Chapter 2

Computer-Assisted Modeling of Antibody Variable Domains

Oscar H.P. Ramos

Abstract

Antibody modeling is an interesting option to gain structure–function insights and to allow rational engineering of these molecules in the absence of experimental data. Among a diversity of algorithms, software packages, and specialized Web servers, the protocol described here presents the following main features: (1) nonautomatic modeling process guided by direct application of antibody modeling principles; (2) local generation of molecular models using free software which can be used in most common operational systems; and (3) the resulting model quality is comparable to models generated by Web servers which represent the current standard of antibody modeling. Briefly, hybrid models of heavy- and light-chain variable domains are separately built by grafting segments from homologous templates (framework regions and complementarity-determining regions). Next, hybrid models are mutated to comply with the target's sequence and associated by fitting into a template structure that closely matches the predicted packing angle for the target variable domains. After a few cycles of energy minimization the model can be submitted to CDR-H3 optimization or its quality can be directly assessed.

Key words: Antibody, Variable domains, Modeling, Antigen combining site, Structure prediction

Abbreviations

VD	Variable domain
FR	Framework region
CDR	Complementarity-determining region
HV	Antibody heavy-chain variable domain
LV	Antibody light-chain variable domain

1. Introduction

The increasing relevance of antibodies in biotechnology and therapeutics experienced in the last decades correlates with progress in recombinant protein expression, increasing amount of sequence/structure data, and rising of bioinformatics. Since it is recognized

Patrick Chames (ed.), *Antibody Engineering: Methods and Protocols, Second Edition*, Methods in Molecular Biology, vol. 907, DOI 10.1007/978-1-61779-974-7_2,

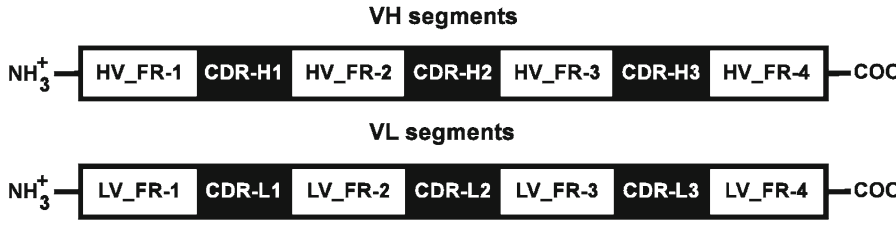

Fig. 1. Schematic representation of heavy-chain variable domain (VH) and light-chain variable domain (VL). Segments corresponding to conserved framework regions (FRs, *white background*) and hypervariable complementarity-determining regions (CDRs, *black background*). *L* indicates antibody light chain; *H* indicates antibody heavy chain. *Numbers* indicate consecutive positions of hypervariable loops in the antibody sequence from N- to C-terminus of variable domains.

that structure is intrinsically related to molecular properties and function, at the crossroad of these scientific advances emerges antibody engineering, which encompasses methods to predict conformations of antibody variable domains (VDs) that are prone to happen in life-compatible environments. Concerning predictive approaches, two main streams are available: comparative homology modeling (see Note 1) and *ab initio* (see Note 2) predictions. Regarding antibody variable domains, homology modeling is still considered the method of choice since most of the sequence is highly conserved and hundreds of high-resolution structure are available (see Note 3). The conserved sequence mainly corresponds to the framework regions (FRs, Fig. 1).

Furthermore, analysis of antibody sequence and structure has provided "canonical rules" that frequently allow to transpose sequence information of five hypervariable complementarity-determining regions (CDR-L1, L2, CDR-L3, CDR-H1, and CDR-H2) into structural information based on key residues present on CDR itself and on framework regions. In opposition to the cited hypervariable loops, nature has reserved significantly higher diversity for the last one, CDR-H3, making it more difficult to establish canonical rules and constituting an important obstacle for generating more reliable models of antibody variable domains (see Note 4). Ironically, CDR-H3 is located at the center of antigen binding site and its conformation is generally crucial for antigen recognition. Thus, high-resolution modeling of CDR-H3 is very desirable in order to allow good predictions of complex formation with antigens.

Some state-of-the-art algorithms implemented in antibody modeling servers have demonstrated their values, such as WAM (1) (http://antibody.bath.ac.uk/, see Note 5), RosettaAntibody (2) (http://antibody.graylab.jhu.edu/, see Note 6), and PIGS (3) (http://arianna.bio.uniroma1.it/pigs/, see Note 7). Their interest lies on the fact that they require almost no prior modeling knowledge and they generally provide good results at the end (see Note 8), so

they are frequently considered as first choice and the reader is encouraged to try them. Nevertheless, some particular sequences fail to be automatically processed or, sometimes, it takes more than one week for job completion (see Note 9). In these cases, when the researcher looks for more detailed control of the modeling process (see Note 10) or simply wants to do it locally, other options are available including a considerable amount of commercial and non-commercial software (see Note 11).

The modeling protocol described here is intended for local generation of models with quality that is comparable to automatized servers (see Note 12). It stands out among a great diversity of available approaches because of its direct application of antibody modeling concepts and its broad applicability (see Note 13). By attempting to compile detailed step-by-step instructions, it is expected to be applicable even by researchers with low modeling or bioinformatics knowledge. Briefly, structural relevant elements are identified from sequences analysis of variable domains, namely, FRs and CDRs (Fig. 1). Conformation templates are searched and selected for structure elements based on sequence similarity and resolution. For CDR elements, the length and the respect of canonical rules also constitute criteria of primary relevance and should be considered. Afterwards, using the selected templates, chimeric domains are independently constructed for VL and VH which are then mutated to comply the target's sequence. Finally, VH and VL primary models are packaged and the resulting structure is submitted to a few steps of energy minimization. The primary model can be submitted to further refinements, such as simulated annealing of the CDR-H3, or its quality can be directly verified (Fig. 2). The required software are free and can run on Linux, Mac, or Windows (see Note 14).

2. Materials

Antibody VH and VL sequences to be modeled.

Computer connected to the Internet.

Swiss-PdbViewer (SPDBV)—Deep Viewer application.

3. Methods

3.1. Creating an Organized Structure of Directories for Antibody Modeling

1. Create an organized directory structure intended for the modeling process inside a new directory named after the antibody to be modeled (see Note 15).

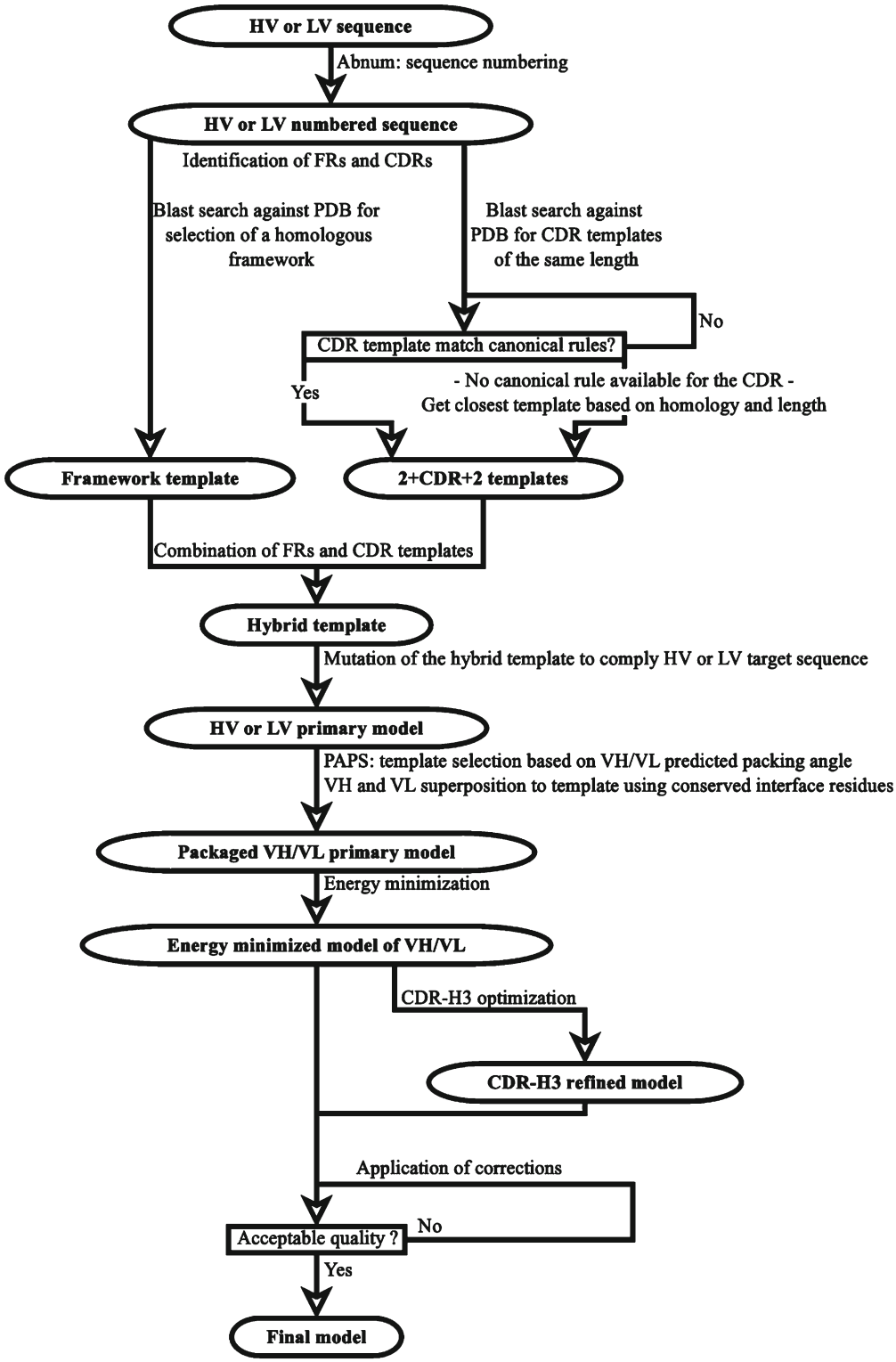

Fig. 2. Flowchart illustrating the antibody modeling protocol described here.

3.2. Sequence Analysis of Antibody VDs

Numbering schemes provide useful standardization of residues' position among homologous sequences, making analysis and comparison readily accessible. In past years, several VD numbering schemes were proposed, such as Kabat (4) (based exclusively on antibody sequences), Chothia (5, 6) (based on antibody three-dimensional structures), a modified Chothia scheme (aiming to correct insertion sites in framework regions) (7), and AHo (8) (based on three-dimensional structure of antibodies and non-antibodies immunoglobulin VDs). Considering the specific objective of modeling antibody VDs, the Chothia scheme is particularly convenient (see Note 16).

1. Number VH and VL sequences by Chothia scheme using Abnum server (7) available at http://www.bioinf.org.uk/abs/abnum/ and save VH and VL numbered sequences in the "Info" directory for future reference.
2. Analyze numbered sequences in order to identify four FRs and three CDRs for each chain. Annotate corresponding segments (see Note 17) using CDR definitions presented in Table 1.

3.3. Homology Search

1. Search for homologous FR and CDR (see Note 18) sequences using protein BLAST (9) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against Protein Data Bank (pdb) database.

3.4. VH and VL Primary Model Building Using SPDBV (10)

1. Select a template for each segment (see Note 19). Concerning CDR segments, it is preferable to select templates that belong to the same canonical class (see Tables 2 and 3: Potential residues related to canonical CDR conformations).
2. Open SPDBV (see Note 20) and import PDB file (see Notes 21, 22, 23) corresponding to each selected template.

   Import PDB file: "File" => "Import …", enter PDB code, and click on PDB file.

Table 1
CDR definitions in Chothia numbering scheme

CDR	Residues[a]
L1	L24…L34
L2	L50…L56
L3	L89…L97
H1	H26…H32
H2	H52…56
H3	H95…102

[a]Residue positions correspond to Chothia numbering scheme

Table 2
LV canonical classes of CDR conformations based on potentially related key residues

L1																
Class[A]	1	2	3	4	5	6	5λ	6λ	7λ							
Class[B]	10A	11A	17A	16A/B/C	15A	12A	13A	14A	14B	11B	14C/D	14E	14F	15B	16B	16C
Length	10	11	17	16	15	12	13	14	14	11	14	14	14	15	16	16
L2	I^{1}	$I^{3}(L^{2})$	I^{3}	$V^{3}I^{2}L^{5}$	I^{2}	N^{5}	S^{2}	S^{2}	$A^{2}Q^{2}$		S^{2}	S^{2}	S^{2}	I^{2}	V^{2}	V^{2}
L25	$A^{1}(S^{1})$	A^{3}	(S^{3})	$S^{3}P^{5}$	A^{2}	A^{5}	G^{2}	(G^{2})	(S^{2})	A^{2}				A^{2}	S^{2}	S^{2}
L29	$V^{1}(I^{1}L^{1})$	$V^{3}I^{3}(L^{2})$	$(V^{5}I^{2})L^{3}$	$I^{5}L^{3}$	V^{2}	V^{5}		N^{2}	$A^{2}T^{2}$	P^{2}	D^{2}	D^{2}	N^{2}	V^{2}	L^{2}	I^{2}
L30					D^{2}		I^{2}	I^{2}	V^{2}		V^{2}	V^{2}	I^{2}	$D^{2}S^{2}$		
L30D			$(E^{5}Q^{5}S^{5})$	G^{5}											G^{2}	G^{2}
L33	$M^{3}L^{1}$	$V^{4}I^{5}L^{3}$	L^{3}	$F^{5}L^{3}$	L^{2}	L^{5}	V^{2}	V^{2}	A^{2}	A^{2}	V^{2}	I^{2}	V^{2}	$M^{2}I^{2}$	L^{2}	L^{2}
L71	$Y^{1}(F^{1})$	$Y^{3}F^{4}$	$(Y^{2})F^{3}$	F^{3}	F^{2}	$Y^{5}(V)$	A^{2}	A^{2}	A^{2}	V^{2}	A^{2}	A^{2}	A^{2}	F^{2}	F^{2}	F^{2}
L2																
Class[A]	1															
Class[B]	7A															
Length	7															
L48	$I^{1}V^{1}$															
L64	G^{1}															

L3													
Class[A]	1	2	3	4	5	4λ	5λ						
Class[B]	9A	9B	8A	7A	10A	9C	11A	8B	9D	9E	9F	10B/C/D	11B
Length	9	9	8	7	10	9	11	8	9	9	9	10	11
L90	$Q^{1}N^{1}H^{3}$	Q^{3}	Q^{3}	Q^{5}	Q^{5}	$L^{2}S^{2}$	A^{2}	Q^{2}	L^{2}	Q^{2}	A^{2}	S^{2}	A^{2}
L94	$D^{2}Y^{2}T^{2}V^{2}L^{2}H^{2}$ $N^{2}I^{2}W^{2}P^{2}S^{2}$	P^{1}				$N^{2}S^{2}$	$(D^{2}N^{2}G^{2})$		N^{2}	Y^{2}	S^{2}		
L95	P^{1}	L^{1}	(P^{3})			$H^{2}L^{2}$			$H^{2}L^{2}$	P^{2}	A^{2}		
L96	$P^{2}L^{2}Y^{2}$ $R^{2}I^{2}W^{2}F^{2}$	I^{2}			$I^{2}(P^{5})$	$W^{2}R^{2}$			W^{2}	L^{2}	S^{2}	F^{2}	
L97	$T^{5}(S^{5})$	T^{5}		(S^{5})	T^{5}	$I^{2}V^{2}$	$V^{2}G^{2}$		V^{2}	T^{2}	I^{2}	V^{2}	V^{2}

[A] Chothia and cols class name
[B] Martin and Thornton class name
[1] Chothia and Lesk, 1987 (5)
[2] Martin and Thornton, 1996 (29)
[3] Chothia et al. 1989 (30)
[4] Barré et al. 1994 (31)
[5] Tomlinson et al. 1995 (32)
(X): Residue X was not found to be relevant based on Martin and Thornton analyses
Classes that could not be distinguished based on allowed residues for Chothia proposed key positions were assembled into a unique class (CDR-L1 16A/B/C and 14 C/D; CDR-L3 10B/C/D)

Table 3
HV canonical classes of CDR conformations based on potentially related key residues

H1							
Class[A]	1	2	3				
Class[B]	10A/B/C/D	11A	12A				
Length	10	11	12				
H24	$A^6V^6T^6G^6S^4$	$V^6(F^6)$	$V^6F^6(G^6)$				
H26	G^1	(G^6)	G^6				
H27[&]	$Y^1F^1G^6T^4$	$Y^6F^6G^6$					
H29	$I^6F^1L^3(V^6)S^2$	$I^6(L^6)$	$I^6L^6(V^6)$				
H34	$I^3V^3M^1W^6(T^6L^6)$	$W^6(C^6)$	W^6V^6				
H94	$R^1K^1G^3S^4$ $H^4N^4(T^6A^6)$	(H^6R^6)	(H^6R^6)				
H2							
Class[A]	1	2	3	4			
Class[B]	9A	10A/C/D	10B	12A	10C	10F	12B
Length	9	10	10	12	10	10	12
H52a[&]		$P^3T^3A^6$	$D^3P^3S^3$	N^3			
H54		$N^2S^2T^2K^2D^2G^2$	$S^3G^1(N^1D^1)$	$(K^3S^6N^3)$	N^2G^2	Y^2	
H55	$G^1(D^1)$	$(G^6S^3T^3)$	(G^3S^3)	(Y^3)			
H71	$R^6K^6V^6(I^6)$	$V^2A^3L^3(T^6)$	R^3	R^3	A^2R^2	V^2	

[A] Chothia and cols class name
[B] Martin and Thornton class name
[1] Chothia and Lesk, 1987 (5)
[2] Martin and Thornton, 1996 (29)
[3] Chothia et al. 1989 (30)
[4] Barré et al. 1994 (31)
[6] Chothia et al. 1992 (33)
[&] Nonrelevant position according to Martin and Thornton, 1996
(X) not relevant residue based on Martin and Thornton analyses
Classes that could not be distinguished based on allowed residues for Chothia proposed key positions were assembled into a unique class (CDR-H1 10A/B/C/D, CDR-H2 10A/C/D)

3. Select template residues matching target sequence segment, invert selection, and remove non-matching residues.

   Select template residues: "Edit" => "Find sequence …", copy template sequence matching to the target segment, and paste into the sequence field. Click OK. For noncontinuous selections, such as framework regions, "Wind =>Control panel" can be used.

   Invert selection: "Select" => "Inverse Selection".

   Remove residues: "Build" => "Remove Selected Residues …".

4. Use CDR anchor residues to fit these segments into FR template using backbone atoms. After fitting all CDRs into FR segments, create a merged layer comprising FRs and CDRs without anchor residues. Save the merged layer.

   Select anchor residues and fit atoms: Select anchor and corresponding FR residues using "Control panel". Then fit segments

using backbone atoms: menu "Fit" => "Fit Molecules (from selection)". Choose FR layer as reference and CDR layer as mobile.

Create a merged layer: Select all residues in the FR template layer and CDR residues (do not include anchor residues) and "Edit" => "Create Merged Layer from Selection (by layer)".

Save the merged layer.

Save merged layer: Be sure that created merged layer (_merge_) is the active layer (its name should appear in the top of the control panel). Save merged layer: "File" => "Save" => "Current Layer …" and choose destination file name and directory.

5. Close all layers.

   Close all layer: "File" => "Close All Layers".

6. Using a text editor of your choice, modify the saved file to put atoms corresponding to CDR residues in their correct sequence order (see Note 24).

7. Open the modified PDB file, select all residues, and renumber residues.

   Open PDB file: "File" => "Open PDB File …", and select location and PDB file to be loaded.

   Select all residues: "Select" => "All".

   Renumber residues: "Edit" => "Rename Current Layer …". In the field "Rename Chain of Selected Groups" type H or L (depending on which VD to be modeled, VH or VL, respectively). In the "Renumber Selected Groups from" type 1.

8. Load VH or VL sequence into the workspace and check if the merged template and the loaded sequence are correctly aligned.

   Load VH or VL sequence: "SwissModel" => "Load Raw Sequence from Amino Acids …" and select a file containing VH or VL sequence in FASTA format.

   Inspect and modify the alignment: "Wind" => "Alignment". If necessary, the alignment can be edited by selecting the residues to be displaced and using "Ctrl" + "Space" keys to move it to the right (N- to C-terminus) or "Ctrl" + "Backspace" keys to move it to the left (C- to N-terminus).

9. Set the merged template as modeling reference, create a primary model for the loaded VH or VL sequence (VH-PM or VL-PM) by mutation of the merged/hybrid template, and save the primary model.

   Set merged template as modeling reference: Make sure that the merged template layer is the active layer by selecting its name in the drop-down list available at the top of "Control panel". Execute the following menu commands "SwissModel" => "Set Current

Layer as Modelling Reference", then "Wind" => "Layers info", and click over the cell corresponding to the intersection of the merged template line and the "mdl" column (now, a "v" letter should be exhibited).

Create primary model: "SwissModel" => "Build Preliminary Model for Selected Layers …".

Save the model: Be sure that the created model is the active layer (the name imported from the fasta sequence file should appear in the top of the control panel) and execute "File" => "Save" => "Current Layer …" and choose destination file name and directory.

3.5. VH and VL Association Using Predicted Packing Angle

1. Using PAPS server find the template that most closely matches the predicted packing angle for the association of target VH and VL (11) and download the proposed PDB file.

   Find suitable template: Access "www.bioinf.org.uk/abs/paps/" and enter the sequences for VH and VL (see Note 25).

2. Load the downloaded template, VH-PM and VL-PM structures into SPDBV workspace (*see* "Open PDB file" section of Subheading 3.4, step 7).

3. Select template residues H36–H39 and H89–H92 and the corresponding residues in VH-PM (see Notes 26 and 27). Fit selected residues using backbone atoms.

   Fit residues: "Fit" => "Fit Molecules (from selection)". Use backbone atoms only. Assume template layer as reference structure (fixed) and VH-PM layer as mobile structure.

4. Apply the same procedure to VL using template residues L35–L38 and L85–L88.

5. Dismiss the template structure, select all variable domain residues, and create a merged layer comprising VH and VL.

   Dismiss the template structure: Select the template as an active layer (*see Note 23*) *and execute "File" => "Close".*

   Select all residues: Select VH-PM or VL-PM as active layer (*see Note 23*) *and select all residues.*

   Create a merged layer: "Edit" => "Create a Merged Layer from Selection (by layer)".

6. Set the energy minimization parameters, perform energy minimization, and save merged layer (see "Save merged layer" section of Subheading 3.4, step 4).

   Set energy minimization parameters: "Prefs" => "Energy Minimization …". For guiding purposes 50 steps of Steepest Descent are recommended but best values are model dependent.

Perform energy minimization: "Tools" => "Energy Minimization …" and wait for the job to finish. SwissPDBViewer energy minimization computations are done in vacuum and use an implementation of GROMOS96 force field.

3.6. CDR-H3 Refinement

In opposition to the other CDRs (see Note 28), long CDR-H3 loops frequently adopt better conformations after optimization. Several methods can be useful for CDR-H3 optimization including cyclic coordinate descent (ccd) (12) and kinematic closure (kic) (13) loop redesign implemented in ROSETTA (14); vacuum, implicit, or explicit solvent molecular dynamics (MD) implemented in packages such as GROMACS (15), NAMD (16), and CHARMM (17); CAMAL algorithm (18) that uses CONGEN program (19); and loop modeling implemented in MODELLER (20) (server: http://modbase.compbio.ucsf.edu/modloop/). It is a hard task to predict which method will provide better results for each case.
Next, the application of a simulated annealing MD (see Note 29) protocol implemented using a Linux shell script (see Note 30) and GROMACS (see Note 31) will be presented. In the mentioned protocol only CDR-H3 coordinates are updated during simulation while the remaining residues are kept frozen.

1. Download the Linux shell script file to the "*path*/VDs/H3-SA" directory (path: directory path preceding VDs directory) and copy the structure file from the Subheading 3.5, step 6, to the same directory.
2. Edit the script if necessary (see Note 32).
3. In a system terminal, enter the "*path*/VDs/H3-SA" directory and run the script by typing "./EM-SA" (see Note 33). Wait the job to finish and verify the resulting PDB structure in the "SimAnneal" directory. Check for problematic regions and visually inspect the resulting model (see Subheadings 3.7 and 3.8) (see Note 34).

3.7. Verification of Potentially Problematic Conformations

The normality of protein stereochemistry and atomic environments can be assessed by uploading the PDB file to the SAVES server (http://nihserver.mbi.ucla.edu/SAVES/), which allows concomitant analyses by PROCHECK (21) and WHAT IF (22) packages, among others (see Note 35). It should be noted that template-derived problematic regions are a common source of problems regarding homology-based models. Other problems can be frequently corrected by side-chain rotamer search (see Note 36) and/or local energy minimization involving problematic residues.

3.8. Model Visualization

Three-dimensional structure visualization and analysis can be carried out using software such as Pymol (23), VMD (24), and SPDBV (10).

3.9. Quality Perspective of the Resulting Models: RMSD-to-Native Test

VH and VL sequences of three randomly selected antibody structures available at PDB were *de novo* modeled using the protocol described here, WAM, RosettaAntibody, or PIGS Web server. According to RMSD-to-native criteria, the results presented in Table 4 suggest that the models generated by the present approach are expected to be comparable to those generated by current algorithms that are implemented in antibody modeling servers (Fig. 3).

Table 4
Comparison of RMSD-to-native backbone atom distances of models generated by WAM, RosettaAntibody, and PIGS antibody modeling servers and the protocol described here

PDB code	WAM[a]	Rosetta Antibody	PIGS	SPDBV + EM + H3SA	SPDBV + EM	SPDBV
3NTC	1.063	1.434	1.084	1.004	1.065	1.075
1UYW	0.816	0.262[b]	0.869	1.104	1.100	1.105
1LMK	1.089	1.477	1.766	1.450	1.624	1.634

Backbone RMSD values of equivalent regions comprising packed VH/VL are presented for three random test cases. WAM—Web Antibody Modeling Server (http://antibody.bath.ac.uk/); RosettaAntibody server (http://antibody.graylab.jhu.edu/; only the model with best energy was considered); PIGS—Prediction of Immunoglobulin Structure Server (http://arianna.bio.uniroma1.it/pigs/)
SPDBV model that was not submitted to energy minimization, *SPDBV+EM* energy minimized model, *SPDBV+EM+H3SA* Energy minimized model after CDR-H3 optimization by simulated annealing using GROMACS
[a]WAM server does not indicate templates used during modeling process; thus the selection of the crystal structures of the target sequences as template could not be ruled out
[b]PDB 1UYW was automatically selected by RosettaAntibody server as template for the modeling of all FRs and CDRs of HV and LV, compromising in this case the relevance of the RMSD value for the modeling of sequences without solved structures. No option was available to reject 1UYW as template

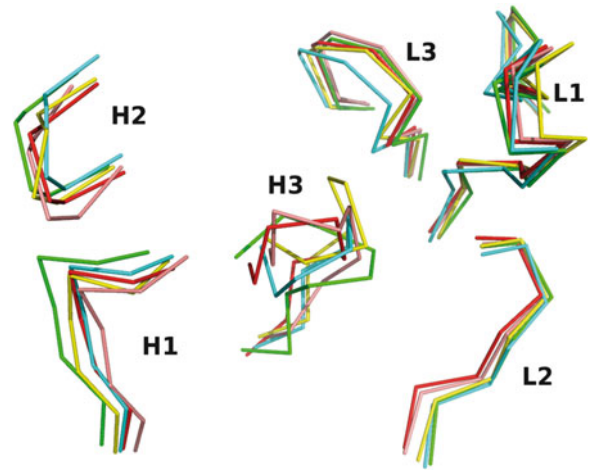

Fig. 3. Superposition of models to the crystallographic structure of KD-247 Fab (PDB 3NTC, solved at 1.55 Å resolution). Structures were superposed using backbone atoms (Ca, N, O, C) of anchor residues and CDRs are shown. 3NTC (*red*), WAM (*blue*), RosettaAntibody (*green*), PIGS (*yellow*) and protocol here described (*salmon*).

4. Notes

1. The reasoning behind comparative modeling is that similar sequences tend to adopt similar structural conformations under similar conditions. It includes homology modeling and protein threading.
2. *Ab initio* predictions do not use structural data of related molecules for model calculation.
3. Thousands of structures are available if lower resolution (above 2 angstrons) and redundancies are considered. For updated information the reader is invited to visit pdb Internet site: www.pdb.org.
4. Even if canonical rules for long CDR-H3 loops are not available, several works propose that their base can be predicted by the nature of residues at key positions (25–27).
5. Usually, more than one week is necessary to download the resulting model. User intervention is necessary (two residues following CDR-L1, CDR-L3, CDR-H1, and CDR-H3 have to be provided when opting for the autoalign method; otherwise users have to manually align its sequences with some templates). In order to obtain the required password to download each model it is necessary to fill a form and send it back to University of Bath (UK). Payment is required for commercial users. Templates are automatically selected.
6. Usually, more than one week is necessary to download the resulting model (it depends on the amount of pending jobs). Low user intervention (selection of parameters). Templates are automatically selected.
7. Models are readily generated. Allow the modeling of multiple sequences. Low user intervention (selection of parameters). Template selection is available.
8. Recent results comparing WAM and RosettaAntibody indicate that WAM's final models tend to fit better experimentally solved structures than the best RosettaAntibody models. On the other hand, RosettaAntibody usually has a model that is closer to the solved structure among its best ten models and docking with the ensemble of structures tends to give better complexes with antigen than WAM's best model (28).
9. It is often the case for WAN and RosettaAntibody servers, but not for PIGS server.
10. Such as analyses of the effect of different templates on the final model, optimization of modeling parameters, and test of different refinement strategies.
11. For instance, Modeller, Discovery Studio, Rosetta, Tinker, Yasara, and SwissPDB-Viewer.

12. Comparisons of models generated by the method delineated here and Modeller or Rosetta package (please note that recent versions of Rosetta package do not include the implementation of the specific algorithm used by RosettaAntibody server) using standard parameters indicate that, in the particular case of antibody modeling, the former tends to yield results that are closer to experimentally solved structures and comparable to WAM, RosettaAntibody, and PIGS servers.
13. The method described here is broadly applicable regarding the required computational resources (hardware and software) and the associated cost. It is also widely applicable in the sense that a wide diversity of antibody VDs can be predicted.
14. For the better understanding of each step, the files corresponding to the modeling process of the three test cases described here can be downloaded at https://docs.google.com/open?id=0B_zucHwbv9VAblpIeVhkZWFvNDA.
15. Ready-to-use directory structure is available at https://docs.google.com/open?id=0B_zucHwbv9VAWW42VmRfb2pEaG8.
16. Optionally, Chothia modified-scheme is also suitable to be applied.
17. Annotation of HV and LV segments is a good modeling practice that facilitates nonautomatic modeling processes and prevents modeling errors due to mistaken assignments of variable domain segments. A template file for comprehensive annotation of HV and LV segments is available at https://docs.google.com/open?id=0B_zucHwbv9VAOU1tb0R4WlI4LXM.
18. To search for homologous template of CDRs use the CDR sequence itself plus two anchor residues at N-terminus and two anchor residues at C-terminus (2res + CDR + 2res). Eventually, H3 loops of appropriate size are difficult to find using standard BLAST parameters. In these cases, PHI-BLAST can be used to exclude hits that do not match loop size (*see*: http://www.ncbi.nlm.nih.gov/blast/html/PHIsyntax.html for syntax of PHI-BLAST patterns). Similarly, keywords that would limit the search to antibody-related molecules can be entered in the "Entrez Query" field (i.e., antibody, immunoglobulin, Fab, scfv, minibody, etc.).
19. For template selection consider the respect of known canonical rules, sequence similarity, and experimental resolution of the structure. Dr. Andrew C.R. Martin's Group developed a system for automatic assignment of canonical classes to CDRs (Web server available at http://www.bioinf.org.uk/abs/chothia.html).
20. Swiss-PDBViewer—Deep View software (SPDBV)—can run on Windows, Mac, and Linux. Download page: http://spdbv.vital-it.ch/disclaim.html.

21. PDB file documentation available at http://www.wwpdb.org/docs.html.
22. Sequentially opening FRs, CDR-1, CDR-2, and CDR-3 template structures can facilitate the work making it more organized.
23. A given structure can be defined as active layer by selecting its corresponding code in the top of "Control panel."
24. Atoms descriptions lines (beginning with ATOM) from different templates should be subsequently listed in the merged file. Cut and paste atoms descriptions from different templates to the correct sequence order (i.e., FR-1 atoms followed by CDR-1, FR-2, CDR-2, FR-3, CDR-3, and FR-4 atoms).
25. Generally, the use of template structures derived from the predicted VH/VL packing angle allows the generation of final models that are slightly closer to the expected structure than models derived from templates selected based on framework regions homology.
26. Aiming to facilitate the selection of matching residues, template and primary model sequences can be manually aligned using Alignment Window (see "Inspect and modify the alignment" section of Subheading 3.4, step 8).
27. Exactly the same number of matching residues should be selected in each structure.
28. Refinement of structures derived from canonical templates generally results in worse conformation.
29. In brief, Simulated Annealing implies simulation of the effect of heat on molecular dynamics which allows bad conformations to surmount energy barriers corresponding to local-minimum conformations followed by a gradual decrease in the system temperature expecting to achieve a new conformation that is more energetically favorable.
30. Even if the shell script was designed for Linux, it can be adapted for other systems or can be used to guide command-line procedures. Sample script available at https://docs.google.com/open?id=0B_zucHwbv9VAemx1VDdxNWdxSTQ.
31. GROMACS packages are available for most popular Linux distributions and Mac OS. Binaries are also available for Windows. Optionally, platform-specific binaries can be built from the source code.
32. At least the target file name should be verified (next to "setenv MOL" statement). It is recommended to number consecutively VD residues on the packaged VDs model (over both chains, see section "Renumber residues" of Subheading 3.4, step 7), so the residue range upstream and downstream the CDR-H3 can be easily defined as frozen while loop residues will be defined as

mobile during simulated annealing. Change residues range if necessary in the index file section of the script.

33. To be run as Linux shell script, the file must have execution permission. To allow the file to be executed use chmod + x "file name" (e.g., "chmod + x EM + SA").
34. GROMACS package provides several programs allowing to analyze the atomic trajectories during simulation. Atomic trajectories can be visualized using VMD (http://www.ks.uiuc.edu/Research/vmd/).
35. Optionally, PROCHECK can be downloaded at http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/download.html and WHAT_CHECK at http://swift.cmbi.ru.nl/gv/whatcheck/whatcheck_8.3_beta.tar.bz2.
36. SPDBV offers rotamer search in conjunction with its mutation tool (http://spdbv.vital-it.ch/mutation_guide.html).

References

1. Whitelegg NR, Rees AR (2000) Wam: an improved algorithm for modelling antibodies on the web. Protein Eng 13:819–824
2. Sircar A, Kim ET, Gray JJ (2009) Rosettaantibody: antibody variable region homology modeling server. Nucleic Acids Res 37:W474–W479
3. Marcatili P, Rosi A, Tramontano A (2008) Pigs: automatic prediction of antibody structures. Bioinformatics 24:1953–1954
4. Kabat EA, Wu TT (1991) Identical v region amino acid sequences and segments of sequences in antibodies of different specificities. relative contributions of vh and vl genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. J Immunol 147:1709–1719
5. Chothia C, Lesk AM (1987) Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 196:901–917
6. Al-Lazikani B, Lesk AM, Chothia C (1997) Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 273:927–948
7. Abhinandan KR, Martin ACR (2008) Analysis and improvements to kabat and structurally correct numbering of antibody variable domains. Mol Immunol 45:3832–3839
8. Honegger A, Plückthun A (2001) Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. J Mol Biol 309:657–670
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
10. Guex N, Peitsch MC (1997) Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
11. Abhinandan KR, Martin ACR (2010) Analysis and prediction of vh/vl packing in antibodies. Protein Eng Des Sel 23:689–697
12. Canutescu AA, Dunbrack RLJ (2003) Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci 12: 963–972
13. Mandell DJ, Coutsias EA, Kortemme T (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods 6:551–552
14. Kaufmann KW, Lemmon GH, Deluca SL, Sheehan JH, Meiler J (2010) Practically useful: what the rosetta protein modeling suite can do for you. Biochemistry 49:2987–2998
15. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) Gromacs: fast, flexible, and free. J Comput Chem 26:1701–1718
16. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with namd. J Comput Chem 26:1781–1802
17. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B,

Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) Charmm: the biomolecular simulation program. J Comput Chem 30:1545–1614

18. Martin AC, Cheetham JC, Rees AR (1989) Modeling antibody hypervariable loops: a combined algorithm. Proc Natl Acad Sci U S A 86:9268–9272
19. Bruccoleri RE, Karplus M (1987) Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26:137–168
20. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9:1753–1773
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
22. Vriend G (1990) What if: a molecular modeling and drug design program. J Mol Graph 8(52–6):29
23. Schrödinger L The pymol molecular graphics system
24. Humphrey W, Dalke A, Schulten K (1996) Vmd: visual molecular dynamics. J Mol Graph 14(33–8):27–28
25. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM (1998) Conformations of the third hypervariable region in the vh domain of immunoglobulins. J Mol Biol 275: 269–294
26. Shirai H, Kidera A, Nakamura H (1996) Structural classification of cdr-h3 in antibodies. FEBS Lett 399:1–8
27. Kuroda D, Shirai H, Kobori M, Nakamura H (2008) Structural classification of cdr-h3 revisited: a lesson in antibody modeling. Proteins 73:608–620
28. Sivasubramanian A, Sircar A, Chaudhury S, Gray JJ (2009) Toward high-resolution homology modeling of antibody fv regions and application to antibody-antigen docking. Proteins 74:497–514
29. Martin AC, Thornton JM (1996) Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. J Mol Biol 263:800–815
30. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR et al (1989) Conformations of immunoglobulin hypervariable regions. Nature 342:877–883
31. Barré S, Greenberg AS, Flajnik MF, Chothia C (1994) Structural conservation of hypervariable regions in immunoglobulins evolution. Nat Struct Biol 1:915–920
32. Tomlinson IM, Cox JP, Gherardi E, Lesk AM, Chothia C (1995) The structural repertoire of the human v kappa domain. EMBO J 14:4628–4638
33. Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G (1992) Structural repertoire of the human vh segments. J Mol Biol 227:799–817

http://www.springer.com/978-1-61779-973-0

Antibody Engineering
Methods and Protocols, Second Edition
Chames, P. (Ed.)
2012, XVI, 734 p. 98 illus., 41 illus. in color., Hardcover
ISBN: 978-1-61779-973-0
A product of Humana Press