

RESEARCH ARTICLE

Improving Loop Modeling of the Antibody Complementarity-Determining Region 3 Using Knowledge-Based Restraints

Jessica A. Finn¹, Julia Koehler Leman⁷, Jordan R. Willis⁵, Alberto Cisneros, III⁵, James E. Crowe, Jr^{1,3,5,6}, Jens Meiler^{2,4,5*}

1 Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Department of Chemistry, Vanderbilt University, Nashville, Tennessee, United States of America, **3** Department of Pediatrics, Vanderbilt University, Nashville, Tennessee, United States of America, **4** Department of Pharmacology, Vanderbilt University, Nashville, Tennessee, United States of America, **5** Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee, United States of America, **6** Vanderbilt Vaccine Center, Vanderbilt University, Nashville, Tennessee, United States of America, **7** Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America

* jens@meilerlab.org



OPEN ACCESS

Citation: Finn JA, Koehler Leman J, Willis JR, Cisneros A, III, Crowe JE, Jr, Meiler J (2016) Improving Loop Modeling of the Antibody Complementarity-Determining Region 3 Using Knowledge-Based Restraints. PLoS ONE 11(5): e0154811. doi:10.1371/journal.pone.0154811

Editor: Mitchell Ho, National Cancer Institute, NIH, UNITED STATES

Received: February 4, 2016

Accepted: April 19, 2016

Published: May 16, 2016

Copyright: © 2016 Finn et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Institute of Allergy and Infectious Diseases grants R56 AI110750 and U19 AI117905, and the National Institute of Allergy and Infectious Diseases Contract HHSN272201400024C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Structural restrictions are present even in the most sequence diverse portions of antibodies, the complementary determining region (CDR) loops. Previous studies identified robust rules that define canonical structures for five of the six CDR loops, however the heavy chain CDR 3 (HCDR3) defies standard classification attempts. The HCDR3 loop can be subdivided into two domains referred to as the “torso” and the “head” domains and two major families of canonical torso structures have been identified; the more prevalent “bulged” and less frequent “non-bulged” torsos. In the present study, we found that Rosetta loop modeling of 28 benchmark bulged HCDR3 loops is improved with knowledge-based structural restraints developed from available antibody crystal structures in the PDB. These restraints restrict the sampling space Rosetta searches in the torso domain, limiting the ϕ and ψ angles of these residues to conformations that have been experimentally observed. The application of these restraints in Rosetta result in more native-like structure sampling and improved score-based differentiation of native-like HCDR3 models, significantly improving our ability to model antibody HCDR3 loops.

Introduction

The field of antibody-mediated immunity has long benefited from structural studies of protein-protein interactions, in most cases through the determination of co-crystal structures of antibodies in complex with their antigens. Such studies often reveal the molecular mechanism of pathogen neutralization [1–4]. However, the size and complexity of the antibody repertoire coupled with the substantial resources needed for experimental structure determination

prohibit such studies on a comprehensive scale. B cell development leads to the generation of a large population of unique antibody proteins, and it is theorized that this diverse antibody repertoire may contain 10^{11} or more different protein sequences [5,6]. Recent studies determined that the circulating antibody repertoire contains at least 10^6 unique sequences, a number still far too large for comprehensive experimental structural studies [7,8].

Analysis of antibody structures determined by X-ray crystallography revealed conservation of structural features even in the regions of the antibody with the most sequence diversity, the six complementarity determining region (CDR) loops, which are responsible for antigen binding. Three of these loops are contributed by the heavy chain component of the fragment variable (Fv) domain of the antibody (HCDRs), and three are contributed by the light chain Fv domain (LCDRs). Two studies have identified robust rules that define canonical structures for five of the six CDR loops [9,10]. However, the HCDR3 defies classification attempts. The HCDR3 is encoded by the junction of three gene segments (V, D and J genes) connected by random nucleotide additions or deletions that are not encoded in the antibody germline gene segments, but rather introduced by the host enzyme terminal deoxynucleotidyl transferase during antibody gene recombination. The HCDR3 is therefore significantly more diverse in sequence length and composition than the other CDR loops, which are encoded by either a single gene segment (heavy and light chain CDRs 1 and 2) or by a simplified junction (LCDR3) [11–13]. As a result a large and diverse conformational space is observed for HCDR3s. Accordingly, HCDR3 is often especially important for antigen recognition and binding as has been revealed in previous structural studies [14].

The Rosetta software suite for macromolecular modeling can *de novo* predict the structure of a protein or portions thereof. The tertiary structure of a protein is determined from its primary sequence by pairing effective sampling techniques with knowledge-based energy functions. These energy functions for the most part assume that optimal geometries within proteins can be derived from a statistical analysis of the available structural information stored in the Protein Data Bank [15,16]. Similar approaches are used during comparative modeling, when structurally divergent regions (typically loops) of otherwise homologous proteins must be predicted [17]. Rosetta is capable of predicting antibody structures with low root mean square deviation (RMSD) to experimental structures outside the HCDR3; however accurately modeling the HCDR3 loop remains a challenge [18–21].

In an effort to classify canonical structures of the HCDR3 loop, prior work has subdivided it into two domains: the less diverse “torso” and the more variable “head” (Fig 1) [9,10]. Two major families of canonical torso structures have been identified, and are referred to as “bulged” and “non-bulged” torsos [10]. In this study, the geometries of the bulged torso domain have been used to develop restraints that restrict the sampling space of the HCDR3 torso and result in more native-like models when *de novo* modeling the entire HCDR3 loop.

Previous studies have used restraints to model the bulged HCDR3 torso, following rules previously described by Shirai et al. wherein a pseudodihedral angle restraint was calculated from the $C\alpha$ atoms of residues T5, T6, T7 and the following initial residue of Framework 4 to define the bulged or non-bulged torso [18,21–24]. Weitzner et al. [18] utilized RosettaAntibody implemented within the Rosetta 3 framework to predict the structures of 11 previously unpublished antibody structures for the second antibody modeling assessment (AMA-II) [21]. The longest HCDR3 loop in AMA-II contained 16 residues, and was predicted by the RosettaAntibody team with an RMSD of 3.70Å to the native HCDR3 loop [18, 21]. Shirai et al. also competed in AMA-II, and used their torso restraint rules to filter results generated by a pipeline that includes both Spanner and OSCAR for loop structure prediction; in comparison to the RosettaAntibody team described above, their best model for the longest HCDR3 loop had an RMSD of 3.29Å to the native HCDR3 loop [21, 23].

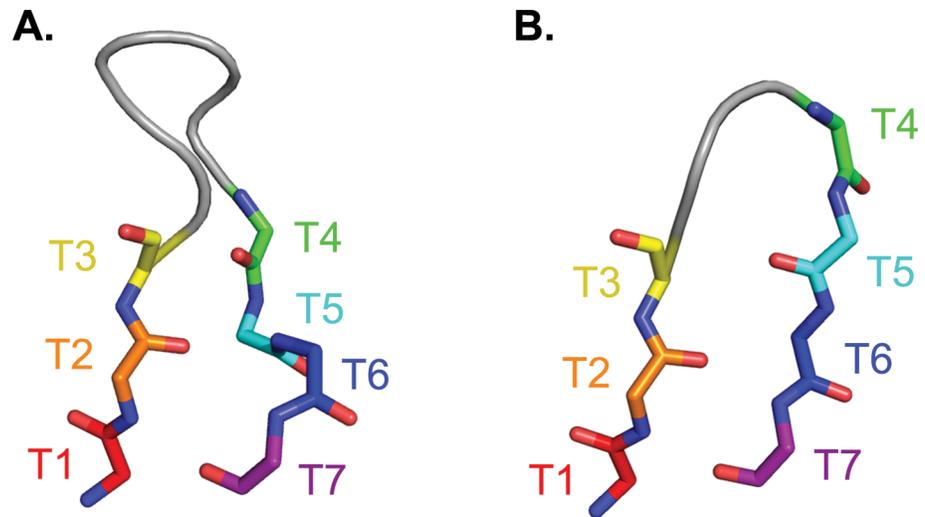


Fig 1. Defining the HCDR3 torso. The torso is defined as the first three and last four residues of the HCDR3 loop, numbered from T1 to T7. Main chain atoms are shown for bulged (panel A; PDBID 1UYW) and non-bulged (panel B; PDBID 2J88) torsos. In many (but not all) bulged torsos, a side-chain interaction between T2 and T6 causes the C-terminal side of the torso to bulge outward; the lack of such an interaction in non-bulged torsos leaves the beta-strand structure intact.

doi:10.1371/journal.pone.0154811.g001

In this study, a novel set of restraints was tested on 28 previously crystallized human antibodies with HCDR3 loops of increasing length and structural complexity. We expect that these restraints will improve modeling of antibodies for which no structural information is available, providing a means by which comprehensive structural studies of antibodies may be accomplished.

Results

Measuring bulged and non-bulged torso dihedral angles

An annotated list of antibodies was used to cull experimentally derived structures from the Protein Data Bank (PDB), expanding upon the list published by North et al. [10]. Following the IMGT conventions for defining the HCDR3, where the first HCDR3 residue occurs immediately following the V-gene residue Cys104 and the last HCDR3 residue occurs immediately preceding the J-gene residue Trp118, the torso is defined as the first three and the last four residues of the HCDR3 [10,25]. Accordingly, torso domain regions were pulled from these structures as two short peptide fragments (T1-T3 and T4-T7) and clustered using Rosetta at a threshold of 2 Å to separate bulged and non-bulged torsos. Previous studies identified a sequence motif (Arg or Lys at T2 and Asp at T6) that contributes to bulged torso formation in some but not all cases; these key residues were conserved in our bulged cluster, with 80% of bulged structures presenting Arg or Lys at T2, 73% presenting Asp at T6, and 65% retaining the complete T2/T6 sequence motif (S1 Fig) [9,10]. We found that germline-encoded regions of the antibody sequence often contribute these critical residues, as the end of the V gene segment contributes the first two to three torso residues while the J gene segment contributes the last four torso residues. The T2/T6 sequence motif that is often found in bulged torsos is present in 73% of naïve V and 92% of J germline gene allele segments (S1 Fig).

The ϕ and ψ angles of the seven torso residues of each antibody structure were measured, with key differences between bulged and non-bulged torsos identified in the ψ angles of residues T4 and T6 (Table 1). However, upon further study of previously defined torso clusters we

Table 1. Bulged and non-bulged dihedral angle measurements.

Torso Residue	Bulged		Non-bulged	
	ϕ	ψ	ϕ	ψ
T1	-145 ± 9	148 ± 12	-146 ± 12	145 ± 16
T2	-101 ± 22	142 ± 13	-109 ± 20	136 ± 26
T3	-107 ± 32	137 ± 33	-119 ± 44	138 ± 51
T4	-121 ± 49	161 ± 48	-82 ± 49	3 ± 59
T5	-95 ± 35	98 ± 26	-126 ± 43	136 ± 53
T6	-87 ± 18	-30 ± 26	-118 ± 34	129 ± 24
T7	-126 ± 14	134 ± 10	-125 ± 19	136 ± 11

The average and standard deviation of ϕ and ψ angles were calculated from existing human and mouse antibody crystal structures available in the PDB. Torso structures were clustered as bulged ($n = 218$) and non-bulged ($n = 38$) using a cluster radius of 2 Å.

doi:10.1371/journal.pone.0154811.t001

observed that the ψ angle of T4 is able to form two distinct conformations in both bulged and non-bulged torso clusters, and the T4 ψ angle does not distinguish between bulged and non-bulged torso clusters; the differences we observed when comparing all bulged antibodies to all non-bulged antibodies were due to the limited sample size of structures available in the PDB for these sub-conformations (S2 Fig) [10]. This is in contrast to for example T5, where a larger standard deviation is observed but still a statistically significant preference for a smaller ψ angle in a bulged torso exists. Average ϕ and ψ angles were calculated as follows:

$$\text{atan2}\left(\frac{\sum \sin\alpha}{n}, \frac{\sum \cos\alpha}{n}\right) \tag{1}$$

An approximate standard deviation was found using the following equations. For the vector v :

$$\vec{v} = \left(\frac{\sin\alpha}{n}, \frac{\cos\alpha}{n}\right) \tag{2}$$

Approximate standard deviation is calculated using:

$$\sqrt{2 \times [1 - \vec{v}]} \tag{3}$$

It is worth noting that straightforward average and standard deviation calculations are insufficient when handling circular values such as dihedral angles.

Derivation of restraints for bulged torso conformation

It has been observed that Rosetta rarely samples the bulged torso conformation when modeling HCDR3 loops [14]. Due to this limitation, coupled with the greater amount of experimentally derived structural data available for bulged torsos than non-bulged torsos and the fact that bulged torsos are more prevalent in the human antibody repertoire, we chose to focus on developing restraints to improve modeling of HCDR3 loops with bulged torsos. Rosetta uses a defined format to read in experimentally derived restraints. We used our measurements to generate dihedral angle restraints following a circular harmonic scoring function. Since the ψ angle measurement of T4 varies by 180 degrees between known bulged torso clusters, this measurement was omitted from our calculated restraints (S2 Fig).

Table 2. Experimentally derived antibodies used to benchmark bulged torso restraints.

PDB ID	HCDR3 Length	Resolution (Å)	Source
1WT5	11	2.10	Humanized
2G75	11	2.28	Human
4G5Z	11	1.83	Human
3QRG	12	1.70	Human
4G6K	12	1.90	Humanized
4LLU	12	2.16	Human
1FVC	13	2.20	Humanized
3HI5	13	2.50	Human
4HFW	13	2.60	Human
4FQH	14	2.05	Human
4NM4	14	2.65	Human
8FAB	14	1.80	Human
3G6A	15	2.10	Human
3TNM	15	1.85	Human
3W9D	15	2.32	Human
1AQK	16	1.84	Human
1DQL	16	2.60	Human
1OM3	16	2.20	Human
1U6A	17	2.81	Human
3AAZ	17	2.20	Humanized
4M5Y	17	1.55	Human
3INU	18	2.50	Human
3QEH	18	2.59	Human
4F58	18	2.49	Human
1HZH	20	2.70	Human
4LKC	22	2.20	Human
1RHH	24	1.90	Human
4FNL	26	2.30	Human

28 high-resolution antibody structures solved by X-ray crystallography were used to benchmark the bulged torso restraints. Each of these antibody structures was solved in the absence of antigen (*i.e.*, apo structures) and all residues in the HCDR3 loops were resolved.

doi:10.1371/journal.pone.0154811.t002

Modeling HCDR3 loops using bulged torso restraints

Following the protocol capture outlined in Supplemental Information, these restraints were used to model and score the HCDR3 loops from 28 benchmark antibodies whose structures had been previously determined by X-ray crystallography (Table 2). These 28 benchmark structures represent HCDR3 lengths from 11 to 26 residues, with a mean length of 16 residues, spanning a range regularly observed in human antibody repertoires that also have a mean HCDR3 length of 16 amino acids [26]. Each of the benchmark antibodies was crystallized in the absence of an antigen (*i.e.*, apo) in order to avoid attempts to model conformations achieved by induced fit with a binding partner.

Restraints function as a penalty during Rosetta’s scoring protocol, *i.e.*, a positive energy value is added when a dihedral angle leaves the allowed range. In this case, models formed with native-like bulged torso dihedral angles would have no (or only a very small) penalty from the restraint term, whereas models that deviated from the bulged torso dihedral angles would be penalized with a positive energy score. When restraints were applied during modeling, we

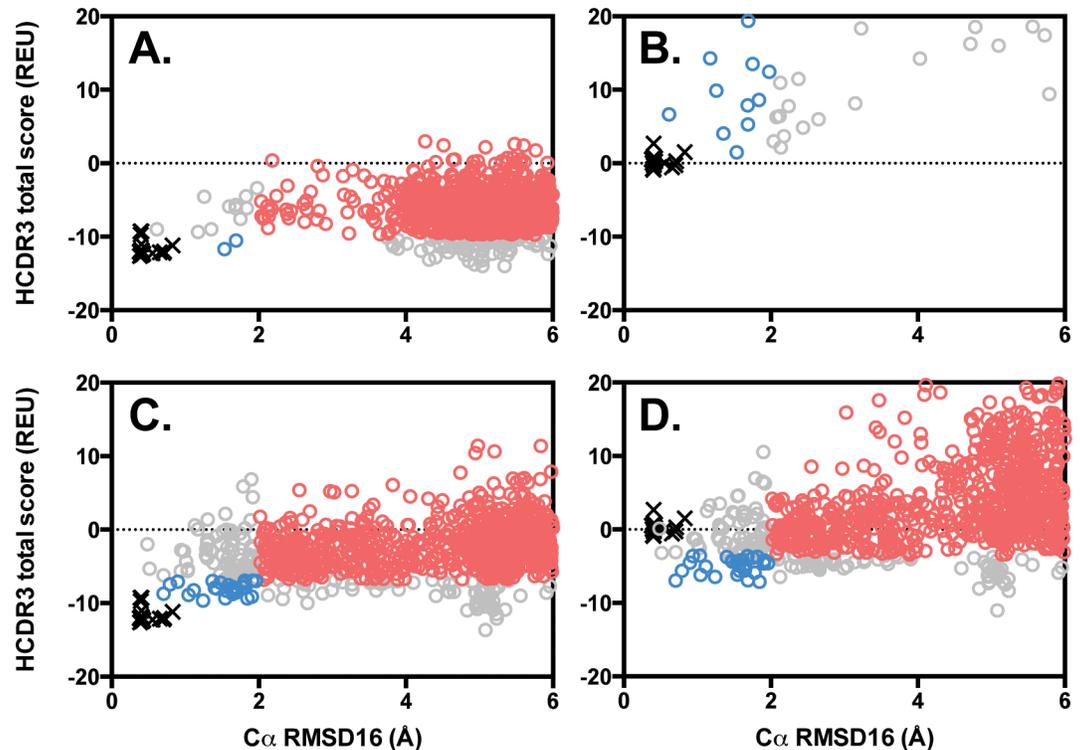


Fig 2. Bulged torso restraints improve native-like HCDR3 sampling and recovery. Using Rosetta LoopModel, 1,000 models of the benchmark antibody 4G5Z (circles) were generated with or without bulged restraints and these models were then scored with or without bulged restraints (panel A, modeled and scored without restraints; panel B, modeled without but scored with restraints; panel C, modeled with but scored without restraints; panel D, modeled and scored with restraints). The native crystal structure 4G5Z was also minimized using Rosetta FastRelax, generating 20 structures (black x's). The total HCDR3 score (in Rosetta Energy Units, or REU) is shown versus the $C\alpha$ root mean square deviation of the HCDR3 loop, normalized to that of a protein loop containing 16 residues (RMSD16, in Å) to the native crystal structure. Models with scores ranked in the top 10% and $RMSD16 \leq 2$ Å have been colored blue, while models with scores ranked below the top 10% and $RMSD16 > 2$ Å have been colored red. Improved native-like HCDR3 sampling is observed as a greater density of low RMSD16 models (blue circles) in comparison to Panel A, while improved model recovery is defined as a greater correlation between RMSD16 and score (colored vs. gray circles) in comparison to Panel A, as seen in panels C and D.

doi:10.1371/journal.pone.0154811.g002

observed a higher density of low-scoring, low-RMSD models (Fig 2C, blue circles, $n = 26$) than when modeling without restraints (Fig 2A, blue circles, $n = 2$). These low-scoring, low-RMSD models are defined as scoring in the top 10% of models, with $C\alpha$ RMSD16 to the native structure of ≤ 2 Å (represented as blue circles, whereas models scoring below the top 10% of models with $C\alpha$ RMSD16 > 2 Å are represented as red circles in Fig 2 and S2 Fig). When restraints were applied during scoring but not during modeling (Fig 2B) we found that the resulting models incur substantial restraint penalties due to non-native-like sampling of the torso domain, however the correlation between score and RMSD16 is improved. During application of this protocol wherein a native structure is unavailable, the ability to identify native-like models by score alone is extremely valuable. When applying restraints during both modeling and scoring, Rosetta generates a model population where an increased number of native-like structures correlate with low scores (Fig 2D, blue circles, $n = 30$; also see S1 File) as compared to experiments modeled and scored without restraints (Fig 2A, blue circles, $n = 2$; also see S1 File). Finally, we found that the application of these restraints results in more models whose backbone structures agree with bulged torso measurements defined in the literature ($n = 719$ with restraints, $n = 33$ without restraints; see S3 Fig) [14,22].

The results of modeling the 28 benchmark HCDR3 loops with or without bulged torso restraints can be found in Figs 3–5. We observed changes in both conformational sampling and in model discretion by score when restraints were applied. To analyze improvements in conformational sampling, models were ranked by RMSD16 to their native structure (Fig 3) and to study changes in scoring discretion the models were ranked by HCDR3 score (Fig 4). Finally, models were clustered using a package called Calibur and the best cluster by average HCDR3 score was analyzed (Fig 5).

Bulged HCDR3 restraints improve native-like conformational sampling

Modeling with bulged torso restraints improved native-like conformational sampling (the number of models with RMSD16 below 2 Å) in 26 out of 28 benchmark cases (Fig 3); in the remaining case of benchmark antibody 1RHH with an HCDR3 loop of 24 residues, no models below 2 Å were observed when modeling with or without restraints, and in the case of 4FNL with an HCDR3 loop of 26 residues, 2 models below 2 Å were observed when modeling without restraints, compared to no models sampled below 2 Å when modeling with restraints. On average, 90 models below 2 Å were generated with restraints, compared to only 12 models below 2 Å without restraints. The best RMSD sampled using bulged torso restraints was below 1 Å in 18 out of 28 cases with restraints, compared to 10 out of 28 cases without restraints. The average difference in the best RMSD sampled was 0.33 Å lower when restraints were applied during modeling. Furthermore, the average RMSD16 of the most native-like 10% of models (when ranked by RMSD16) is below 1 Å in 11 out of 28 cases when restraints are applied, compared to just 1 of 28 cases without restraints, revealing improved depth of high-resolution native-like sampling.

State-of-the-art computational methods to construct loop regions in proteins work reliably until about eight residues, and provide good results from some loops up to twelve residues [18–21]. Beyond this limit, the conformational space often becomes too large to be sampled exhaustively. Many HCDR3 loops are longer and specialized methods are needed to limit the conformational space. Our analyses describe better sampling of native-like structures during modeling of these diverse HCDR3 loops when our torso restraints are used, with qualitative changes in performance observed at 14 and 18 amino acids.

Bulged HCDR3 restraints improve scoring discretion

The ability to identify native-like HCDR3 loops by score when *de novo* modeling using Rosetta is of critical importance. Unfortunately, we found the predictive ability of Rosetta's scoring function in the absence of restraints to be lacking; when ranking models by HCDR3 score, only 2 of 28 benchmark cases resulted in a top-scoring model with RMSD16 < 2 Å (Fig 4). However when restraints were applied, ranking models by score resulted in 7 of 28 cases with an RMSD16 below 2 Å and two of those with RMSD16 below 1 Å (antibody 3QRG, 12 amino acids long and 4FQH, 14 amino acids long). On average, the RMSD16 of the best scoring model improved by 0.84 Å when restraints were used during modeling and scoring. Because restraints improve sampling, there was also a marked improvement in the average RMSD16 of the top 10 models ranked by score; when restraints are applied, the average is below 2 Å in 9 out of 28 cases, but no results below 2 Å were found when restraints were not used. On average, there is an improvement of 1.22 Å in the average RMSD16 of the top 10 models ranked by score. The average rank of the first model below 2 Å is 17 when restraints are applied and in 8 of 28 cases the first-ranking model is below 2 Å, compared to only 2 out of 28 cases resulting in a first-ranking model below 2 Å and an average rank of 82 when restraints are not used. Altogether these analyses reveal that the bulged torso restraints improve scoring discretion of native-like structures, but that further improvement to the scoring of HCDR3 loops is needed [27].

PDB ID	Length	Without Restraints			With Restraints		
		Best RMSD16 sampled	Average RMSD16 of top 10 by RMSD16	Models below 2Å RMSD16	Best RMSD16 sampled	Average RMSD16 of top 10 by RMSD16	Models below 2Å RMSD16
1WT5	11	0.72	1.24	31	0.74	0.84	235
2G75	11	0.37	1.11	23	0.48	0.55	449
4G5Z	11	0.61	1.44	13	0.48	0.80	100
3QRG	12	0.72	0.84	26	0.61	0.69	167
4G6K	12	0.93	1.19	50	0.92	1.09	201
4LLU	12	1.16	1.20	44	0.59	0.75	296
1FVC	13	0.81	1.27	37	0.94	1.00	245
3HI5	13	1.37	1.58	22	0.95	1.40	90
4HFW	13	2.07	2.24	0	0.64	0.80	99
4FQH	14	0.75	1.98	4	0.46	0.70	25
4NM4	14	0.61	2.03	3	0.64	0.91	38
8FAB	14	1.39	1.52	24	0.95	1.16	146
3G6A	15	1.10	1.77	10	0.77	0.87	99
3TNM	15	0.76	1.80	8	0.76	1.12	30
3W9D	15	1.44	1.92	5	1.33	1.53	34
1AQK	16	1.39	1.78	8	0.96	1.24	65
1DQL	16	1.47	1.91	6	1.44	1.73	22
1OM3	16	1.84	2.14	1	1.10	1.39	31
1U6A	17	2.11	2.40	0	0.91	2.17	1
3AAZ	17	1.40	1.92	5	1.26	1.59	24
4M5Y	17	0.88	1.34	19	0.68	0.94	79
3INU	18	1.73	2.15	2	1.35	1.73	16
3QEH	18	1.90	2.30	2	1.76	2.05	4
4F58	18	2.15	2.48	0	1.45	2.02	2
1HZH	20	1.71	2.07	4	1.32	1.80	9
4LKC	22	2.04	2.38	0	0.71	1.66	7
1RHH	24	2.62	2.84	0	2.32	2.57	0
4FNL	26	1.75	2.33	2	2.09	2.25	0
Average	16	1.35	1.83	12	1.02	1.33	90

Fig 3. Torso restraints improve sampling of bulged HCDR3 loops. For each benchmark antibody structure, 1,000 models were generated with or without bulged torso restraints. The number of models below 2 Å RMSD16 to the native structure, the best RMSD16 sampled, and the average RMSD16 of the best 10 models ranked by RMSD16 are provided. For RMSD16-containing cells, blue shading represents $\text{RMSD16} \leq 1 \text{ \AA}$; yellow shading represents RMSD16 between 1 and 2 Å; red represents $\text{RMSD16} > 2 \text{ \AA}$. For cells containing the number of models below 2 Å, blue shading represents ≥ 100 models; yellow shading represents ≥ 10 models; red shading represents fewer than 10 models.

doi:10.1371/journal.pone.0154811.g003

PDB ID	Length	Without Restraints			With Restraints		
		RMSD16 of best scoring model	Average RMSD16 of top 10 by score	Rank of first model <2Å RMSD16	RMSD16 of best scoring model	Average RMSD16 of top 10 by score	Rank of first model <2Å RMSD16
1WT5	11	3.39	3.63	3	3.18	2.31	9
2G75	11	3.63	3.93	40	1.54	1.45	1
4G5Z	11	6.04	5.49	26	5.08	4.32	7
3QRG	12	4.05	2.56	3	0.85	1.15	1
4G6K	12	1.33	2.35	1	1.39	1.35	1
4LLU	12	3.50	3.34	23	2.52	1.99	4
1FVC	13	1.40	2.96	1	2.03	2.05	10
3HI5	13	2.23	3.67	66	1.80	1.72	1
4HFW	13	3.75	3.42	N/A	3.31	2.01	2
4FQH	14	4.26	4.66	27	0.47	1.83	1
4NM4	14	2.59	4.23	15	2.89	1.73	1
8FAB	14	4.21	4.02	4	1.71	1.59	1
3G6A	15	5.18	4.71	44	2.03	2.12	6
3TNM	15	3.14	3.13	2	3.29	3.18	15
3W9D	15	2.89	3.72	12	2.62	2.50	12
1AQK	16	4.48	4.23	25	1.81	1.94	1
1DQL	16	2.09	3.70	2	3.03	2.90	33
1OM3	16	3.48	3.70	281	2.65	2.87	51
1U6A	17	3.14	3.76	N/A	2.93	2.94	183
3AAZ	17	3.11	3.31	7	3.58	2.77	8
4M5Y	17	2.33	2.07	3	2.43	2.54	8
3INU	18	3.47	4.00	656	4.70	3.70	5
3QEH	18	2.48	3.63	176	2.49	3.12	16
4F58	18	5.01	4.75	N/A	4.69	3.28	19
1HZH	20	3.87	3.59	412	2.02	3.26	14
4LKC	22	4.82	4.12	N/A	3.08	3.44	39
1RHH	24	5.78	4.39	N/A	4.32	3.75	N/A
4FNL	26	3.86	3.97	46	3.33	3.00	N/A
Average	16	3.55	3.75	82	2.71	2.53	17

Fig 4. Torso restraints improve recovery of native-like bulged HCDR3 loops. For each benchmark antibody structure, 1,000 models were generated with or without bulged torso restraints. The number of models below 2 Å RMSD16 to the native structure, best RMSD16 sampled, average RMSD16 of the best 10 models ranked by RMSD16, RMSD16 of the best model ranked by Rosetta score, average RMSD16 of the top 10 models ranked by Rosetta score, and the rank of the first model below 2 Å when sorted by Rosetta score are provided. For RMSD16-containing cells, blue shading represents RMSD16 ≤ 1 Å; yellow shading represents RMSD16 between 1 and 2 Å; red represents RMSD16 > 2 Å. For rank-containing cells, blue shading represents rank 1; yellow shading represents ranks 2 to 10; red shading represents ranks > 10.

doi:10.1371/journal.pone.0154811.g004

PDB ID	Length	Without Restraints				With Restraints			
		Best Average Cluster Score	Cluster Size (Rank)	Average Cluster RMSD16 (Rank)	Best RMSD16 in Cluster	Best Average Cluster Score	Cluster Size (Rank)	Average Cluster RMSD16 (Rank)	Best RMSD16 in Cluster
1WT5	11	-10.24	32 (5)	3.99 (7)	2.52	-3.92	178 (3)	1.93 (1)	0.75
2G75	11	-8.88	38 (4)	4.31 (6)	3.79	-4.84	527 (1)	1.49 (1)	0.48
4G5Z	11	-7.91	11 (7)	3.61 (2)	3.10	-0.36	370 (2)	1.98 (1)	0.37
3QRG	12	-8.30	15 (12)	3.13 (5)	2.01	-9.45	132 (1)	0.91 (1)	0.61
4G6K	12	-7.40	124 (2)	1.78 (1)	0.77	-8.13	129 (1)	1.17 (1)	0.85
4LLU	12	-7.95	43 (6)	3.12 (3)	2.62	-5.94	12 (13)	1.83 (4)	1.66
1FVC	13	-11.02	40 (8)	2.68 (2)	1.62	-7.68	404 (1)	1.91 (1)	0.94
3HI5	13	-9.88	64 (4)	1.92 (1)	1.21	-6.94	86 (1)	1.54 (1)	0.84
4HFW	13	-10.36	17 (11)	4.34 (14)	3.58	-5.18	11 (11)	3.53 (7)	3.02
4FQH	14	-11.26	14 (18)	4.64 (11)	3.99	-10.94	18 (12)	1.20 (1)	0.43
4NM4	14	-8.80	14 (12)	4.69 (14)	4.17	-3.68	16 (11)	2.58 (2)	1.80
8FAB	14	-11.70	11 (16)	4.29 (10)	3.61	-8.81	37 (4)	1.66 (2)	1.36
3G6A	15	-11.33	32 (7)	4.11 (9)	3.29	-5.45	238 (1)	2.17 (3)	0.77
3TNM	15	-14.16	10 (18)	2.91 (2)	2.51	-9.52	12 (9)	3.34 (5)	2.95
3W9D	15	-10.26	10 (6)	3.80 (5)	3.46	-6.39	24 (9)	2.81 (3)	2.04
1AQK	16	-10.82	19 (1*)	4.07 (1*)	3.60	-8.98	39 (1)	2.33 (2)	1.90
1DQL	16	-12.76	11 (1*)	2.84 (1*)	2.42	-9.44	66 (3)	2.70 (4)	1.74
1OM3	16	-10.91	19 (8)	3.27 (5)	2.71	-7.07	10 (21)	2.94 (7)	2.18
1U6A	17	-16.88	18 (7)	3.08 (2)	2.42	-3.53	71 (2)	3.29 (5)	2.75
3AAZ	17	-12.46	12 (1*)	4.25 (1*)	3.96	-12.60	29 (4)	2.52 (3)	1.41
4M5Y	17	-17.06	11 (16)	2.02 (1)	1.09	-13.59	113 (2)	1.92 (1)	0.71
3INU	18	-15.69	16 (3)	3.58 (2)	2.95	-13.19	26 (2)	4.26 (10)	3.35
3QEH	18	-12.27	11 (4)	4.33 (4)	4.05	-12.13	11 (1*)	3.57 (1*)	2.99
4F58	18	N/A	N/A	N/A	N/A	-7.89	13 (12)	2.88 (2)	2.40
Average	15	-11.11	28 (9)	3.48 (5)	2.77	-7.54	111 (6)	2.30 (3)	1.54

Fig 5. Cluster analysis of bulged HCDR3 loop modeling. Calibur was used to cluster the 1,000 models generated with or without bulged torso restraints for each antibody, using a threshold of 2.0. Clusters containing less than 1% of the total models were omitted from analysis; models generated for benchmark antibodies 4F58, 1HZH, 4LKC, 1RHH and 4FNL did not produce any large clusters upon analysis (N/A). Average Rosetta score was calculated for each cluster, and the cluster with the lowest average score was selected as the “correct” cluster. The size of this correct cluster (and its rank among cluster sizes), its average RMSD16 to the native structure (and rank among average RMSD16 measurements) are provided. Cells containing rank data are shaded blue if the value represents the top rank, yellow for ranks 2–3, and red for ranks >3; if only one cluster (1*) was found, the cell is shaded gray. For RMSD16-containing cells, blue shading represents RMSD16 ≤ 1 Å; yellow shading represents RMSD16 between 1 and 2 Å; red represents RMSD16 > 2 Å. Values were omitted from column averages if ≤1 cluster was found.

doi:10.1371/journal.pone.0154811.g005

Clustering bulged HCDR3 loop models

Using the clustering package Calibur [28], we analyzed the HCDR3 models generated with and without bulged restraints (Fig 5). Only clusters containing >1% of models (10 or more) were

considered. For models made based on structures with 20 or more amino acids in the HCDR3 loop, no sufficiently large clusters were found. For the other benchmark structures, clusters were sorted by average cluster HCDR3 score, with the lowest average HCDR3 score being chosen as the “correct” cluster. This approach to selecting the “correct” conformation is common when *de novo* modeling HCDR3 loops, as the native structure of the loop is not known outside of benchmark studies. When restraints were used during modeling, the rank of the cluster size (how large a cluster is compared to other clusters) improved in 18 out of 24 cases over experiments where restraints were not used. When restraints were applied during modeling, the average RMSD16 of the correct cluster improved in 21 out of 24 cases. The average RMSD16 for the best cluster by score was top-ranking in 9 out of 24 cases when restraints were applied during modeling, compared to just 3 out of 24 cases when restraints were not used, which reveals the predictive power of our scoring metrics when restraints are applied.

Discussion

There is a growing body of work surrounding canonical structures of antibody CDR loops, first described by Chothia and colleagues and updated as recently as 2011 by the Dunbrack group [9,10]. These groups have shown that five of the six CDR loops take on canonical structures, and that the remaining HCDR3 forms only a few canonical classes of structure in its torso domain. Our work builds upon this background, and has led to the development of knowledge-based structural restraints from available crystal structures of HCDR3 loops with bulged torsos. We have shown that these restraints can be used to restrict the sampling space Rosetta searches during *de novo* loop modeling, limiting the torso domain to the ϕ and ψ angles of these residues that have been experimentally observed. These torso restraints improve native-like structure sampling and score-based differentiation of native-like HCDR3 models. We have also shown that such structural restraints improve Rosetta’s ability to model longer HCDR3 loops than previously possible, extending the range of the technique to cover more biologically relevant HCDR3 loop lengths.

While this study focuses on benchmarking new knowledge-based restraints against antibodies whose structures have been experimentally determined, the true value of these restraints is in their ability to improve *de novo* antibody modeling. Such antibody structural predictions are a more rapid approach than experimental structural techniques, and can improve our understanding of host-pathogen interactions, provide insight into mechanisms of viral infection, and may lead to new monoclonal antibody therapeutics or vaccine candidates. Combined with our prior understanding of canonical CDR loops, which had made it possible to homology model much of the functional surface of the antibody (the “paratope”) using Rosetta, we can now predict the remaining HCDR3 which is critical in many antibody-antigen interactions. The central dogma of structural biology, that structure dictates function, lets us expect that improved accuracy in modeling HCDR3 will lead to improved accuracy in modeling antibody/antigen interactions which in turn leads to improved prediction of antibody function. We recognize that further experiments would be needed to prove this. Finally, upcoming advances in antibody sequencing, including the ability to sequence endogenously paired heavy and light chains, will provide the last critical insight in antibody modeling; we must now come to understand restrictions at the heavy and light chain interface that alter the paratope, and incorporate such restrictions into our structural predictions.

Although we have applied this approach to improving human antibody modeling, we recognize that this approach to structural restraint development is applicable to many other protein families in which structurally diverse surface loops with key functional importance are supported upon more structurally restricted framework regions [27]. Obvious examples include

proteins with the PDZ domain and peptidase C1 domain protein families, which were found to use bulged HCDR3-like loops to recognize and bind their substrates [14]. Finally, we have shown that knowledge-based structural restraints can be calculated easily and applied to improve modeling of novel loops not previously solved by experimental techniques, provided enough experimentally derived structural data is available for framework regions of functional loops in other protein families, and that canonical classes of those regions can be defined.

Materials and Methods

Calculating bulged and non-bulged torso dihedral angles

A collection of antibody heavy chain variable domains was manually curated from the PDB, building upon a published list [10] (S2 File). The torso residues of these structures were extracted from the PDB files and were clustered using Rosetta Cluster with a cluster radius of 2 Å to separate bulged and non-bulged antibody torsos. ϕ and ψ dihedral angles of the seven torso residues were found using Biopython [29], with average and approximate standard deviation calculated using Eqs 1 and 2 (Table 1).

Generating HCDR3 loop models

The complete protocol for generating the HCDR3 loop models using Rosetta is described in S3 File and example file input and output is provided in S4 File. In brief, structure files for each benchmark antibody were downloaded from the PDB and were cleaned such that only a single variable domain remained. Input files for loop modeling were generated with the assistance of a suite of python scripts, and fragments were selected using the fragment picker. Centroid loop modeling was accomplished using cyclic coordinate descent (CCD), followed by a kinematic closure (KIC) full-atom refinement [30–32].

HCDR3 torso sequence analysis

Sequences of the seven torso residues were taken from each of the PDB files of the bulged antibody torso cluster found above and used to generate a WebLogo using the default webserver settings [33] (S1 Fig). A second WebLogo was generated using the sequences of the torso residues taken from the IMGT human V_H and J_H gene segments [34] (S1 Fig).

Supporting Information

S1 Fig. Bulged torso structures share similar sequences, which are germline-encoded. Previous studies identified a sequence motif in bulged torso structures, which are formed primarily via a side-chain interaction between either Arg or Lys (R/K) at T2 and Asp (D) at T6. A consensus sequence from bulged torsos culled from the PDB shows the prevalence of these residues at these positions (panel A). These residues are germline-encoded, as observed in a consensus sequence of the V_H and J_H gene segments that contribute to the torso domain (panel B).

(TIF)

S2 Fig. Average ϕ and ψ angles observed for each torso residue in known bulged and non-bulged clusters. North et al. [10] defined seven canonical torso conformations from experimentally-determined antibody structures. Two of these clusters are considered bulged (H3-anchor-1 and H3-anchor-3; blue) and two are considered non-bulged (H3-anchor-2 and H3-anchor-5; red). ϕ and ψ angles are well defined for both bulged and non-bulged HCDR3 torso residues. Bulged and non-bulged torsos are differentiated by their ψ angle at T6. The ψ

angle at T4 is bimodal for both bulged and non-bulged HCDR3 torsos, with ~180 degrees separating the two clusters within each definition.

(TIF)

S3 Fig. Bulged torso restraints improve sampling of HCDR3 torso angles. Using Rosetta LoopModel, 1,000 models of the benchmark antibody 4G5Z were generated without (red) or with (blue) bulged restraints. The τ_{101} angle and α_{101} dihedral angle defined by Weitzner et al. [14] were calculated for each model. Gray regions of the plot denote $\pm 3\sigma$ of the mean angles calculated for bulged HCDR3 torsos by Weitzner et al. [14]. Improved recovery of bulged torsos was observed as a greater density of points in the center gray region when restraints were applied ($n = 719$), versus when no restraints were applied ($n = 33$).

(TIF)

S1 File. Bulged torso restraints improve native-like HCDR3 sampling and recovery. As in Fig 2, 1,000 models of each benchmark antibody were generated and scored with or without bulged restraints using Rosetta LoopModel (comparable to Fig 2A and 2D). Models with scores ranked in the top 10% and $\text{RMSD}_{16} \leq 2 \text{ \AA}$ have been colored blue, while models with scores ranked below the top 10% and $\text{RMSD}_{16} > 2 \text{ \AA}$ have been colored red. The native crystal structure was also minimized using Rosetta FastRelax, generating 20 structures (black x's). The total HCDR3 score vs. the HCDR3 C α RMSD₁₆ to the native crystal structure is shown.

(PDF)

S2 File. HCDR3 definitions file. This file contains two comma separated value tables. The first table represents the non-bulged antibody structures used to calculate dihedral angle values, and lists the PDB file, chain ID, HCDR3 start residue and HCDR3 end residue when each chain in the PDB file has been renumbered sequentially starting from 1. The second file lists these values for the bulged antibody structures used to calculate the dihedral angle values.

(TXT)

S3 File. Rosetta protocol. A complete protocol has been provided, including Rosetta version number, for individuals who wish to utilize our methodology.

(PDF)

S4 File. Rosetta protocol capture. This archive contains example input and output files needed to run the Rosetta protocol described in S3 File.

(ZIP)

Author Contributions

Conceived and designed the experiments: JAF JRW JEC JM. Performed the experiments: JAF JRW AC. Analyzed the data: JAF JKL JRW AC JEC JM. Contributed reagents/materials/analysis tools: JAF JKL AC. Wrote the paper: JAF JEC JM. Independently verified the protocol: JKL.

References

1. Hashiguchi T, Fusco ML, Bornholdt ZA, Lee JE, Flyak AI, Matsuoka R, et al. Structural basis for Marburg virus neutralization by a cross-reactive human antibody. *Cell*. 2015; 160: 904–912. doi: [10.1016/j.cell.2015.01.041](https://doi.org/10.1016/j.cell.2015.01.041) PMID: [25723165](https://pubmed.ncbi.nlm.nih.gov/25723165/)
2. Hong M, Lee PS, Hoffman RMB, Zhu X, Krause JC, Laursen NS, et al. Antibody recognition of the pandemic H1N1 Influenza virus hemagglutinin receptor binding site. *J Virol*. American Society for Microbiology; 2013; 87: 12471–12480. doi: [10.1128/JVI.01388-13](https://doi.org/10.1128/JVI.01388-13) PMID: [24027321](https://pubmed.ncbi.nlm.nih.gov/24027321/)
3. Whittle JRR, Zhang R, Khurana S, King LR, Manischewitz J, Golding H, et al. Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. *Proc Natl*

- Acad Sci USA. National Acad Sciences; 2011; 108: 14216–14221. doi: [10.1073/pnas.1111497108](https://doi.org/10.1073/pnas.1111497108) PMID: [21825125](https://pubmed.ncbi.nlm.nih.gov/21825125/)
4. Li Y, O'Dell S, Walker LM, Wu X, Guenaga J, Feng Y, et al. Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01. *J Virol. American Society for Microbiology*; 2011; 85: 8954–8967. doi: [10.1128/JVI.00754-11](https://doi.org/10.1128/JVI.00754-11) PMID: [21715490](https://pubmed.ncbi.nlm.nih.gov/21715490/)
 5. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci USA. National Acad Sciences*; 2009; 106: 20216–20221. doi: [10.1073/pnas.0909775106](https://doi.org/10.1073/pnas.0909775106) PMID: [19875695](https://pubmed.ncbi.nlm.nih.gov/19875695/)
 6. Trepel F. Number and distribution of lymphocytes in man. A critical analysis. *Klin Wochenschr.* 1974; 52: 511–515. PMID: [4853392](https://pubmed.ncbi.nlm.nih.gov/4853392/)
 7. Arnaout R, Lee W, Cahill P, Honan T, Sparrow T, Weiland M, et al. High-resolution description of antibody heavy-chain repertoires in humans. Reindl M, editor. *PLoS ONE. Public Library of Science*; 2011; 6: e22365. doi: [10.1371/journal.pone.0022365](https://doi.org/10.1371/journal.pone.0022365) PMID: [21829618](https://pubmed.ncbi.nlm.nih.gov/21829618/)
 8. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med. NIH Public Access*; 2009; 1: 12ra23. PMID: [20161664](https://pubmed.ncbi.nlm.nih.gov/20161664/)
 9. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J Mol Biol.* 1998; 275: 269–294. doi: [10.1006/jmbi.1997.1442](https://doi.org/10.1006/jmbi.1997.1442) PMID: [9466909](https://pubmed.ncbi.nlm.nih.gov/9466909/)
 10. North B, Lehmann A, Dunbrack RL. A new clustering of antibody CDR loop conformations. *J Mol Biol.* 2011; 406: 228–256. doi: [10.1016/j.jmb.2010.10.030](https://doi.org/10.1016/j.jmb.2010.10.030) PMID: [21035459](https://pubmed.ncbi.nlm.nih.gov/21035459/)
 11. Finn JA, Crowe JE. Impact of new sequencing technologies on s'tudies of the human B cell repertoire. *Curr Opin Immunol.* 2013; 25: 613–618. doi: [10.1016/j.coi.2013.09.010](https://doi.org/10.1016/j.coi.2013.09.010) PMID: [24161653](https://pubmed.ncbi.nlm.nih.gov/24161653/)
 12. Fanning LJ, Connor AM, Wu GE. Development of the immunoglobulin repertoire. *Clin Immunol Immunopathol.* 1996; 79: 1–14. PMID: [8612345](https://pubmed.ncbi.nlm.nih.gov/8612345/)
 13. Tonegawa S. Somatic generation of antibody diversity. *Nature.* 1983; 302: 575–581. PMID: [6300689](https://pubmed.ncbi.nlm.nih.gov/6300689/)
 14. Weitzner BD, Dunbrack RL, Gray JJ. The origin of CDR H3 structural diversity. *Structure.* 2015; 23: 302–311. doi: [10.1016/j.str.2014.11.010](https://doi.org/10.1016/j.str.2014.11.010) PMID: [25579815](https://pubmed.ncbi.nlm.nih.gov/25579815/)
 15. Kaufmann KW, Lemmon GH, Deluca SL, Sheehan JH, Meiler J. Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry. American Chemical Society*; 2010; 49: 2987–2998. doi: [10.1021/bi902153g](https://doi.org/10.1021/bi902153g) PMID: [20235548](https://pubmed.ncbi.nlm.nih.gov/20235548/)
 16. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol.* 1997; 268: 209–225. doi: [10.1006/jmbi.1997.0959](https://doi.org/10.1006/jmbi.1997.0959) PMID: [9149153](https://pubmed.ncbi.nlm.nih.gov/9149153/)
 17. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins. Wiley Subscription Services, Inc., A Wiley Company*; 2004; 55: 656–677. doi: [10.1002/prot.10629](https://doi.org/10.1002/prot.10629) PMID: [15103629](https://pubmed.ncbi.nlm.nih.gov/15103629/)
 18. Weitzner BD, Kuroda D, Marze N, Xu J, Gray JJ. Blind prediction performance of RosettaAntibody 3.0: grafting, relaxation, kinematic loop modeling, and full CDR optimization. *Proteins.* 2014; 82: 1611–1623. doi: [10.1002/prot.24534](https://doi.org/10.1002/prot.24534) PMID: [24519881](https://pubmed.ncbi.nlm.nih.gov/24519881/)
 19. Sircar A, Kim ET, Gray JJ. RosettaAntibody: antibody variable region homology modeling server. *Nucleic Acids Res. Oxford University Press*; 2009; 37: W474–9. doi: [10.1093/nar/gkp387](https://doi.org/10.1093/nar/gkp387) PMID: [19458157](https://pubmed.ncbi.nlm.nih.gov/19458157/)
 20. Almagro JC, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, Butenhof K, et al. Antibody modeling assessment. *Proteins. Wiley Subscription Services, Inc., A Wiley Company*; 2011; 79: 3050–3066. doi: [10.1002/prot.23130](https://doi.org/10.1002/prot.23130) PMID: [21935986](https://pubmed.ncbi.nlm.nih.gov/21935986/)
 21. Almagro JC, Teplyakov A, Luo J, Sweet RW, Kodangattil S, Hernandez-Guzman F, et al. Second antibody modeling assessment (AMA-II). *Proteins.* 2014; 82: 1553–1562. doi: [10.1002/prot.24567](https://doi.org/10.1002/prot.24567) PMID: [24668560](https://pubmed.ncbi.nlm.nih.gov/24668560/)
 22. Shirai H, Kidera A, Nakamura H. Structural classification of CDR-H3 in antibodies. *FEBS Lett.* 1996; 399: 1–8. PMID: [8980108](https://pubmed.ncbi.nlm.nih.gov/8980108/)
 23. Shirai H, Ikeda K, Yamashita K, Tsuchiya Y, Sarmiento J, Liang S, et al. High-resolution modeling of antibody structures by a combination of bioinformatics, expert knowledge, and molecular simulations. *Proteins.* 2014; 82: 1624–1635. doi: [10.1002/prot.24591](https://doi.org/10.1002/prot.24591) PMID: [24756852](https://pubmed.ncbi.nlm.nih.gov/24756852/)
 24. Whitelegg NR, Rees AR. WAM: an improved algorithm for modelling antibodies on the WEB. *Protein Eng.* 2000; 13: 819–824. PMID: [11239080](https://pubmed.ncbi.nlm.nih.gov/11239080/)

25. Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol*. 2003; 27: 55–77. PMID: [12477501](#)
26. Briney BS, Willis JR, Hicar MD, Thomas JW, Crowe JE. Frequency and genetic characterization of V (DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology*. 2012; 137: 56–64. doi: [10.1111/j.1365-2567.2012.03605.x](#) PMID: [22612413](#)
27. Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem*. Annual Reviews; 2008; 77: 363–382. doi: [10.1146/annurev.biochem.77.062906.171838](#) PMID: [18410248](#)
28. Li SC, Ng YK. Calibur: a tool for clustering large numbers of protein decoys. *BMC Bioinformatics*. BioMed Central Ltd; 2010; 11: 25. doi: [10.1186/1471-2105-11-25](#) PMID: [20070892](#)
29. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. Oxford University Press; 2009; 25: 1422–1423. doi: [10.1093/bioinformatics/btp163](#) PMID: [19304878](#)
30. Wang C, Bradley P, Baker D. Protein-protein docking with backbone flexibility. *J Mol Biol*. 2007; 373: 503–519. doi: [10.1016/j.jmb.2007.07.050](#) PMID: [17825317](#)
31. Canutescu AA, Dunbrack RL. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci*. Cold Spring Harbor Laboratory Press; 2003; 12: 963–972. doi: [10.1110/ps.0242703](#) PMID: [12717019](#)
32. Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods*. Nature Publishing Group; 2009; 6: 551–552. doi: [10.1038/nmeth0809-551](#) PMID: [19644455](#)
33. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. Cold Spring Harbor Lab; 2004; 14: 1188–1190. doi: [10.1101/gr.849004](#) PMID: [15173120](#)
34. Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res*. Oxford University Press; 2005; 33: D256–61. doi: [10.1093/nar/gki010](#) PMID: [15608191](#)