that a common antibody-binding interface be the subject of design. For RECON MSD, the antibody interface of interest is based on a known antibody−antigen complex structure, such that any novel binding interfaces are based on the superimposition of target antigens to the known antibody−antigen complex. RECON MSD is novel with respect to other MSD protocols in that rather than treating design as a combinatorial problem, it reduces the design of a large conformation space by treating each structure, or state, included in the design as a separate design problem, thus making RECON MSD very efficient. More specifically, design sampling identifies the lowest-relative free energy sequence for each single conformation but will accept a redesigned sequence only if the sequence has the lowest average energy across all states. RECON MSD assumes that the native sequence is close to the sequence that is ideal for conformational flexibility or polyspecificity and encourages the selection by using a convergence restraint to favor the selection of native sequences. Convergence is further encouraged by using multiple rounds (typically four rounds) of design. To converge on a common sequence, a sequence similarity restraint is introduced. The restraint is kept small in early rounds of design to sample a broad sequence and conformational space specific to each antigen and ramped up in later rounds of design to find convergence over multiple antigens. In the case in which selection of a sequence does not converge for a designed position, the last step in the protocol forces a selection based on the lowest fitness over all sampled amino acids for nonconverging positions. In the end, this sequence convergence encouraged through restraints is hypothesized to find minima in the energy landscape more rapidly (Figure 5B). The independent sequence search allows trajectories to adopt sequences that are favorable in one state but might not be in another state, which in contrast to classic MSD algorithms prevents the exclusion of these intermediate states. Thus, the encouraged convergence bypasses high-energy states. RECON was benchmarked in comparison to the traditional Rosetta MSD, where it showed improved performance to recapitulate evolutionary sequence profiles, a metric chosen to represent polyspecificity.[6] RECON was further refactored to run in parallel on separate processors using message passing interface (MPI) communication, which enables massive parallel design against a large number of antigens.[5] It was applied to design broad influenza hemagglutinin H1 antibodies based on the C05−H3 complex structure[114] and could propose mutations that showed an enhanced breadth against additional virus

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

strains, including a strain with a known escape mutation.[5] In this work, criteria that yield greater success in design were identified. For example, a high drop of energy for some antigens, especially the antigen that is bound by the antibody in the original complex structure, indicates nonfavorable mutations.[5] Mutations that establish new hydrogen bonds, relieve clashes with the antigen, or create more van der Waals interactions are favorable. To increase the sampling space, the protocol can be combined with backrub moves, which creates a backbone ensemble and enables the sampling of a larger sequence space.[6] Generally, the evaluation criteria are similar to a single design task, however, considering only such amino acid changes that improve predicted binding affinity (e.g., interface energy) for all multistate design targets while not compromising protein stability (total_score). An example protocol for multistate design with the RECON design protocol can be found in the Supporting Information.

The BROAD (BReadth Optimization for Antibody Design) algorithm has been developed to enhance MSD performance further than RECON MSD. The RECON protocol becomes computationally expensive when designing antibodies against large panels of antigens, or many different conformations of a protein. BROAD includes support-vector machines to classify antibody binders versus nonbinders and optimizes breadth through the use of integer linear programming. This method is very fast and can be applied to large sets of antigens (e.g., a large panel of different viral strains). The method has been tested computationally, but the protocol has not yet been applied to an experimental application.[115]

Multistate Design for Negative Design Tasks. While both RECON and BROAD are highly optimized protocols for the design of broadly binding antibodies, the generic MSD protocol in Rosetta can be used for a broader range of applications.[116] Here, we illustrate the use of the MSD protocol with an example of negative state design for the generation of bispecific antibodies.

Bispecific antibodies carry two different variable domains that simultaneously recognize two different antigens, which can be used in applications such as cancer and immunotherapy.[117] However, the production of bispecific antibodies can suffer from large losses due to undesired heavy and light chain pairings; avoiding this problem usually requires extensive protein engineering. Using Rosetta's MSD protocol, multiple IgG constant regions have been remodeled to exclusively bind to the heterologous heavy chain and prevent light chain mispairing by incorporating specific light chain−heavy chain interfaces.[8,36,118] This requires the interface redesign of the constant regions, both between the light chain and heavy chain and between heterologous heavy chains, resulting in designs that when co-expressed would form monomeric bispecific antibodies without unwanted byproducts. Throughout a range of design rounds, C_H1/C_λ, $C_H1/C\kappa$, and C_H3 heavy chain constant domains were designed by applying MSD with negative design states. As Leaver-Fay et al. described in depth, the design of nonbinders requires a larger amount of redocking and repacking due to the fact that a designed amino acid might clash in the original proteins state, but upon redocking and repacking, the protein accommodates the introduced change.[36,116] As repacking is one of the most computationally expensive steps in Rosetta protocols, Leaver-Fay et al. introduced the idea of running multiple rounds of MSD, in which all conformations of the negative state (nonbinding conformations) are collected and added to a set of states that will be used in a further round of MSD.[36] With this growing ensemble of negative states, Rosetta's design algorithm will more likely find mutations that will not be overcome by conformational changes.[36] These designs were also experimentally tested on known antibodies, with a high success rate of 93% in the case of bispecific antibody formation.[36] A schematic overview of MSD using negative states can be found in Figure 5C.

Computational Design of Non-Immunoglobin Epitope Binders. The interactions determined by antibody−antigen complex structures can be used as a blueprint for the design of non-antibody protein binders. This process can be seen as a parallel approach to antibody design, transplanting the binding paratope onto a non-immunoglobulin domain. Illustrative examples include two efforts targeting influenza hemagglutinin (HA) binding, one to the stem region and one to the head region.

The HA stem-binding proteins were based on the interactions of the two broadly neutralizing antibodies, CR6261 and F10, known to bind to all influenza A group 1 HAs.[119,120] The interacting residues were treated as "hot spots", and scaffold proteins that supported the positioning of the hot spot residues were determined with a shape-complementation search. Docking and minimization steps followed by rounds of design maintaining the key hot spot interactions led to a number of unique designs that were tested experimentally. The protocol yielded two binding proteins that underwent directed evolution to cause the maturation of their affinity.[121,122]

The HA head binder was designed on the basis of the interaction profile of the broadly neutralizing C05 antibody.[114] C05 maintains a well-characterized binding mode that requires an extensive hydrogen backbone network in its HCDR3 loop. On the basis of HCDR3 residues 7−15, a backbone similarity search against a large scaffold library resulted in a number of hits that could be subsequently optimized through sequence design and shape complementarity optimization. Yeast surface display was used to identify successfully binding proteins with a hit rate of 5 of 80. As HA naturally occurs as a trimer, the monomeric head domain binders were linked by a trimeric scaffold protein to result in a cap for the HA trimer. This trimeric construct showed viral neutralization activity and was also tested in a mouse model for protection against influenza virus challenge.[123]

The approaches of epitope-focused design (see below) can also be inverted for general protein binder design, where it is the paratope (rather than the epitope) that is transferred into a new scaffold. This can potentially allow whole loops to be placed into a non-immunoglobin context.

In contrast to redesigning immunoglobins, these paratope grafting approaches require more extensive sampling and experimental testing. Such grafting approaches (both paratope and epitope) require not only computational throughput but also experimental screening for correctly folded proteins. The chance that an antibody with redesigned interface residues expresses, folds, and binds will likely be far greater than for a complete *de novo*-designed protein.

An approach similar to that described above was undertaken by Sevy et al., using the HCDR3 loop of the broadly neutralizing antibody C05, which targets the HA receptor-binding site.[124] The loop was redesigned by cyclization with a disulfide bond and fixed backbone design to stabilize the peptide. Designs were controlled with folding experiments on the obtained sequences, yielding a selection of peptides with favorable energy scores. While the wild-type peptide was not able bind to HA, two of

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

eight redesigned cyclized forms bound to the C05-binding site.[124]

Methods for Antigen Design. Rational vaccine design has changed the field of vaccinology drastically in recent years, and new methods for facilitating targeted vaccine design have been introduced. With new concepts such as epitope-focused immunogen design, lineage-based immunogen design, and germline targeting, vaccinology has been reinvented with the aim of enabling vaccine design for diseases for which the traditional method failed to provide long-term immune protection.[125] These same methods can also be used to design probes for immunization studies and antibody repertoire characterization.

Epitope-Focused Immunogen Design with Rosetta. It is not always possible to raise an immune response against one particular epitope in the presence of other epitopes. The concept of the immunodominance of certain epitopes over others and its implications for the escape of viruses from antibody recognition have raised the idea of presenting immune subdominant epitopes to trigger the desired antibody responses, particularly in the fields of HIV and influenza vaccine research, where standard vaccine development strategies fail to induce immune responses with long-term protection.[126−128]

Epitope-focused approaches include all methods that enable the presentation of only the epitope of interest to enhance targeted antibody responses. The classic epitope transplantation for antibody focusing was first described by Ofek et al. for the HIV gp41 2F5 epitope,[128] by Correia et al. for the HIV 4E10 epitope,[129] and by McLellan et al., who showed the successful transplantation of the motavizumab epitope of respiratory syncytial virus (RSV) glycoprotein onto a smaller scaffold protein, which could raise an immune response in mice, but no neutralizing activity.[12] In all cases, Rosetta was the software of choice for grafting epitopes onto scaffold proteins. Today, multiple design strategies are available in Rosetta. Here, we will discuss these methods, compare their strengths, and define design problems for which they are most suitable.

The basic idea behind classic epitope focusing is the transplantation of an epitope onto a smaller scaffold protein. All available protocols are today executed via RosettaScripts and therefore can be easily adapted to the user's requirements. We will here refer to the epitope as the "motif", the antibody-interacting residues of the motif as "hot spots", the acceptor protein of the graft as the "scaffold", and the motif-binding proteins (including nAbs) as the "context" or "binder", which is the terminology that is used throughout Rosetta's grafting protocols.[39]

Side chain grafting and backbone grafting both use the same Mover, called `MotifGraftMover`, which needs three inputs: (1) the motif (epitope), which specifies the residues to be grafted, (2) the context, which is the protein (antibody) binder, and (3) a list of scaffolds, onto which the motif will be inserted. Furthermore, so-called hot spots can be defined, which are a user-based selection of key residues interacting with the protein binder. During the sampling process, the protocol will alter the surrounding sequence to adjust for the incorporation of the motif but maintain the identity of the hot spot residues, which are crucial for binding.[39]

Scaffold library selection is one of the most important considerations in the overall protocol. While the MotifGraft-Mover can be applied over a large number of prepared scaffolds, in many cases it is desirable to prefilter the scaffold set to minimize sampling time and increase scaffold quality. Different methods for filtering scaffolds exist, including MASTER, a tertiary RMSD-based search algorithm,[130] or a shape complementarity filter within Rosetta.[121] Overall, the most important features for selected scaffolds include the high resolution (<2.5 Å) of the experimental structure, which raises confidence in the backbone and side chain atom placement, the lack of ligands, and the requirement for a single-chain protein. Depending on the preferences of the users, the ability to express proteins in a desired system can also be considered.[39]

Side chain grafting transplants only the hot spot residues that are crucial for the interaction onto the scaffold protein. In contrast, backbone grafting transplants a whole peptide or peptide fragments, while maintaining their backbone geometry. To guarantee the mimicry of the motif, side chain grafting requires that the motif and scaffold backbone be superimposed with high accuracy, as the scaffold backbone will not be changed. This, therefore, puts stringent limits on the suitability of scaffold proteins (normally an RMSD of <0.5 Å between the motif and scaffold backbone segment). In backbone grafting, the backbone conformation of the motif will be transferred to the scaffold protein, which allows the use of a wider range of scaffolds. However, the alignment in the N- and C-terminal regions of the motifs with those from the scaffold has to match closely, because any breaks in the geometry of the protein chain can reduce protein stability and the robustness of the design protocol.

A comprehensive manual about side chain and backbone grafting can be found in ref 39, which overlaps to a large extent with the example protocols presented in the Supporting Information.

Both side chain and backbone grafting have been applied successfully in multiple cases. Side chain grafting was used to transplant the 4E10 epitope from the HIV viral glycoprotein, which resulted in high-affinity binders.[129] For the motavizumab epitope located on the prefusion RSV F protein, a side chain graft was performed onto a discontinuous region on a scaffold protein (termed a multisegment side chain graft), resulting in a number of motavizumab binders. However, the immunogens themselves failed to elicit neutralizing antibody titers.[12] A first linear backbone graft in combination with side chain grafting was performed by using the 4E10 and 2F5 epitopes located on the HIV glycoprotein. The backbone-grafted designs bound with higher potency to the respective antibodies.[10] A discontinuous backbone grafting approach has been used by Azoitei et al. to graft the HIV gp120 epitope of cross-neutralizing antibody b12 onto a smaller scaffold protein. Determining the structure of the design by X-ray crystallography confirmed that the design could recapitulate the interactions seen between the natural epitope and the antibody. The `MultiGraftMatch` algorithm used in earlier studies has been updated and is now available as backbone grafting based on `MotifGraft-Mover`.[37] Another use case of side chain grafting has been shown in ref 131, in which inhibitors for the Epstein-Barr viral Bcl-2 protein were designed for a helical epitope.

The new protocol FunFolDes (Functional Folding and Design), which succeeds the FoldFromLoops protocol by Correia et al.,[11] tries to overcome scaffold limitations by grafting the motif onto a protein fold instead of a specific scaffold protein. The fold of the scaffold protein is represented as a topology defined by distance, angle, and dihedral constraints. Instead of a mere motif transplantation, the motif is transferred into the scaffold topology and the remaining protein is folded into the defined topology. This combined grafting and *ab initio* folding is implemented by the `NubInitioMover`, (nub refers to the

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

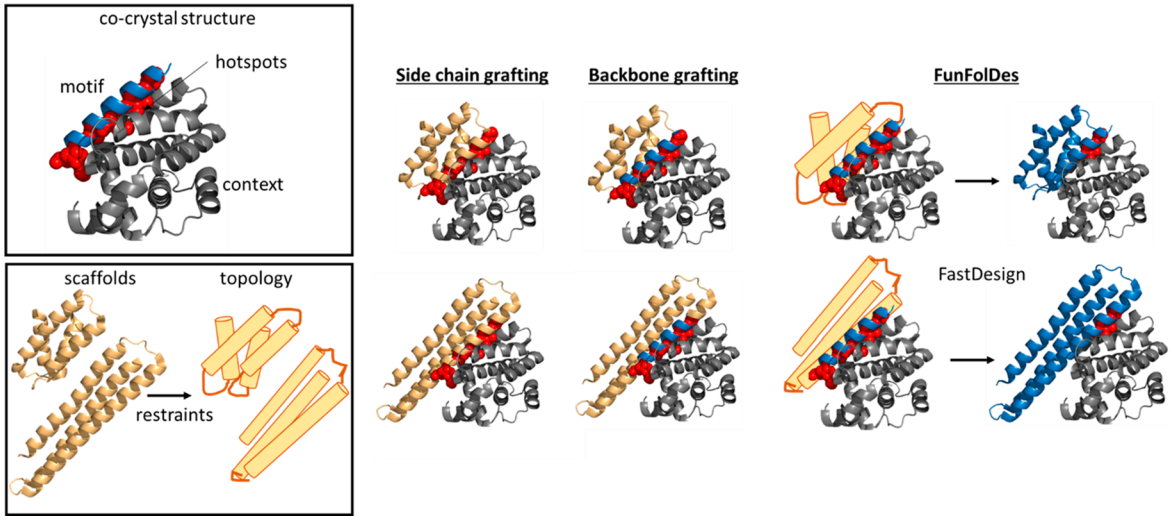


Figure 6. Overview of the different grafting protocols in Rosetta. The nomenclature commonly used in grafting protocols defines the binder as the context, the epitope as the motif, and the interface residues as hot spots. Side chain grafting transplants only the hot spots onto a scaffold, while backbone grafting transfers the motif backbone with hot spots. FunFolDes (Functional Folding and Design) abstracts the scaffold topology using distance restraints and refolds the protein from the motif into the topology, followed by a design step (PDB entries 2WH6,[132] 3LHP,[129] and 3FBL[133]).

motif as the known peptide). Subsequently, the protein is redesigned using FastDesign. The major benefit of FunFolDes is that the scaffold can compensate for the motif insertion by adapting the protein backbone. FunFolDes can furthermore be used for loops, which notoriously have low numbers of suitable templates from the PDB. The FunFolDes protocol is implemented in RosettaScripts and can be combined with other design tools or filters.[40] The new implementation has been benchmarked computationally on 14 known design cases and focused on two cases, including a redesign of a BHRF1 binder,[131] for which the previous version of the protocol showed large numbers of unsuitable designs. FunFolDes performed significantly better on this target than its previous implementation, largely through restricting the search space to more native-like folds and good interface energies.[13,40] To address a major problem in *de novo* design, the FunFolDes protocol was used to fold and design custom-tailored protein backbones to structurally irregular epitopes of RSV, which showed the ability to elicit neutralizing antibodies *in vivo*.[13] FunFolDes is a powerful tool and not only can be employed to design immunogens but also represents a universal tool for motif-guided protein design, and a protocol capture can be found in the Supporting Information.

One of the greatest challenges in epitope-focused immunogen design is identifying suitable scaffold proteins. Although the number of proteins in the PDB is constantly growing, it might not be possible to find a scaffold with the correct topology to support complex epitopes. Additionally, protein engineering techniques in Rosetta can be applied to further improve scaffold thermostability and to reduce its off-target immune stimulation. Correia et al. showed how the HIV gp41-specific 4E10 epitope immunogen could be optimized into a well-behaved protein that displayed the desired topology. The major techniques that were applied include the removal of an unnecessary domain, thermostabilization using Rosetta Remodel, and extensive resurfacing through Rosetta Design.[38]

The Rosetta protocols in the Supporting Information include a section on Rosetta Remodel, a very diverse and powerful application in Rosetta. Remodel can be used to extend and shrink protein C- and N-termini, insert and delete domains, form disulfides, construct new backbone topologies based on secondary structure, and many more. Rosetta Remodel relies on a so-called blueprint file that defines residue-specific secondary structure and amino acid identities. The sampling relies heavily on fragment insertion guided by the blueprint specification.[41]

Another option to create new scaffolds is represented by the structure extension with native substructure graphs (SEWING) method. SEWING uses connected or disconnected pieces of experimentally determined structures and recombines them for protein design. In contrast to more classic design strategies, which tend to result in idealized protein folds, SEWING successfully samples proteins with higher backbone diversity.[42] Protocols for the use of SEWING can be found in ref 43.

For many immunogen design tasks, a number of protocols could be used. In general, it is wise to proceed with the most conservative approach. With fewer perturbations to the scaffold protein structure, the chances that the designed protein will adopt the predicted fold and structure improve. Nevertheless, when an immunogen design project is started, all available protocols should be assessed for their suitability to solve the task. FunFolDes is especially suitable for complex epitopes and loops, while side chain and backbone grafting should be employed for grafting problems with well-defined secondary structure. Depending on the complexity of the design task, it will be necessary to sample increased numbers of models, and also to conduct multiple optimization rounds until converging onto a reasonable solution. The number of models necessary can be tens of thousands. Different metrics can be used to assess the designs, including a number of filters that are provided within Rosetta, such as cavity and packing filters. All three methods are available as tutorials in the Supporting Information. A side-by-side comparison can be viewed in Figure 6.

Vaccine Design through Thermostabilization. A major challenge of vaccine design is the flexibility and instability of immunogenic proteins. For now, computational generation of novel epitope-presenting proteins, such as with the methods described above, requires several rounds of testing and optimization both computationally and experimentally. Another approach is to simply stabilize a protein of interest.[44] In the latter, the needs of thermostabilizing a protein structure and

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

maintaining its function were achieved through amino acid changes guided by information about the protein sequence's evolutionary diversity. The rationale is that evolution does not allow for destabilizing mutations as those would render the protein inactive. The protocol represents the evolutionary diversity with a PSSM, which it uses to sample possible mutations. The effect of a mutation on the stability of the protein and its interactions are evaluated by a $\Delta\Delta G$ calculation in Rosetta. Stabilization is achieved by a combinatorial search of groups of amino acid changes that can have an additive effect on protein stability. This protocol was benchmarked on multiple proteins to predict known stabilizing mutations without choosing known destabilizing mutations.[44] Additionally, the protocol was tested for thermostabilization of human acetylcholinesterase (hAChE), which is usually expressed in eukaryotic cells and could be obtained in large amounts in *Escherichia coli* expression. Of the five chosen designs that had 17−67 mutations in total, four maintained activity while having higher deactivation temperatures.[44] The protocol is available as a Web server, called the Protein Repair One Stop Shop (PROSS, http://pross.weizmann.ac.il).

This approach also has been applied to a vaccine design project, namely, the thermostabilization of *Plasmodium falciparum* reticulocyte-binding protein homologue 5, a relevant target for malaria vaccine development. In total, 18 mutations were introduced and yielded a design that was expressed in *E. coli* and showed higher stability, while maintaining its immunogenicity. An experimentally determined structure proved that the design was very similar to the original protein.[134]

Designing and Refining Glycans in Rosetta. Viral proteins are often covered with glycans to circumvent immune recognition. Use of glycan masking to focus immune system attention represents an alternative approach to motif transplantation techniques. Additional glycosylation sites are introduced on the protein surface, which cover immunodominant sides on the immunogen, thus leading to the focused recognition of the epitope of interest. This method has been applied to focus the immune response to the CD4-binding site on the HIV gp120 immunogen eOD-GT8 by introducing five additional glycans that covered sites of undesired immune recognition,[127] and also to shield the nonspecific immune response toward the *Plasmodium vivax* Duffy-binding protein.[135] However, upon introduction of additional glycosylation sites, the protein could be destabilized or new glycans could interfere with other already existing glycan patches.

Due to their flexible nature, glycans are rarely solved in experimental structures. When glycans are found in experimental density, it is often not the natural glycan, as the glycan identity and linkage patterns are species-dependent and heterogeneous and therefore derived from the expression system rather than the original host cell. Sometimes, glycosylation sites are even removed before running crystallization trials to increase protein homogeneity. With the resolution revolution in cryo-EM, more structures with larger glycan patches are becoming available, and a need to integrate glycan modeling, structure prediction, and design became obvious.

Earlier efforts using a combination of Rosetta scoring and external preparation scripts were successful in modeling glycoproteins, such as the ternary complex of HIV gp120 reported by Pancera et al.[136] The RosettaCarbohydrate framework, however, represents an integration of all of the basic tools to read, score, sample, and process glycans directly within Rosetta. Critical to this framework was updating the core Rosetta code to support glycan branching, nomenclature, and stereochemistry, as well as the attachment of glycans to the protein side chains. Glycans can have branched structures, which increases combinatorial complexity, different stereochemical properties, the ability to be substituted, and a wide range of possible interaction partners. Information about the glycans can be found in Rosetta's CarbohydrateInfo object, which stores the identity, stereochemistry, and functional properties of the respective saccharide. Recent improvements in the PDB have addressed the lack of standardized glycan nomenclature,[137] and Rosetta can read most glycans in PDB files and automatically detect and score them. It is also able to work with some of the most common file formats in saccharide computational science, such as GLYCAM and GlycoWorkbench (.gws). Rosetta scoring was updated to accommodate glycans by incorporating a sugar backbone term (sugar_bb). Glycans now not only can be read and scored but also can undergo processing, such as relax and backbone sampling. Some glycan-specific tools are also available, such as the glycosylation tool to insert or generate a glycan at a possible glycosylation site, loop modeling of glycans, especially to fill gaps in glycans from missing density in experimental structures, and docking of glycans primarily in a protein environment. Many of these methods are described in detail in ref 45, in which they are accompanied by case and code examples. More applications for glycan modeling are already available in Rosetta and can be found on the RosettaCommons Wiki alongside notes and tips for use (https://www.rosettacommons.org/docs/latest/application_documentation/carbohydrates/WorkingWithGlycans).

A special application of the existing RosettaCarbohydrate framework is the refinement of glycans into low-resolution cryo-EM and X-ray density maps. Most existing density refinement tools have limitations in modeling glycans, resulting in poorly modeled glycans in deposited structures. Rosetta has shown success refining protein structures in low-resolution densities,[138−140] and with the integration of the RosettaCarbohydrate framework, it became possible to both correct existing glycoprotein structures and provide a method for generating more reliable structures in the future.[141]

Due to the increasing importance of being able to work with glycosylated structures and to glyco-engineer proteins for tailored functions, we provide a tutorial on glycan modeling into a cryo-EM density in the Supporting Information. Note that this methodology has not been previously published and is based on in-development protocols. More tutorials on working with EM densities are part of ref 15.

Antigen Carrier Design. Even if an epitope-focused immunogen can produce a protein with tight binding to a desired antibody, expressed proteins based on these designs may fail to induce neutralizing antibody responses in mammalian hosts, e.g., the epitope scaffold mimicking prefusion RSV F design from McLellan et al.[12] Display of immunogens of the surface of nanoparticles is one option to enhance the immunorecognition of such immunogens. A common carrier is ferritin, a naturally occurring self-assembling nanocage for iron storage, which can be decorated with immunogens.[142] However, other self-assembling nanoparticle systems exist. In particular, a number of self-assembling particles have been created with Rosetta by redesigning small protein surfaces already forming small homomers, e.g., dimers, trimers, or pentamers, to assemble together with a heteromeric partner. A number of different particle topologies were sampled, including designs of Platonic

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

solids formed by the smaller homomers.[143−146] Even a particle that is large enough to encapsulate its own RNA has been reported.[147] These proteins behave well, are highly stable, and express in high yields from *E. coli*, which facilitates their handling.

For vaccine design, these particles have been used as carriers to display viral glycoproteins, as recently shown for a stabilized prefusion RSV F protein and an HIV envelope protein.[148,149] In both cases, the soluble part of the trimeric viral protein was docked computationally to a range of trimeric assembly units. After linker design, the viral glycoprotein fused to the carrier protein was expressed and assembled with the second component to yield particles displaying 20 copies of a viral glycoprotein.[148,149] These particles show the ability to optimally display immunogens, as in this case the viral glycoproteins.

Sesterhenn et al. used an epitope-focused immunogen, derived from RSV site II,[11] and combined it with the RSV nucleoprotein, which not only forms nanorings but also presents RSV-directed T-cell epitopes and would therefore be ideal for enhancing the immunogenicity of the RSV immunogen. They successfully established a nanoring displaying the immunogen and showed enhanced enrichment of specific antibodies for the RSV site II epitope as compared to prefusion RSV F, with greater neutralization titers.

CONCLUSION

For many years, the development of structure-based computational methods has been proceeding, resulting in a multitude of protocols in Rosetta. For an unexperienced user with only preliminary knowledge of computational structural biology, it can be quite overwhelming to choose the appropriate protocol. The Meiler lab hosts a semiannual workshop at Vanderbilt University that trains new users in using Rosetta and since 2016 has regularly held one specific to methods for antibody structure prediction and design. These workshops have made us aware of the interest in using these Rosetta methods. Therefore, we present here the collected tutorials from the 2019 workshop together with a literature review to support users in modeling and designing their own proteins of interest with the help of Rosetta.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.0c00912.

Step-by-Step tutorials (S1) and associated inputs (S2) for the following protocols as discussed in this work: antibody structure prediction, HCDR3 modeling, antibody−antigen docking, antibody affinity maturation, RosettaAntibodyDesign (RAbD), GlycanTreeModeler, and Scaffolding and epitope grafting techniques (PDF)

Additional information (ZIP)

Further information about Rosetta and the included tutorials can be found at https://www.rosettacommons.org/docs/latest/Home or on the Meiler lab homepage (http://www.meilerlab.org/index.php/rosetta-tutorials).

AUTHOR INFORMATION

Corresponding Authors

Rocco Moretti − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States;* orcid.org/0000-0003-2162-1116; Phone: +1 (615) 936-6594; Email: rocco.moretti@vanderbilt.edu

Jens Meiler − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States; Institute for Drug Discovery, University Leipzig Medical School, 04103 Leipzig, Germany*; Phone: +1 (615) 936-6594; Email: jens@meilerlab.org

Authors

Clara T. Schoeder − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*

Samuel Schmitz − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*

Jared Adolf-Bryfogle − *Department of Immunology and Microbiology and IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California 92037, United States*

Alexander M. Sevy − *Center for Structural Biology and Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37240-7917, United States; Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, Tennessee 37232-0417, United States*

Jessica A. Finn − *Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37240-7917, United States; Vanderbilt Vaccine Center and Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee 37232-0417, United States*

Marion F. Sauer − *Center for Structural Biology and Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37240-7917, United States; Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, Tennessee 37232-0417, United States*

Nina G. Bozhanova − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*

Benjamin K. Mueller − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*

Amandeep K. Sangha − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*

Jaume Bonet − *Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

Jonathan H. Sheehan − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*

Georg Kuenze − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States; Institute for Drug Discovery, University Leipzig Medical School, 04103 Leipzig, Germany*

Brennica Marlow − *Center for Structural Biology and Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37240-7917, United States*

Shannon T. Smith − *Center for Structural Biology and Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37240-7917, United States*

Hope Woods − *Center for Structural Biology and Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37240-7917, United States*

Brian J. Bender − *Center for Structural Biology and Department of Pharmacology, Vanderbilt University, Nashville, Tennessee 37240-7917, United States*
Cristina E. Martina − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*
Diego del Alamo − *Center for Structural Biology and Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee 37240-7917, United States;* orcid.org/0000-0003-1757-9971
Pranav Kodali − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*
Alican Gulsevin − *Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37212, United States*
William R. Schief − *Department of Immunology and Microbiology and IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California 92037, United States*
Bruno E. Correia − *Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*
James E. Crowe, Jr. − *Vanderbilt Vaccine Center, Department of Pathology, Microbiology, and Immunology, and Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee 37232-0417, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.0c00912

Author Contributions

The manuscript was mainly written by C.T.S. and S.S. with contributions of all authors. All authors have given approval to the final version of the manuscript. C.T.S. and S.S. contributed equally to this work.

Funding

This work was supported by a grant from the National Institutes of Health (NIH) (R01 GM073151) and RosettaCommons. Work in the Meiler laboratory is supported by the NIH (U01 AI150739 and U19 AI117905).

Notes

The authors declare the following competing financial interest(s): J.E.C. has served as a consultant for Takeda Vaccines, Sanofi-Aventis U.S., Pfizer, Novavax, Lilly, and Luna Biologics, is a member of the Scientific Advisory Boards of CompuVax and Meissa Vaccines, and is Founder of IDBiologics. The Crowe laboratory at Vanderbilt University Medical Center has received sponsored research agreements from IDBiologics. All other authors declare no conflict of interest.

■ ABBREVIATIONS

AMA, Antibody Modeling Assessment; bnAb, broadly neutralizing antibodies; CAPRI, Critical Assessment of Prediction of Interactions; CCD, cyclic coordinate descent; CDR, complementarity-determining region; FR, framework; HA, hemagglutinin; HIV, human immunodeficiency virus; L-DOPA, L-3,4-dihydroxyphenylalanine; MPI, message passing interface; MSD, multistate design; PDB, Protein Data Bank; PSSM, position-specific scoring matrix; RAbD, RosettaAntibodyDesign; RECON, restrained convergence; RMSD, root-mean-square deviation; ROSIE, Rosetta Online Server That Includes Everyone; RSV, respiratory syncytial virus; VH, variable heavy chain domain; VL, variable light chain domain.

■ REFERENCES

(1) Andreano, E., Seubert, A., and Rappuoli, R. (2019) Human monoclonal antibodies for discovery, therapy, and vaccine acceleration. *Curr. Opin. Immunol. 59*, 130−134.
(2) Castelli, M. S., McGonigle, P., and Hornby, P. J. (2019) The pharmacology and therapeutic applications of monoclonal antibodies. *Pharmacol. Res. Perspect. 7*, e00535.
(3) Lu, R. M., Hwang, Y. C., Liu, I. J., Lee, C. C., Tsai, H. Z., Li, H. J., and Wu, H. C. (2020) Development of therapeutic antibodies for the treatment of diseases. *J. Biomed. Sci. 27*, 1.
(4) Weitzner, B. D., Jeliazkov, J. R., Lyskov, S., Marze, N., Kuroda, D., Frick, R., Adolf-Bryfogle, J., Biswas, N., Dunbrack, R. L., Jr., and Gray, J. J. (2017) Modeling and docking of antibody structures with Rosetta. *Nat. Protoc. 12*, 401−416.
(5) Sevy, A. M., Wu, N. C., Gilchuk, I. M., Parrish, E. H., Burger, S., Yousif, D., Nagel, M. B. M., Schey, K. L., Wilson, I. A., Crowe, J. E., Jr., and Meiler, J. (2019) Multistate design of influenza antibodies improves affinity and breadth against seasonal viruses. *Proc. Natl. Acad. Sci. U. S. A. 116*, 1597−1602.
(6) Sevy, A. M., Jacobs, T. M., Crowe, J. E., and Meiler, J. (2015) Design of protein multi-specificity Uusing an independent sequence search reduces the barrier to low energy sequences. *PLoS Comput. Biol. 11*, e1004300.
(7) Sevy, A. M., and Meiler, J. (2014) Antibodies: Computer-aided prediction of structure and design of function. *Microbiol. Spectrum 2*, n/a.
(8) Froning, K. J., Leaver-Fay, A., Wu, X., Phan, S., Gao, L., Huang, F., Pustilnik, A., Bacica, M., Houlihan, K., Chai, Q., Fitchett, J. R., Hendle, J., Kuhlman, B., and Demarest, S. J. (2017) Computational design of a specific heavy chain/kappa light chain interface for expressing fully IgG bispecific antibodies. *Protein Sci. 26*, 2021−2038.
(9) Leman, J. K., Weitzner, B. D., Lewis, S. M., Adolf-Bryfogle, J., Alam, N., Alford, R. F., Aprahamian, M., Baker, D., Barlow, K. A., Barth, P., Basanta, B., Bender, B. J., Blacklock, K., Bonet, J., Boyken, S. E., Bradley, P., Bystroff, C., Conway, P., Cooper, S., Correia, B. E., Coventry, B., Das, R., De Jong, R. M., DiMaio, F., Dsilva, L., Dunbrack, R., Ford, A. S., Frenz, B., Fu, D. Y., Geniesse, C., Goldschmidt, L., Gowthaman, R., Gray, J. J., Gront, D., Guffy, S., Horowitz, S., Huang, P. S., Huber, T., Jacobs, T. M., Jeliazkov, J. R., Johnson, D. K., Kappel, K., Karanicolas, J., Khakzad, H., Khar, K. R., Khare, S. D., Khatib, F., Khramushin, A., King, I. C., Kleffner, R., Koepnick, B., Kortemme, T., Kuenze, G., Kuhlman, B., Kuroda, D., Labonte, J. W., Lai, J. K., Lapidoth, G., Leaver-Fay, A., Lindert, S., Linsky, T., London, N., Lubin, J. H., Lyskov, S., Maguire, J., Malmström, L., Marcos, E., Marcu, O., Marze, N. A., Meiler, J., Moretti, R., Mulligan, V. K., Nerli, S., Norn, C., Ó'Conchúir, S., Ollikainen, N., Ovchinnikov, S., Pacella, M. S., Pan, X., Park, H., Pavlovicz, R. E., Pethe, M., Pierce, B. G., Pilla, K. B., Raveh, B., Renfrew, P. D., Burman, S. S. R., Rubenstein, A., Sauer, M. F., Scheck, A., Schief, W., Schueler-Furman, O., Sedan, Y., Sevy, A. M., Sgourakis, N. G., Shi, L., Siegel, J. B., Silva, D. A., Smith, S., Song, Y., Stein, A., Szegedy, M., Teets, F. D., Thyme, S. B., Wang, R. Y., Watkins, A., Zimmerman, L., and Bonneau, R. (2020) Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods 17*, 665−680.
(10) Azoitei, M. L., Ban, Y. E., Julien, J. P., Bryson, S., Schroeter, A., Kalyuzhniy, O., Porter, J. R., Adachi, Y., Baker, D., Pai, E. F., and Schief, W. R. (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J. Mol. Biol. 415*, 175−192.
(11) Correia, B. E., Bates, J. T., Loomis, R. J., Baneyx, G., Carrico, C., Jardine, J. G., Rupert, P., Correnti, C., Kalyuzhniy, O., Vittal, V., Connell, M. J., Stevens, E., Schroeter, A., Chen, M., Macpherson, S., Serra, A. M., Adachi, Y., Holmes, M. A., Li, Y., Klevit, R. E., Graham, B. S., Wyatt, R. T., Baker, D., Strong, R. K., Crowe, J. E., Jr., Johnson, P. R., and Schief, W. R. (2014) Proof of principle for epitope-focused vaccine design. *Nature 507*, 201−206.

(12) McLellan, J. S., Correia, B. E., Chen, M., Yang, Y., Graham, B. S., Schief, W. R., and Kwong, P. D. (2011) Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus. *J. Mol. Biol. 409*, 853−866.
(13) Sesterhenn, F., Yang, C., Bonet, J., Cramer, J. T., Wen, X., Wang, Y., Chiang, C. I., Abriata, L. A., Kucharska, I., Castoro, G., Vollers, S. S., Galloux, M., Dheilly, E., Rosset, S., Corthésy, P., Georgeon, S., Villard, M., Richard, C. A., Descamps, D., Delgado, T., Oricchio, E., Rameix-Welti, M. A., Más, V., Ervin, S., Eléouët, J. F., Riffault, S., Bates, J. T., Julien, J. P., Li, Y., Jardetzky, T., Krey, T., and Correia, B. E. (2020) De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science 368*, eaay5051.
(14) Koehler Leman, J., Weitzner, B. D., Renfrew, P. D., Lewis, S. M., Moretti, R., Watkins, A. M., Mulligan, V. K., Lyskov, S., Adolf-Bryfogle, J., Labonte, J. W., Krys, J., Bystroff, C., Schief, W., Gront, D., Schueler-Furman, O., Baker, D., Bradley, P., Dunbrack, R., Kortemme, T., Leaver-Fay, A., Strauss, C. E. M., Meiler, J., Kuhlman, B., Gray, J. J., and Bonneau, R. (2020) Better together: Elements of successful scientific software development in a distributed collaborative community. *PLoS Comput. Biol. 16*, e1007507.
(15) Bender, B. J., Cisneros, A., 3rd, Duran, A. M., Finn, J. A., Fu, D., Lokits, A. D., Mueller, B. K., Sangha, A. K., Sauer, M. F., Sevy, A. M., Sliwoski, G., Sheehan, J. H., DiMaio, F., Meiler, J., and Moretti, R. (2016) Protocols for molecular modeling with Rosetta3 and RosettaScripts. *Biochemistry 55*, 4748−4763.
(16) Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., and Baker, D. (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One 6*, e20161.
(17) Chaudhury, S., Lyskov, S., and Gray, J. J. (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics 26*, 689−691.
(18) DeLuca, S., Khar, K., and Meiler, J. (2015) Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS One 10*, e0132508.
(19) Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Jr., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. (2017) The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput. 13*, 3031−3048.
(20) Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y. E., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popovic, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol. 487*, 545−574.
(21) Koehler Leman, J., Mueller, B. K., and Gray, J. J. (2017) Expanding the toolkit for membrane protein modeling in Rosetta. *Bioinformatics 33*, 754−756.
(22) Weinstein, J. Y., Elazar, A., and Fleishman, S. J. (2019) A lipophilicity-based energy function for membrane-protein modelling and design. *PLoS Comput. Biol. 15*, e1007318.
(23) Renfrew, P. D., Choi, E. J., Bonneau, R., and Kuhlman, B. (2012) Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS One 7*, e32637.
(24) Kappel, K., and Das, R. (2019) Sampling native-like structures of RNA-protein complexes through Rosetta folding and docking. *Structure 27*, 140−151.
(25) Weitzner, B. D., Kuroda, D., Marze, N., Xu, J., and Gray, J. J. (2014) Blind prediction performance of RosettaAntibody 3.0: grafting, relaxation, kinematic loop modeling, and full CDR optimization. *Proteins: Struct., Funct., Genet. 82*, 1611−1623.
(26) Sivasubramanian, A., Sircar, A., Chaudhury, S., and Gray, J. J. (2009) Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins: Struct., Funct., Genet. 74*, 497−514.
(27) Norn, C. H., Lapidoth, G., and Fleishman, S. J. (2017) High-accuracy modeling of antibody structures by a search for minimum-energy recombination of backbone fragments. *Proteins: Struct., Funct., Genet. 85*, 30−38.
(28) Song, Y., DiMaio, F., Wang, R. Y., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013) High-resolution comparative modeling with RosettaCM. *Structure 21*, 1735−1742.
(29) Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol. 331*, 281−299.
(30) Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., and Gray, J. J. (2011) Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One 6*, e22477.
(31) Chaudhury, S., and Gray, J. J. (2008) Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. *J. Mol. Biol. 381*, 1068−1087.
(32) Marze, N. A., Jeliazkov, J. R., Roy Burman, S. S., Boyken, S. E., DiMaio, F., and Gray, J. J. (2017) Modeling oblong proteins and water-mediated interfaces with RosettaDock in CAPRI rounds 28−35. *Proteins: Struct., Funct., Genet. 85*, 479−486.
(33) Sircar, A., and Gray, J. J. (2010) SnugDock: paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. *PLoS Comput. Biol. 6*, e1000644.
(34) Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B. D., Hu, X., Adachi, Y., Schief, W. R., and Dunbrack, R. L., Jr. (2018) RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLoS Comput. Biol. 14*, e1006112.
(35) Lapidoth, G. D., Baran, D., Pszolla, G. M., Norn, C., Alon, A., Tyka, M. D., and Fleishman, S. J. (2015) AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins: Struct., Funct., Genet. 83*, 1385−1406.
(36) Leaver-Fay, A., Froning, K. J., Atwell, S., Aldaz, H., Pustilnik, A., Lu, F., Huang, F., Yuan, R., Hassanali, S., Chamberlain, A. K., Fitchett, J. R., Demarest, S. J., and Kuhlman, B. (2016) Computationally designed bispecific antibodies using negative state repertoires. *Structure 24*, 641−651.
(37) Azoitei, M. L., Correia, B. E., Ban, Y. E., Carrico, C., Kalyuzhniy, O., Chen, L., Schroeter, A., Huang, P. S., McLellan, J. S., Kwong, P. D., Baker, D., Strong, R. K., and Schief, W. R. (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science 334*, 373−376.
(38) Correia, B. E., Ban, Y. E., Friend, D. J., Ellingson, K., Xu, H., Boni, E., Bradley-Hewitt, T., Bruhn-Johannsen, J. F., Stamatatos, L., Strong, R. K., and Schief, W. R. (2011) Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design. *J. Mol. Biol. 405*, 284−297.
(39) Silva, D. A., Correia, B. E., and Procko, E. (2016) Motif-driven design of protein-protein interfaces. *Methods Mol. Biol. 1414*, 285−304.
(40) Bonet, J., Wehrle, S., Schriever, K., Yang, C., Billet, A., Sesterhenn, F., Scheck, A., Sverrisson, F., Veselkova, B., Vollers, S., Lourman, R., Villard, M., Rosset, S., Krey, T., and Correia, B. E. (2018) Rosetta FunFolDes - A general framework for the computational design of functional proteins. *PLoS Comput. Biol. 14*, e1006623.
(41) Huang, P. S., Ban, Y. E., Richter, F., Andre, I., Vernon, R., Schief, W. R., and Baker, D. (2011) RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One 6*, e24109.
(42) Jacobs, T. M., Williams, B., Williams, T., Xu, X., Eletsky, A., Federizon, J. F., Szyperski, T., and Kuhlman, B. (2016) Design of structurally distinct proteins using strategies inspired by evolution. *Science 352*, 687−690.
(43) Guffy, S. L., Teets, F. D., Langlois, M. I., and Kuhlman, B. (2018) Protocols for requirement-driven protein design in the Rosetta modeling program. *J. Chem. Inf. Model. 58*, 895−901.

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

(44) Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., and Fleishman, S. J. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell 63*, 337−346.
(45) Labonte, J. W., Adolf-Bryfogle, J., Schief, W. R., and Gray, J. J. (2017) Residue-centric modeling and design of saccharide and glycoconjugate structures. *J. Comput. Chem. 38*, 276−287.
(46) Finn, J. A., Koehler Leman, J., Willis, J. R., Cisneros, A., 3rd, Crowe, J. E., Jr., and Meiler, J. (2016) Improving loop modeling of the antibody complementarity-determining region 3 using knowledge-based restraints. *PLoS One 11*, e0154811.
(47) North, B., Lehmann, A., and Dunbrack, R. L., Jr. (2011) A new clustering of antibody CDR loop conformations. *J. Mol. Biol. 406*, 228−256.
(48) Harris, L. J., Larson, S. B., Hasel, K. W., and McPherson, A. (1997) Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry 36*, 1581−1597.
(49) Bangaru, S., Lang, S., Schotsaert, M., Vanderven, H. A., Zhu, X., Kose, N., Bombardi, R., Finn, J. A., Kent, S. J., Gilchuk, P., Gilchuk, I., Turner, H. L., Garcia-Sastre, A., Li, S., Ward, A. B., Wilson, I. A., and Crowe, J. E., Jr. (2019) A site of vulnerability on the Influenza virus hemagglutinin head domain trimer interface. *Cell 177*, 1136−1152.
(50) Tu, C., Terraube, V., Tam, A. S., Stochaj, W., Fennell, B. J., Lin, L., Stahl, M., LaVallie, E. R., Somers, W., Finlay, W. J., Mosyak, L., Bard, J., and Cunningham, O. (2016) A combination of structural and empirical analyses delineates the key contacts mediating stability and affinity increases in an optimized biotherapeutic single-chain Fv (scFv). *J. Biol. Chem. 291*, 1267−1276.
(51) Decanniere, K., Muyldermans, S., and Wyns, L. (2000) Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes? *J. Mol. Biol. 300*, 83−91.
(52) Adolf-Bryfogle, J., Xu, Q., North, B., Lehmann, A., and Dunbrack, R. L., Jr. (2015) PyIgClassify: a database of antibody CDR structural classifications. *Nucleic Acids Res. 43*, 432−438.
(53) Al-Lazikani, B., Lesk, A. M., and Chothia, C. (1997) Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol. 273*, 927−948.
(54) Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M., and Poljak, R. J. (1989) Conformations of immunoglobulin hypervariable regions. *Nature 342*, 877−883.
(55) Morea, V., Tramontano, A., Rustici, M., Chothia, C., and Lesk, A. M. (1998) Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol. 275*, 269−294.
(56) Kelow, S. P., Adolf-Bryfogle, J., and Dunbrack, R. L. (2020) Hiding in plain sight: structure and sequence analysis reveals the importance of the antibody DE loop for antibody-antigen binding. *MAbs 12*, 1840005.
(57) Honegger, A., and Pluckthun, A. (2001) Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *J. Mol. Biol. 309*, 657−670.
(58) Brochet, X., Lefranc, M. P., and Giudicelli, V. (2008) IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res. 36*, 503−508.
(59) Lefranc, M. P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., Hadi-Saljoqi, S., Sasorith, S., Lefranc, G., and Kossida, S. (2015) IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res. 43*, 413−422.
(60) Marze, N. A., Lyskov, S., and Gray, J. J. (2016) Improved prediction of antibody VL-VH orientation. *Protein Eng., Des. Sel. 29*, 409−418.
(61) Cisneros, A., Nargi, R. S., Parrish, E. H., Haliburton, C. M., Meiler, J., and Crowe, J. E., Jr (2019) (2019) Role of antibody heavy and light chain interface residues in affinity maturation of binding to HIV envelope glycoprotein. *Mol. Sys. Des. Eng. 4*, 737−746.
(62) Wang, C., Bradley, P., and Baker, D. (2007) Protein-protein docking with backbone flexibility. *J. Mol. Biol. 373*, 503−519.
(63) Canutescu, A. A., and Dunbrack, R. L., Jr. (2003) Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci. 12*, 963−972.
(64) Stein, A., and Kortemme, T. (2013) Improvements to robotics-inspired conformational sampling in Rosetta. *PLoS One 8*, e63090.
(65) Mousa, J. J., Sauer, M. F., Sevy, A. M., Finn, J. A., Bates, J. T., Alvarado, G., King, H. G., Loerinc, L. B., Fong, R. H., Doranz, B. J., Correia, B. E., Kalyuzhniy, O., Wen, X., Jardetzky, T. S., Schief, W. R., Ohi, M. D., Meiler, J., and Crowe, J. E., Jr. (2016) Structural basis for nonneutralizing antibody competition at antigenic site II of the respiratory syncytial virus fusion protein. *Proc. Natl. Acad. Sci. U. S. A. 113*, E6849−E6858.
(66) Sangha, A. K., Dong, J., Williamson, L., Hashiguchi, T., Saphire, E. O., Crowe, J. E., Jr., and Meiler, J. (2017) Role of non-local interactions between CDR loops in binding affinity of MR78 antibody to Marburg virus glycoprotein. *Structure 25*, 1820−1828.
(67) Long, F., Fong, R. H., Austin, S. K., Chen, Z., Klose, T., Fokine, A., Liu, Y., Porta, J., Sapparapu, G., Akahata, W., Doranz, B. J., Crowe, J. E., Jr., Diamond, M. S., and Rossmann, M. G. (2015) Cryo-EM structures elucidate neutralizing mechanisms of anti-chikungunya human monoclonal antibodies with therapeutic activity. *Proc. Natl. Acad. Sci. U. S. A. 112*, 13898−13903.
(68) Krause, J. C., Ekiert, D. C., Tumpey, T. M., Smith, P. B., Wilson, I. A., and Crowe, J. E., Jr. (2011) An insertion mutation that distorts antibody binding site architecture enhances function of a human antibody. *mBio 2*, e00345-10.
(69) Tsibane, T., Ekiert, D. C., Krause, J. C., Martinez, O., Crowe, J. E., Jr., Wilson, I. A., and Basler, C. F. (2012) Influenza human monoclonal antibody 1F1 interacts with three major antigenic sites and residues mediating human receptor specificity in H1N1 viruses. *PLoS Pathog. 8*, e1003067.
(70) Flyak, A. I., Ruiz, S., Colbert, M. D., Luong, T., Crowe, J. E., Jr., Bailey, J. R., and Bjorkman, P. J. (2018) HCV broadly neutralizing antibodies use a CDRH3 disulfide motif to recognize an E2 glycoprotein site that can be targeted for vaccine design. *Cell Host Microbe 24*, 703−716.
(71) Bangaru, S., Zhang, H., Gilchuk, I. M., Voss, T. G., Irving, R. P., Gilchuk, P., Matta, P., Zhu, X., Lang, S., Nieusma, T., Richt, J. A., Albrecht, R. A., Vanderven, H. A., Bombardi, R., Kent, S. J., Ward, A. B., Wilson, I. A., and Crowe, J. E., Jr. (2018) A multifunctional human monoclonal neutralizing antibody that targets a unique conserved epitope on influenza HA. *Nat. Commun. 9*, 2669.
(72) Lammens, A., Baehner, M., Kohnert, U., Niewoehner, J., von Proff, L., Schraeml, M., Lammens, K., and Hopfner, K. P. (2013) Crystal structure of human TWEAK in complex with the Fab fragment of a neutralizing antibody reveals insights into receptor binding. *PLoS One 8*, e62697.
(73) Williams, W. B., Zhang, J., Jiang, C., Nicely, N. I., Fera, D., Luo, K., Moody, M. A., Liao, H. X., Alam, S. M., Kepler, T. B., Ramesh, A., Wiehe, K., Holland, J. A., Bradley, T., Vandergrift, N., Saunders, K. O., Parks, R., Foulger, A., Xia, S. M., Bonsignori, M., Montefiori, D. C., Louder, M., Eaton, A., Santra, S., Scearce, R., Sutherland, L., Newman, A., Bouton-Verville, H., Bowman, C., Bomze, H., Gao, F., Marshall, D. J., Whitesides, J. F., Nie, X., Kelsoe, G., Reed, S. G., Fox, C. B., Clary, K., Koutsoukos, M., Franco, D., Mascola, J. R., Harrison, S. C., Haynes, B. F., and Verkoczy, L. (2017) Initiation of HIV neutralizing B cell lineages with sequential envelope immunizations. *Nat. Commun. 8*, 1732.
(74) Nakanishi, T., Tsumoto, K., Yokota, A., Kondo, H., and Kumagai, I. (2008) Critical contribution of VH-VL interaction to reshaping of an antibody: the case of humanization of anti-lysozyme antibody, HyHEL-10. *Protein Sci. 17*, 261−270.
(75) Klein, F., Diskin, R., Scheid, J. F., Gaebler, C., Mouquet, H., Georgiev, I. S., Pancera, M., Zhou, T., Incesu, R. B., Fu, B. Z., Gnanapragasam, P. N., Oliveira, T. Y., Seaman, M. S., Kwong, P. D., Bjorkman, P. J., and Nussenzweig, M. C. (2013) Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell 153*, 126−138.

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

(76) Yacoob, C., Pancera, M., Vigdorovich, V., Oliver, B. G., Glenn, J. A., Feng, J., Sather, D. N., McGuire, A. T., and Stamatatos, L. (2016) Differences in allelic frequency and CDRH3 region limit the engagement of HIV Env immunogens by putative VRC01 neutralizing antibody precursors. *Cell Rep. 17*, 1560−1570.
(77) Saunders, K. O., Nicely, N. I., Wiehe, K., Bonsignori, M., Meyerhoff, R. R., Parks, R., Walkowicz, W. E., Aussedat, B., Wu, N. R., Cai, F., Vohra, Y., Park, P. K., Eaton, A., Go, E. P., Sutherland, L. L., Scearce, R. M., Barouch, D. H., Zhang, R., Von Holle, T., Overman, R. G., Anasti, K., Sanders, R. W., Moody, M. A., Kepler, T. B., Korber, B., Desaire, H., Santra, S., Letvin, N. L., Nabel, G. J., Montefiori, D. C., Tomaras, G. D., Liao, H. X., Alam, S. M., Danishefsky, S. J., and Haynes, B. F. (2017) Vaccine elicitation of high mannose-dependent neutralizing antibodies against the V3-glycan broadly neutralizing epitope in nonhuman primates. *Cell Rep. 18*, 2175−2188.
(78) Scally, S. W., McLeod, B., Bosch, A., Miura, K., Liang, Q., Carroll, S., Reponen, S., Nguyen, N., Giladi, E., Rämisch, S., Yusibov, V., Bradley, A., Lemiale, F., Schief, W. R., Emerling, D., Kellam, P., King, C. R., and Julien, J. P. (2017) Molecular definition of multiple sites of antibody inhibition of malaria transmission-blocking vaccine antigen Pfs25. *Nat. Commun. 8*, 1568.
(79) Wu, Y., Eigenbrot, C., Liang, W. C., Stawicki, S., Shia, S., Fan, B., Ganesan, R., Lipari, M. T., and Kirchhofer, D. (2007) Structural insight into distinct mechanisms of protease inhibition by antibodies. *Proc. Natl. Acad. Sci. U. S. A. 104*, 19784−19789.
(80) Teplyakov, A., Luo, J., Obmolova, G., Malia, T. J., Sweet, R., Stanfield, R. L., Kodangattil, S., Almagro, J. C., and Gilliland, G. L. (2014) Antibody modeling assessment II. Structures and models. *Proteins: Struct., Funct., Genet. 82*, 1563−1582.
(81) Makabe, K., Nakanishi, T., Tsumoto, K., Tanaka, Y., Kondo, H., Umetsu, M., Sone, Y., Asano, R., and Kumagai, I. (2008) Thermodynamic consequences of mutations in vernier zone residues of a humanized anti-human epidermal growth factor receptor murine antibody, 528. *J. Biol. Chem. 283*, 1156−1166.
(82) Wu, N. C., Grande, G., Turner, H. L., Ward, A. B., Xie, J., Lerner, R. A., and Wilson, I. A. (2017) In vitro evolution of an influenza broadly neutralizing antibody is modulated by hemagglutinin receptor specificity. *Nat. Commun. 8*, 15371.
(83) Pantua, H., Diao, J., Ultsch, M., Hazen, M., Mathieu, M., McCutcheon, K., Takeda, K., Date, S., Cheung, T. K., Phung, Q., Hass, P., Arnott, D., Hongo, J. A., Matthews, D. J., Brown, A., Patel, A. H., Kelley, R. F., Eigenbrot, C., and Kapadia, S. B. (2013) Glycan shifting on hepatitis C virus (HCV) E2 glycoprotein is a mechanism for escape from broadly neutralizing antibodies. *J. Mol. Biol. 425*, 1899−1914.
(84) Almagro, J. C., Beavers, M. P., Hernandez-Guzman, F., Maier, J., Shaulsky, J., Butenhof, K., Labute, P., Thorsteinson, N., Kelly, K., Teplyakov, A., Luo, J., Sweet, R., and Gilliland, G. L. (2011) Antibody modeling assessment. *Proteins: Struct., Funct., Genet. 79*, 3050−3066.
(85) Almagro, J. C., Teplyakov, A., Luo, J., Sweet, R. W., Kodangattil, S., Hernandez-Guzman, F., and Gilliland, G. L. (2014) Second antibody modeling assessment (AMA-II). *Proteins: Struct., Funct., Genet. 82*, 1553−1562.
(86) Weitzner, B. D., and Gray, J. J. (2017) Accurate Structure Prediction of CDR H3 Loops Enabled by a Novel Structure-Based C-Terminal Constraint. *J. Immunol. 198*, 505−515.
(87) Lyskov, S., Chou, F. C., Conchuir, S. O., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D., Sripakdeevong, P., Borgo, B., Havranek, J. J., Kuhlman, B., Kortemme, T., Bonneau, R., Gray, J. J., and Das, R. (2013) Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). *PLoS One 8*, e63906.
(88) Moretti, R., Lyskov, S., Das, R., Meiler, J., and Gray, J. J. (2018) Web-accessible molecular modeling with Rosetta: The Rosetta Online Server that Includes Everyone (ROSIE). *Protein Sci. 27*, 259−268.
(89) Sircar, A., Kim, E. T., and Gray, J. J. (2009) RosettaAntibody: antibody variable region homology modeling server. *Nucleic Acids Res. 37*, W474−479.
(90) Bender, B. J., Vortmeier, G., Ernicke, S., Bosse, M., Kaiser, A., Els-Heindl, S., Krug, U., Beck-Sickinger, A., Meiler, J., and Huster, D. (2019) Structural model of ghrelin bound to its G protein-coupled receptor. *Structure 27*, 537−544.
(91) Briney, B. S., Willis, J. R., Hicar, M. D., Thomas, J. W., 2nd, and Crowe, J. E., Jr. (2012) Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology 137*, 56−64.
(92) Pancera, M., McLellan, J. S., Wu, X., Zhu, J., Changela, A., Schmidt, S. D., Yang, Y., Zhou, T., Phogat, S., Mascola, J. R., and Kwong, P. D. (2010) Crystal structure of PG16 and chimeric dissection with somatically related PG9: structure-function analysis of two quaternary-specific antibodies that effectively neutralize HIV-1. *J. Virol. 84*, 8098−8110.
(93) Henderson, K. A., Streltsov, V. A., Coley, A. M., Dolezal, O., Hudson, P. J., Batchelor, A. H., Gupta, A., Bai, T., Murphy, V. J., Anders, R. F., Foley, M., and Nuttall, S. D. (2007) Structure of an IgNAR-AMA1 complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition. *Structure 15*, 1452−1466.
(94) McLellan, J. S., Pancera, M., Carrico, C., Gorman, J., Julien, J. P., Khayat, R., Louder, R., Pejchal, R., Sastry, M., Dai, K., O'Dell, S., Patel, N., Shahzad-ul-Hussan, S., Yang, Y., Zhang, B., Zhou, T., Zhu, J., Boyington, J. C., Chuang, G. Y., Diwanji, D., Georgiev, I., Do Kwon, Y., Lee, D., Louder, M. K., Moquin, S., Schmidt, S. D., Yang, Z. Y., Bonsignori, M., Crump, J. A., Kapiga, S. H., Sam, N. E., Haynes, B. F., Burton, D. R., Koff, W. C., Walker, L. M., Phogat, S., Wyatt, R., Orwenyo, J., Wang, L. X., Arthos, J., Bewley, C. A., Mascola, J. R., Nabel, G. J., Schief, W. R., Ward, A. B., Wilson, I. A., and Kwong, P. D. (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature 480*, 336−343.
(95) Shirai, H., Kidera, A., and Nakamura, H. (1996) Structural classification of CDR-H3 in antibodies. *FEBS Lett. 399*, 1−8.
(96) Weitzner, B. D., Dunbrack, R. L., Jr., and Gray, J. J. (2015) The origin of CDR H3 structural diversity. *Structure 23*, 302−311.
(97) Sircar, A., Sanni, K. A., Shi, J., and Gray, J. J. (2011) Analysis and modeling of the variable region of camelid single-domain antibodies. *J. Immunol. 186*, 6357−6367.
(98) Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A., and Hamers, R. (1994) Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng., Des. Sel. 7*, 1129−1135.
(99) Sivasubramanian, A., Chao, G., Pressler, H. M., Wittrup, K. D., and Gray, J. J. (2006) Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure 14*, 401−414.
(100) Thornburg, N. J., Nannemann, D. P., Blum, D. L., Belser, J. A., Tumpey, T. M., Deshpande, S., Fritz, G. A., Sapparapu, G., Krause, J. C., Lee, J. H., Ward, A. B., Lee, D. E., Li, S., Winarski, K. L., Spiller, B. W., Meiler, J., and Crowe, J. E., Jr. (2013) Human antibodies that neutralize respiratory droplet transmissible H5N1 influenza viruses. *J. Clin. Invest. 123*, 4405−4409.
(101) Willis, J. R., Briney, B. S., DeLuca, S. L., Crowe, J. E., Jr., and Meiler, J. (2013) Human germline antibody gene segments encode polyspecific antibodies. *PLoS Comput. Biol. 9*, e1003045.
(102) Willis, J. R., Sapparapu, G., Murrell, S., Julien, J. P., Singh, V., King, H. G., Xia, Y., Pickens, J. A., LaBranche, C. C., Slaughter, J. C., Montefiori, D. C., Wilson, I. A., Meiler, J., and Crowe, J. E., Jr. (2015) Redesigned HIV antibodies exhibit enhanced neutralizing potency and breadth. *J. Clin. Invest. 125*, 2523−2531.
(103) Smith, C. A., and Kortemme, T. (2008) Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol. 380*, 742−756.
(104) Lehmann, A., Wixted, J. H., Shapovalov, M. V., Roder, H., Dunbrack, R. L., Jr., and Robinson, M. K. (2015) Stability engineering of anti-EGFR scFv antibodies by rational design of a lambda-to-kappa swap of the VL framework using a structure-guided approach. *MAbs 7*, 1058−1071.
(105) Baran, D., Pszolla, M. G., Lapidoth, G. D., Norn, C., Dym, O., Unger, T., Albeck, S., Tyka, M. D., and Fleishman, S. J. (2017) Principles for computational design of binding antibodies. *Proc. Natl. Acad. Sci. U. S. A. 114*, 10900−10905.

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

(106) Warszawski, S., Netzer, R., Tawfik, D. S., and Fleishman, S. J. (2014) A "fuzzy"-logic language for encoding multiple physical traits in biomolecules. *J. Mol. Biol. 426*, 4125−4138.
(107) Sheffler, W., and Baker, D. (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci. 18*, 229−239.
(108) Lawrence, M. C., and Colman, P. M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol. 234*, 946−950.
(109) Xu, J., Tack, D., Hughes, R. A., Ellington, A. D., and Gray, J. J. (2014) Structure-based non-canonical amino acid design to covalently crosslink an antibody-antigen complex. *J. Struct. Biol. 185*, 215−222.
(110) Drew, K., Renfrew, P. D., Craven, T. W., Butterfoss, G. L., Chou, F. C., Lyskov, S., Bullock, B. N., Watkins, A., Labonte, J. W., Pacella, M., Kilambi, K. P., Leaver-Fay, A., Kuhlman, B., Gray, J. J., Bradley, P., Kirshenbaum, K., Arora, P. S., Das, R., and Bonneau, R. (2013) Adding diverse noncanonical backbones to rosetta: enabling peptidomimetic design. *PLoS One 8*, e67051.
(111) Miklos, A. E., Kluwe, C., Der, B. S., Pai, S., Sircar, A., Hughes, R. A., Berrondo, M., Xu, J., Codrea, V., Buckley, P. E., Calm, A. M., Welsh, H. S., Warner, C. R., Zacharko, M. A., Carney, J. P., Gray, J. J., Georgiou, G., Kuhlman, B., and Ellington, A. D. (2012) Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol. 19*, 449−455.
(112) Jung, S., and Pluckthun, A. (1997) Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng., Des. Sel. 10*, 959−966.
(113) Babor, M., and Kortemme, T. (2009) Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility. *Proteins: Struct., Funct., Genet. 75*, 846−858.
(114) Ekiert, D. C., Kashyap, A. K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., Lee, J. H., Dillon, M. A., O'Neil, R. E., Faynboym, A. M., Horowitz, M., Horowitz, L., Ward, A. B., Palese, P., Webby, R., Lerner, R. A., Bhatt, R. R., and Wilson, I. A. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature 489*, 526−532.
(115) Sevy, A. M., Panda, S., Crowe, J. E., Jr., Meiler, J., and Vorobeychik, Y. (2018) Integrating linear optimization with structural modeling to increase HIV neutralization breadth. *PLoS Comput. Biol. 14*, e1005999.
(116) Leaver-Fay, A., Jacak, R., Stranges, P. B., and Kuhlman, B. (2011) A generic program for multistate protein design. *PLoS One 6*, e20937.
(117) Baeuerle, P. A., and Reinhardt, C. (2009) Bispecific T-cell engaging antibodies for cancer therapy. *Cancer Res. 69*, 4941−4944.
(118) Lewis, S. M., Wu, X., Pustilnik, A., Sereno, A., Huang, F., Rick, H. L., Guntas, G., Leaver-Fay, A., Smith, E. M., Ho, C., Hansen-Estruch, C., Chamberlain, A. K., Truhlar, S. M., Conner, E. M., Atwell, S., Kuhlman, B., and Demarest, S. J. (2014) Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface. *Nat. Biotechnol. 32*, 191−198.
(119) Ekiert, D. C., Bhabha, G., Elsliger, M. A., Friesen, R. H., Jongeneelen, M., Throsby, M., Goudsmit, J., and Wilson, I. A. (2009) Antibody recognition of a highly conserved influenza virus epitope. *Science 324*, 246−251.
(120) Sui, J., Hwang, W. C., Perez, S., Wei, G., Aird, D., Chen, L. M., Santelli, E., Stec, B., Cadwell, G., Ali, M., Wan, H., Murakami, A., Yammanuru, A., Han, T., Cox, N. J., Bankston, L. A., Donis, R. O., Liddington, R. C., and Marasco, W. A. (2009) Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol. 16*, 265−273.
(121) Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science 332*, 816−821.
(122) Fleishman, S. J., Corn, J. E., Strauch, E. M., Whitehead, T. A., Karanicolas, J., and Baker, D. (2011) Hotspot-centric de novo design of protein binders. *J. Mol. Biol. 413*, 1047−1062.
(123) Strauch, E. M., Bernard, S. M., La, D., Bohn, A. J., Lee, P. S., Anderson, C. E., Nieusma, T., Holstein, C. A., Garcia, N. K., Hooper, K. A., Ravichandran, R., Nelson, J. W., Sheffler, W., Bloom, J. D., Lee, K. K., Ward, A. B., Yager, P., Fuller, D. H., Wilson, I. A., and Baker, D. (2017) Computational design of trimeric influenza-neutralizing proteins targeting the hemagglutinin receptor binding site. *Nat. Biotechnol. 35*, 667−671.
(124) Sevy, A. M., Gilchuk, I. M., Brown, B. P., Bozhanova, N. G., Nargi, R., Jensen, M., Meiler, J., and Crowe, J. E., Jr. (2020) Computationally designed cyclic peptides derived from an antibody loop increase breadth of binding for Influenza variants. *Structure 28*, 1114−1123.
(125) Burton, D. R. (2017) (2017) What are the most powerful immunogen design vaccine strategies? Reverse Vaccinology 2.0 shows great promise. *Cold Spring Harbor Perspect. Biol. 9*, a030262.
(126) Akram, A., and Inman, R. D. (2012) Immunodominance: a pivotal principle in host response to viral infections. *Clin. Immunol. 143*, 99−115.
(127) Duan, H., Chen, X., Boyington, J. C., Cheng, C., Zhang, Y., Jafari, A. J., Stephens, T., Tsybovsky, Y., Kalyuzhniy, O., Zhao, P., Menis, S., Nason, M. C., Normandin, E., Mukhamedova, M., DeKosky, B. J., Wells, L., Schief, W. R., Tian, M., Alt, F. W., Kwong, P. D., and Mascola, J. R. (2018) Glycan masking focuses immune responses to the HIV-1 CD4-binding site and enhances elicitation of VRC01-class precursor antibodies. *Immunity 49*, 301−311.
(128) Ofek, G., Guenaga, F. J., Schief, W. R., Skinner, J., Baker, D., Wyatt, R., and Kwong, P. D. (2010) Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl. Acad. Sci. U. S. A. 107*, 17880−17887.
(129) Correia, B. E., Ban, Y. E., Holmes, M. A., Xu, H., Ellingson, K., Kraft, Z., Carrico, C., Boni, E., Sather, D. N., Zenobia, C., Burke, K. Y., Bradley-Hewitt, T., Bruhn-Johannsen, J. F., Kalyuzhniy, O., Baker, D., Strong, R. K., Stamatatos, L., and Schief, W. R. (2010) Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure 18*, 1116−1126.
(130) Zhou, J., and Grigoryan, G. (2015) Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci. 24*, 508−524.
(131) Procko, E., Berguig, G. Y., Shen, B. W., Song, Y., Frayo, S., Convertine, A. J., Margineantu, D., Booth, G., Correia, B. E., Cheng, Y., Schief, W. R., Hockenbery, D. M., Press, O. W., Stoddard, B. L., Stayton, P. S., and Baker, D. (2014) A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell 157*, 1644−1656.
(132) Kvansakul, M., Wei, A. H., Fletcher, J. I., Willis, S. N., Chen, L., Roberts, A. W., Huang, D. C., and Colman, P. M. (2010) Structural basis for apoptosis inhibition by Epstein-Barr virus BHRF1. *PLoS Pathog. 6*, e1001236.
(133) Goulet, A., Blangy, S., Redder, P., Prangishvili, D., Felisberto-Rodrigues, C., Forterre, P., Campanacci, V., and Cambillau, C. (2009) Acidianus filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage. *Proc. Natl. Acad. Sci. U. S. A. 106*, 21155−21160.
(134) Campeotto, I., Goldenzweig, A., Davey, J., Barfod, L., Marshall, J. M., Silk, S. E., Wright, K. E., Draper, S. J., Higgins, M. K., and Fleishman, S. J. (2017) One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proc. Natl. Acad. Sci. U. S. A. 114*, 998−1002.
(135) Sampath, S., Carrico, C., Janes, J., Gurumoorthy, S., Gibson, C., Melcher, M., Chitnis, C. E., Wang, R., Schief, W. R., and Smith, J. D. (2013) Glycan masking of *Plasmodium vivax* Duffy Binding Protein for probing protein binding function and vaccine development. *PLoS Pathog. 9*, e1003420.
(136) Pancera, M., Majeed, S., Ban, Y. E., Chen, L., Huang, C. C., Kong, L., Kwon, Y. D., Stuckey, J., Zhou, T., Robinson, J. E., Schief, W. R., Sodroski, J., Wyatt, R., and Kwong, P. D. (2010) Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl. Acad. Sci. U. S. A. 107*, 1166−1171.

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846

(137) Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., Di Costanzo, L., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., Green, R. K., Guranović, V., Guzenko, D., Hudson, B. P., Lawson, C. L., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Persikova, I., Randle, C., Rose, A., Rose, Y., Sali, A., Segura, J., Sekharan, M., Shao, C., Tao, Y. P., Voigt, M., Westbrook, J. D., Young, J. Y., Zardecki, C., and Zhuravleva, M. (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res. 49*, D437−d451.
(138) Terwilliger, T. C., Dimaio, F., Read, R. J., Baker, D., Bunkoczi, G., Adams, P. D., Grosse-Kunstleve, R. W., Afonine, P. V., and Echols, N. (2012) phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics 13*, 81−90.
(139) DiMaio, F., Echols, N., Headd, J. J., Terwilliger, T. C., Adams, P. D., and Baker, D. (2013) Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods 10*, 1102−1104.
(140) Wang, R. Y., Kudryashev, M., Li, X., Egelman, E. H., Basler, M., Cheng, Y., Baker, D., and DiMaio, F. (2015) De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods 12*, 335−338.
(141) Frenz, B., Ramisch, S., Borst, A. J., Walls, A. C., Adolf-Bryfogle, J., Schief, W. R., Veesler, D., and DiMaio, F. (2019) Automatically fixing errors in glycoprotein structures with Rosetta. *Structure 27*, 134−139.
(142) He, L., de Val, N., Morris, C. D., Vora, N., Thinnes, T. C., Kong, L., Azadnia, P., Sok, D., Zhou, B., Burton, D. R., Wilson, I. A., Nemazee, D., Ward, A. B., and Zhu, J. (2016) Presenting native-like trimeric HIV-1 antigens with self-assembling nanoparticles. *Nat. Commun. 7*, 12041.
(143) Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T. O., Gonen, T., King, N. P., and Baker, D. (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science 353*, 389−394.
(144) King, N. P., and Lai, Y. T. (2013) Practical approaches to designing novel protein assemblies. *Curr. Opin. Struct. Biol. 23*, 632−638.
(145) Lai, Y. T., King, N. P., and Yeates, T. O. (2012) Principles for designing ordered protein assemblies. *Trends Cell Biol. 22*, 653−661.
(146) Hsia, Y., Bale, J. B., Gonen, S., Shi, D., Sheffler, W., Fong, K. K., Nattermann, U., Xu, C., Huang, P. S., Ravichandran, R., Yi, S., Davis, T. N., Gonen, T., King, N. P., and Baker, D. (2016) Design of a hyperstable 60-subunit protein dodecahedron. *Nature 535*, 136−139.
(147) Butterfield, G. L., Lajoie, M. J., Gustafson, H. H., Sellers, D. L., Nattermann, U., Ellis, D., Bale, J. B., Ke, S., Lenz, G. H., Yehdego, A., Ravichandran, R., Pun, S. H., King, N. P., and Baker, D. (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature 552*, 415−420.
(148) Marcandalli, J., Fiala, B., Ols, S., Perotti, M., de van der Schueren, W., Snijder, J., Hodge, E., Benhaim, M., Ravichandran, R., Carter, L., Sheffler, W., Brunner, L., Lawrenz, M., Dubois, P., Lanzavecchia, A., Sallusto, F., Lee, K. K., Veesler, D., Correnti, C. E., Stewart, L. J., Baker, D., Lore, K., Perez, L., and King, N. P. (2019) Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for Respiratory Syncytial Virus. *Cell 176*, 1420−1431.
(149) Brouwer, P. J. M., Antanasijevic, A., Berndsen, Z., Yasmeen, A., Fiala, B., Bijl, T. P. L., Bontjer, I., Bale, J. B., Sheffler, W., Allen, J. D., Schorcht, A., Burger, J. A., Camacho, M., Ellis, D., Cottrell, C. A., Behrens, A. J., Catalano, M., Del Moral-Sanchez, I., Ketas, T. J., LaBranche, C., van Gils, M. J., Sliepen, K., Stewart, L. J., Crispin, M., Montefiori, D. C., Baker, D., Moore, J. P., Klasse, P. J., Ward, A. B., King, N. P., and Sanders, R. W. (2019) Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. *Nat. Commun. 10*, 4272.

https://dx.doi.org/10.1021/acs.biochem.0c00912
Biochemistry 2021, 60, 825−846