Automated antibody structure prediction using Accelrys tools: Results and best practices

Marc Fasnacht,* Ken Butenhof, Anne Goupil-Lamy, Francisco Hernandez-Guzman, Hongwei Huang, and Lisa Yan

Accelrys, Inc., San Diego, California 92121

ABSTRACT

We describe the methodology and results from our participation in the second Antibody Modeling Assessment experiment. During the experiment we predicted the structure of eleven unpublished antibody Fv fragments. Our prediction methods centered on template-based modeling; potential templates were selected from an antibody database based on their sequence similarity to the target in the framework regions. Depending on the quality of the templates, we constructed models of the antibody framework regions either using a single, chimeric or multiple template approach. The hypervariable loop regions in the initial models were rebuilt by grafting the corresponding regions from suitable templates onto the model. For the H3 loop region, we further refined models using ab initio methods. The final models were subjected to constrained energy minimization to resolve severe local structural problems. The analysis of the models submitted show that Accelrys tools allow for the construction of quite accurate models for the framework and the canonical CDR regions, with RMSDs to the X-ray structure on average below 1 Å for most of these regions. The results show that accurate prediction of the H3 hypervariable loops remains a challenge. Furthermore, model quality assessment of the submitted models show that the models are of quite high quality, with local geometry assessment scores similar to that of the target X-ray structures.

INTRODUCTION

Knowing the detailed three-dimensional structure of a protein can offer valuable insights into its function and interactions with other molecules. This is of particular importance in the design and optimization of drug candidates. Over the last decade, homology modeling[1] has become an important method for structure prediction of proteins for which no experimental structures are available. The CASP experiments,[2] which have been conducted every 2 years since 1994, have been documenting the significant progress in the field over the last two decades. In general it can be quite difficult to accurately predict a protein structure from its sequence. However, if an X-ray structure for a protein with a high degree of sequence similarity is available, quite accurate models can be built using currently available tools, such as MODELER,[3] RosettaAntibody,[4] or MOE.[5]

Antibody-based therapeutics have become important tools in the treatment of cancer and other diseases.[6,7] Building computational models is frequently an important step in the antibody design process that allows researchers to study antibody properties such as stability, antigenicity, aggregation propensity, solubility, viscosity, and more. In addition, homology models can be used to gain insight into and predict antibody-antigen interactions when used in combination with protein-protein docking methods, such as ZDOCK[8] or SnugDock.[9] The area of antibody design and engineering represents a special case to which homology modeling is particularly well suited, because in general the overall sequence and structural similarity between antibodies is very high. In

particular, the framework regions of antibodies are very well conserved, with most of the variability occurring in the complementarity-determining regions (CDRs).

A blind prediction experiment, similar to CASP, but limited to antibody structure prediction was performed in 2009.[10] The results of Accelrys' participation in this experiment generally validated our template-based modeling approach, including the effectiveness of using chimeric templates (separate templates for the light and heavy chains, oriented by a template containing both a light and a heavy chain). However, it also highlighted some deficiencies in our modeling process. Since the first experiment, we have improved our tools, incorporating a number of lessons learned from the 2009 experiment, as discussed below. The second installment of the antibody prediction experiment was executed in early 2013.[11] Here we discuss what we did well and what can be improved based on the results from our participation in the second Antibody Modeling Assessment (AMA-II).[12]

MATERIALS AND METHODS

The AMA-II prediction experiment consisted of two stages. In the first stage, only the sequences of the 11 Fv targets were available to predictors and the goal was to build models of the Fv region based on this sequence information. For the second stage, the X-ray structures for all target Fv domains, with the H3 CDR residues removed, were made available. The task for the second stage was to predict the conformation of only the H3 loop given the correct crystallographic environment. For details on the targets and a general description of the experiment, consult the description and assessment of the experiment by the organizers.[11] The following is a description of our methods used for model construction for each stage of the model construction process.

Stage 1

Framework template selection

Templates for each of the 11 targets were selected by aligning the target sequences against sequences in a pre-curated database of antibodies extracted from the Protein Data Bank (PDB).[13,14] Alignments were performed using a Hidden Markov Model.[15,16] Based on this alignment, potential templates were identified by calculating the sequence similarity and identity against the target Fv framework region, excluding the CDR loops. The rationale for excluding the CDR loops was that these regions showed a high degree of sequence and structural variability, and therefore might have masked high degrees of similarity for the framework regions. We also identified a separate list of potential templates for the VL and VH domains, again, based on sequence similarity and identity of the framework regions for these domains.

By default, the structures with the highest sequence similarity to the target for the framework regions were selected as the framework templates. However, in cases where several top templates with similar framework sequence similarity were available, team members considered additional criteria such as the similarity of the CDR loops, X-ray resolution, matching organism and germline, structural consensus between templates (i.e., we tried to avoid templates that were obvious outliers in the sense that they had significant structural differences from the majority of the top templates) and binding to antigen or not, etc. to make the final template selection.

For this analysis, the framework region and CDR loops used in the sequence similarity calculations were defined according to the IMGT Unique Numbering Scheme.[17]

Framework model construction

We used three different methods to construct models based on the framework templates identified by the procedure described above:

1. The first approach was to build a model based on a single Fv framework template. Below, we refer to this approach as "single template."
2. The second approach was to build a model based on a chimeric template. The template was assembled from the individual VL and VH templates based on a third interface template that contained both VL and VH domains to determine the relative spatial orientation of the individual VL and VH templates. Below, we refer to this approach as "chimeric template." Note that the VL or VH templates can be identical to the corresponding domain of the interface template.
3. The third approach was to build a model based on five overall Fv framework templates. The models were built based on a multiple sequence alignment of all five templates to the target sequence using the capability of MODELER to construct models based on multiple templates. MODELER uses an additive distance restraint function that peaks at the equivalent distance between atoms in each template. The contribution for each template is weighted by local sequence similarity. For a more detailed description of the MODELER algorithm refer to Ref. 3. Below we refer to this approach as "top5 template" method.

In all cases, 50 models were built using MODELER.[3] The top model as ranked by the MODELER PDF Physical Energy was selected for further refinement.

CDR refinement

The framework models from the previous step were inspected to determine whether further refinement of the hypervariable loop regions (L1, L2, L3, H1, H2, and H3) was required. In cases where the template CDR region of the framework template already had identical sequence to

the target CDR loop, the framework CDR model was kept. However, in most cases the hypervariable loops were rebuilt.

The CDR loop residues selected for refinement were either based on the Chothia[18–20] (targets Ab01, Ab02, Ab03, Ab07, Ab09, Ab10, Ab11) or IMGT[17] CDR definitions (Ab04, Ab05, Ab06, Ab08). Note that for this study, we refer to the range of residues specified by the respective CDR definition (Chothia or IMGT) as "CDR loop," the adjacent residues at either terminal of the CDR loop as "stem residues," and loop and stems together as "CDR region."

CDR loop refinement was performed using the following approach: we identified a set of loop templates for each CDR region based on an alignment of the target sequence to antibody sequences in our antibody database. Template CDRs must be of the same length as the target loops, and templates were ranked by a BLOSUM62 based similarity score of the CDR region, including stems.

H3 refinement

For the modeling of the H3 loops, we used three different approaches:

- The first was purely template based, using the same procedure as described above for L1, L2, L3, H1, and H2. However, in addition to the similarity score, we also took into account the H3 classification described in the work by Kuroda *et al.*[21–23] and selected templates in agreement with these rules whenever available.
- The second method employed was to build the most variable region of the loop with an ab initio approach using Looper.[24] However, we did not usually rebuild the entire H3 region. We determined the range to model by examining the conformation of the templates used in the first approach. Typically, the templates agreed quite well in the stem regions so that only the center of the loop needed to be rebuilt. We identified the residue range for ab initio modeling by determining the residue at which the backbone conformation of the templates started to diverge.
- The third method added a round of neighboring side-chain refinement to the preceding ab initio methodology. The goal of this step was to fix incorrectly positioned neighboring side-chains, which can prevent the ab initio modeling methods from successfully finding a correct solution. For this method, the side-chains of the H3 loop were mutated to Alanine. Then side-chains neighboring the H3 loop were selected and refined using CHIROTOR.[25] The loop side-chains were then mutated back to their original identity, and then a final round of ab initio refinement was performed.

Final minimization

After H3 refinement, all models were inspected for structural problems, such as clashes between atoms. For severe clashes, a limited short energy minimization of the affected regions (typically side-chains) was performed using CHARMm.[26,27] During the minimization, restraints were applied to the backbone and non-affected side-chains as to not perturb the rest of the structure.

Stage 2

For the second stage of the Antibody Modeling Assessment experiment (AMA-II), the X-ray structure of the target (with the exception of the H3 CDR residues), was known. The task was to model the missing section, given the X-ray environment. We used the following approach:

- In the first step, we built the missing region based on homology modeling or simple grafting of the presumed best H3 loop from AMA-II part 1. When using the homology modeling approach, we typically used the same templates as for H3 modeling in the first stage. However, a few of the templates were excluded because of incompatibility of the stem regions in the template structures compared to the known stems in the target X-ray structures.
- The next step was to determine which region to model ab initio. Since accuracy in loop prediction typically decreases with loop length,[24] we tried to restrict the ab initio prediction range as much as possible by using information from the available templates. The stems in H3 loops are often conserved, so we restricted the ab initio loop prediction range to the regions where the available templates started to diverge. This range was determined by visually inspecting a superposition of the target structure with H3 templates.
- We then built 50 models using Looper.[24] Looper provides an energy score and a clustering of the output loops. To pick the models to submit, we identified the clusters with the most low energy structures and submitted the lowest energy model from top ranked clusters. Predictions for a single target typically take less than 30 min on a standard desktop machine.

Manual versus automated predictions

For models submitted for AMA-II first stage, the targets were split up between the six team members, with one member taking the lead for each target. Each team member was working on either one, two, or three structures. Team members used a combination of the methods described above, with each team member manually choosing the methods they thought were most appropriate for their target, based on the available templates.

For framework modeling our strategy was to use a single template approach if a very high sequence similarity template was available for the target, the chimeric approach if templates for the VL and/or VH domains were available that showed significantly higher sequence

similarity than the best corresponding domains of the best overall Fv template, and to use the top5 approach if several Fv templates with high sequence similarity were available. However, apart from target Ab01, where only one good template was available, we did not observe evidence indicating a clear choice as to which method was most suitable for each target. The selected method is mainly a preference of the individual team member who worked on the target. Table I summarizes the methods used for each model submitted.

The data from the manual models submitted for AMA-II stage 1 gives a good indication of the performance of our software in "real-life" situations, where the software is used by domain experts with some degree of manual intervention in choosing methods and the final selection of templates. However the manual interventions make it more difficult to evaluate the performance of the underlying methods. In order to understand which method works best in general, we ran post-experiment automated analysis, where we systematically applied each method to all the targets and simplified template selection to only use those with the highest sequence similarity to make the comparison more consistent. For the models that deviated significantly from the automated methods, more in-depth analysis is provided in the discussion section.

Evaluation of structures

To evaluate our models, we compared them to the corresponding X-ray structures provided by the organizers. Note that most of these structures are not yet released by the Protein Data Bank[14] at the time of writing of this manuscript, so some minor revisions of the final structures are possible.

Models and X-ray structures were compared by calculating root-mean-square deviations (RMSD) between the two structures. For the calculations, we followed the approach used in the assessment of the experiment[11] as closely as possible. For RMSD calculations, structures were superimposed using the β-sheet core, which was defined as follows: VL: 3-13,[†] 18-25, 33-38, 43-49, 61-67, 70-76, 85-90, 97-103; VH: 3-7, 18-24, 34-40, 44-51, 56-59, 67-72, 77-82a, 87-94, 102-110. The CDR loops for the purpose of the RMSD calculations were defined as the following ranges: CDR-L1:27-32, CDR-L2:50-53, CDR-L3: 91-96, CDR-H1: 26-33, CDR-H2: 52-55, CDR-H4: 73-76 and CDR-H3: 95-100x, where 100x refers to whatever residue is just before position 101 in the Chothia numbering scheme.[11] All ranges are based on the Chothia numbering scheme.[18–20] RMSDs were evaluated over different ranges: the CDRs, the β-sheet core,

as defined above, and the framework residues. The framework residues are defined as all residues except CDRs L1, L2, L3, H1, H2 and H3, and the termini.[‡]

Unless otherwise stated, all RMSDs reported in this study were calculated using the peptide carbonyl atoms C and O. The carbonyl RMSD is used by the general evaluation of all models by the organizers[11] and we followed their definition to make the discussion clear. The reason for this choice by the organizers is that carbonyl RMSDs are more sensitive to local structural deviations, such as peptide flip than the more commonly reported C-α or backbone RMSDs. Note however that carbonyl RMSDs cannot directly be compared to C-α RMSDs. On average, carbonyl RMSDs are on the order of 10% higher than the corresponding C-α RMSDs. In cases involving peptide flips, the difference can be much bigger as pointed out in Ref. 11. To allow for easier comparison with results reported elsewhere in the literature, we also include C-α RMSDs for our models in the Supporting Information.

In addition to the RMSDs, we also calculated the deviation in angle of the VL and VH domains between model and X-ray structure. For consistency we follow the approach outlined in Ref. 11. The calculation was done by sequentially superimposing the β-sheet core regions of the model and X-ray VL and VH domains and determining the χ-angle, in spherical polar coordinates (ωφχ), of the second superimposition transformation. Note that the tilt angle calculated this way only partially captures the difference in orientation of VL/VH domains between model and target structure; it leaves out the direction of the tilt as well as any relative translations.

RESULTS AND DISCUSSION

Framework models

Template selection

The templates selected for building models for the first stage are listed in Table I. In order to evaluate our template selection we calculated the framework RMSDs of all Fv structures in our database with respect to X-ray structures of the 11 targets. Figure 1(a) compares the Fv framework RMSD of templates used in building the submitted models (plotted as solid colored circles) to the overall RMSD-distributions of all structures in the database (indicated by the box plots). RMSDs of templates selected for single template or chimeric template models[§] for the manual predictions are plotted as red circles. RMSDs of templates used for manual predictions using the top5 approach are plotted in orange. The plot also

---

[†]The original residue range described in the experiment assessment include residues VL 3-8, but the residues were removed from the superposition range by the organizers because several groups submitted models that had incorrectly placed strands that affected the alignment.[11]

[‡]The terminal residues excluded are the same as for the β-sheet core (i.e. any residues before residue 3 in VL and VH and any residue after residue 103 in VL or after residue 110 in VH.)

[§]For chimeric models, the RMSD of the interface template is plotted.

Table I
Summary of Methods and Templates Used for the Models for Each Target

Model	VH Templ	VL Templ	FV Templ	Framework	H3 Refinement
accAb01m1	4hbc	4hbc	4hbc	Single	Template
accAb01m2	4hbc	4hbc	4hbc	Single	Looper
accAb01m3	3nl4	3nl4	3nl4	Single	Looper
accAb02m1	3umt	3mbx,1sbs,1il1	2gki	Chimeric	Environment
accAb02m2	3umt	3mbx	2gki	Chimeric	Environment
accAb02m3	3umt	3mbx	2gki	Chimeric	Environment
accAb03m1	2xtj	1hez	2xtj	Chimeric	Looper
accAb03m2	3ma9	1dee	2xtj	Chimeric	Looper
accAb03m3	2xtj	2xtj	2xtj	Single	Environment
accAb04m1	3mxv	3mxv	3mxv	Single	Looper
accAb04m2	3mxv	3mxv	3mxv	Single	Environment
accAb04m3	3mxv	3iu4	3mxv	Chimeric	Looper
accAb05m1	2xwt,3mlw,3n9g,4d9l,4fqj	2xwt,3mlw,3n9g,4d9l,4fqj	2xwt,3mlw,3n9g,4d9l,4fqj	Top 5	Template
accAb05m2	3na9	4d9l	3n9g	Chimeric	Template
accAb05m3	2xwt,3mlw,3n9g,4d9l,4fqj	2xwt,3mlw,3n9g,4d9l,4fqj	2xwt,3mlw,3n9g,4d9l,4fqj	Top 5	Looper
accAb06m1	1dee,1hez,2uzi,3bn9,3s34	1dee,1hez,2uzi,3bn9,3s34	1dee,1hez,2uzi,3bn9,3s34	Top 5	Environment
accAb06m2	1dee,1hez,2uzi,3bn9,3s34	1dee,1hez,2uzi,3bn9,3s34	1dee,1hez,2uzi,3bn9,3s34	Top 5	Looper
accAb06m3	3h42	1vge	3bn9	Chimeric	Template
accAb07m1	1nsn,1c12,1wej,3rvu,3rvv	1nsn,2aab,2xqy,3ddg,3pqh	1nsn	Chimeric	Template
accAb07m2	1nsn,1c12,1wej,3rvu,3rvv	1nsn,2aab,2xqy,3ddg,3pqh	1nsn	Chimeric	Looper
accAb07m3	1nsn,1c12,1wej,3rvu,3rvv	1nsn,2aab,2xqy,3ddg,3pqh	1nsn	Chimeric	Looper
accAb08m1	1pkq, 2gki, 2i9l, 3q3g, 3ujt	1pkq, 2gki, 2i9l, 3q3g, 3ujt	1pkq, 2gki, 2i9l, 3q3g, 3ujt	Top 5	Template
accAb08m2	1pkq, 2gki, 2i9l, 3q3g, 3ujt	1pkq, 2gki, 2i9l, 3q3g, 3ujt	1pkq, 2gki, 2i9l, 3q3g, 3ujt	Top 5	Environment
accAb08m3	1pkq, 2gki, 2i9l, 3q3g, 3ujt	1pkq, 2gki, 2i9l, 3q3g, 3ujt	1pkq, 2gki, 2i9l, 3q3g, 3ujt	Top 5	Looper
accAb09m1	3na9	3na9	3na9	Single	Template
accAb09m2	3na9	3na9	3na9	Single	Looper
accAb09m3	3na9	3na9	3na9	Single	Looper
accAb10m1	2gki	2gki	2gki	Single	Environment
accAb10m2	2gki	2gki	2gki	Single	Looper
accAb10m3	2gki	2gki	2gki	Single	Template
accAb11m1	2w9d	2ih3	2w9d	Chimeric	Environment
accAb11m2	2w9d	2ih3	2w9d	Chimeric	Looper
accAb11m3	2w9d	2ih3	2w9d	Chimeric	Template

White background indicates a single template model, light gray a chimeric model, and dark gray a top5 template model.

shows the RMSDs of the five templates in the database with the highest sequence similarity to the target (shown as blue circles). Note that the blue circles correspond to the templates selected for the automated predictions because template selection for the automated prediction was based purely on sequence similarity, whereas the manual predictions (red and orange circles) sometimes used additional criteria, which lead to templates with slightly lower sequence similarity, particularly for targets Ab05, Ab06, and Ab11. These cases will be discussed in more detail below.

For most targets, one of the top five templates ranked by similarity is close to minimum of the RMSD distribution, but the template with the lowest RMSD does not always have the highest sequence similarity, which was not unexpected: In general, pairs of protein structures with high sequence similarity (>50%) align to within ~1 Å or less.[28] However, structural differences remain even at 100% sequence identity.[29] For antibodies, structural variations are typically most prominent in the CDR regions, which were excluded from the RMSD calculation. However, framework-RMSDs of up to ~0.5 Å are observed in cases where multiple structures of an anti-body were solved in the asymmetric unit in a crystal.[10] In addition, antigen binding may affect the tilt angle of the VH and VL domains leading to even larger RMSD differences relative to unbound antibody as indicated in a systematic comparison by Sela-Culang et al.[30] Since our database included both bound and free structures, we expect some variation in template RMSD even for structures with strong similarity.

Note that we made use of both ligand-bound and unbound structures as framework templates, and used antigen binding only as a minor consideration among other factors. In many cases, the highest sequence similarity template was a ligand-bound structure, and the tradeoff between selecting an unbound template with lower sequence similarity over an antigen-bound template with higher sequence similarity is not clearly understood. In fact, for Ab05 model 2, we chose a template without bound ligand as tradeoff for high sequence similarity, which turned out to be a bad choice as seen in Figure 1. Additionally, a post-experiment comparison of antigen-bound templates to unbound templates did not reveal any advantage of systematically selecting unbound structures over bound structures with similar sequence

similarity for the AMA-II targets. However, the data from the 11 AMA-II targets is not sufficient to draw any definitive conclusion about how the presence of antigens in templates influences the model quality. This will involve analysis of the entire antibody structure database and is beyond the scope of current work.

The variation in RMSD among top templates is illustrated in Figure 1(b), which displays the framework RMSD as a function of sequence similarity for the five sequences with the highest sequence similarity to the corresponding target. The top five templates for all targets have more than 90% sequence similarity, with the exception of target from rabbit, Ab01, for which only one template has more than 90% sequence similarity. However, even though the difference in sequence similarities between templates is ∼3% or less for all targets (except Ab01), the spread of template RMSD values can be relatively large. The most extreme case is target Ab03, for which the top template (2XTJ) has 98.9% sequence similarity for an RMSD of 0.6 Å, and the second template (1RZI) has 97.8% sequence similarity with an RMSD of 1.4 Å. For other targets (Ab04, Ab07, Ab08) templates

with slightly lower sequence similarity have lower RMSDs than templates with higher sequence similarity. It therefore appears that picking templates purely based on a single sequence similarity score does not always result in an optimal template.

Nevertheless, it should be noted that most of the five top sequence-similar templates have RMSD values below 1 Å with respect to the target structures, so models built using any of these templates should have RMSDs below 1 Å in most cases.

For most targets, predictors selected the default templates (i.e., the structures with the highest sequence similarity) for building the models. In most cases, this resulted in a template with an RMSD below 1 Å. For three targets (Ab05, Ab6, Ab11), slightly different strategies were used to select the templates. Target Ab06 was a case where a large number of templates with very high sequence similarity to the target sequence (≥95%) were available. The set of templates was pruned by looking for structural consensus among templates and the quality of the selected templates was similar to that of the default selection. For targets Ab05 and Ab11, the RMSD of the manually selected templates was significantly larger than that of the default template with the highest sequence similarity. These two cases are discussed in detail below.
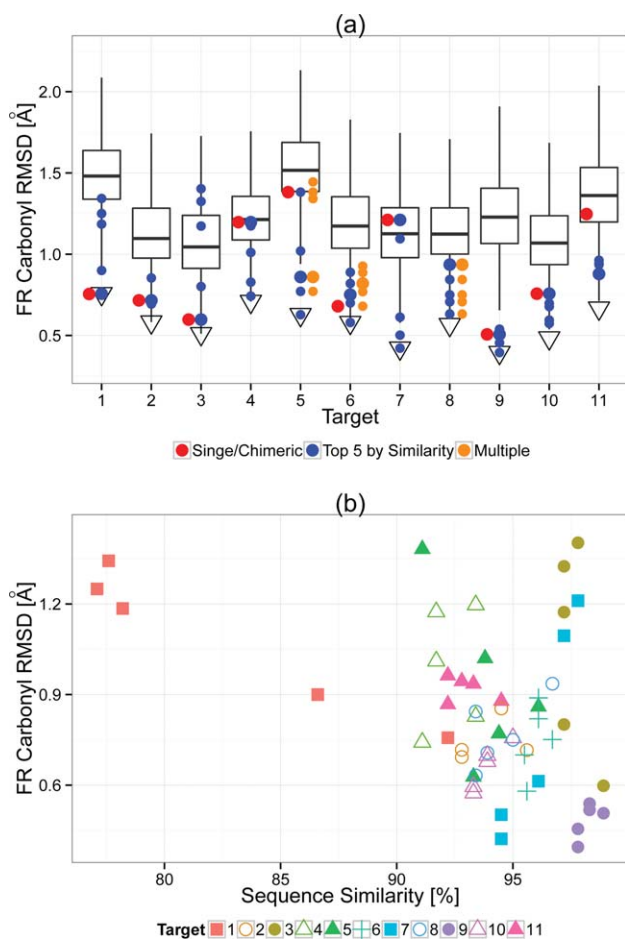

Figure 1.

Figure 1

(a) Framework RMSDs (peptide carbonyl) of potential templates for the eleven targets. The box plots show the distribution of the Framework RMSD with respect to the target for all Fv structures in the antibody database (2099 structures). The horizontal bar inside the box indicates the median of the distribution, the top and bottom of the box are the 75th and 25th percentiles. The tails indicate the highest/lowest RMSD value that fall within a factor of 1.5 times the interquartile distance of the box boundaries. Outliers are not shown, except for the minimum of the distribution which is plotted as a triangle. The blue circles indicate the RMSDs of the top five most sequence-similar templates identified in the database, with the larger circle indicating the structure with the highest sequence similarity (similarity calculated over framework residues only). The red circles show the RMSD of the interface template used in the submitted models, for models built with the single or chimeric template approach. The orange circles indicate the templates used for manual predictions using the top5 template approach, which the largest circle corresponding to the template with the highest sequence similarity to the target. (Note that for the manual top5 predictions (orange circles) in some cases templates were selected using additional criteria. As a result the set of structures used for manual predictions can differ from the five most-sequence similar templates in the database (blue circles).) In general, at least one of the five most sequence similar templates had an RMSD close to the lowest RMSD structure in the database. For models using the single or chimeric template approach, we typically chose the structure with the highest sequence similarity as interface template, except for targets Ab05 and Ab11. (b) Framework RMSD (peptide carbonyl) of the top five templates for each of the targets vs. sequence similarity. The sequence similarity was also calculated considering only framework residues. The plot shows a generally negative correlation between the sequence similarity and the RMSD. However, the correlation is not very strong, especially for similarities greater than 90%, which includes the templates for all targets, except Ab01 (4MA3).
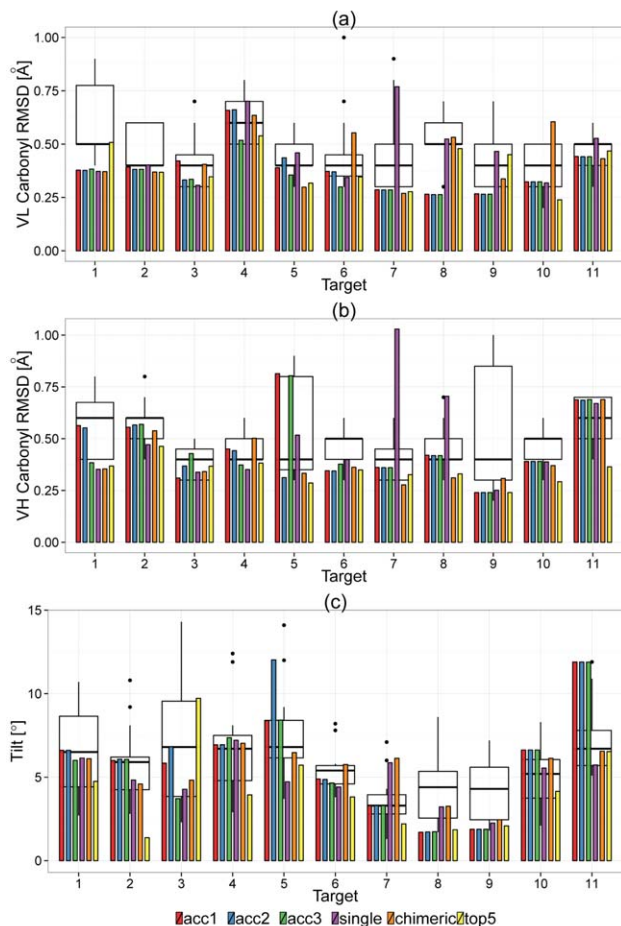
Figure 2

Plots in this figure compare the quality of the framework region of models for each target. In each panel, the results for the Accelrys models submitted for AMA-II are shown as red (model 1), blue (model 2), and green (model 3) bars. The results for models built using different automated approaches in the post-experiment analysis are shown as purple (single template), orange (chimeric template) and yellow (top five templates) bars. In each panel, the box plots in the background indicate the distribution for the models submitted by AMA-II participants (as reported in Ref. 11[‖]). The thick black bar inside the boxes indicates the median, the top and bottom boundaries of the boxes indicate the first and third quartiles (i.e. 25th and 75th percentiles). The tails indicate the highest/lowest RMSD value that fall within a factor of 1.5 times the interquartile distance of the box boundaries. Any outliers fall in the regions beyond the tails are drawn as black circles. (**a**) Plots RMSD of the model structures compared to the X-ray structure calculated over β-core of the VL region. (**b**) Plots the same data for the VH region. (**c**) The results of the Tilt angle.

Submitted models

The results for the framework regions of the models are summarized in Table II and Figure 2. In the table, the columns VL and VH give the RMSD over the β-core regions, as defined above. The column FR lists the

RMSD values of the models when comparing the entire structure, excluding CDRs and termini.

VL domain. From Table II, we can see that the models are very accurate for the VL region, with all but the models for target Ab04 having RMSD values below 0.5 Å. The distribution in Figure 2(a) also shows that the Accelrys models (acc) are generally below the median of the distributions of all submitted AMA-II models, with models for target Ab08 standing out for low RMSDs.

The comparatively high RMSDs for the VL domain for the target Ab04 models seem to be caused by a sub-optimal choice of template; the VL domain of 3MXV, which was used for the single template models accAb04m1 and accAb04m2, is not among the top VL templates for this target. As pointed out in Ref. 11 3MXV has a cis-proline at position 8 in VL, which is replaced by a histidine in Ab04. However, the cis-conformation at position 8 was incorporated into the model accAb04m1 and accAb04m2, which lead to an incorrect conformation of the backbone in that region. Using a separate template for VL seems to help in this case. Chimeric model accAb04m3, which used 3IU4 as VL template, had a VL RMSD of 0.5 Å compared to 0.7 Å for the models based entirely on 3MXV (accAb04m1, accAb04m2).

VH domain. For VH, results are similar with slightly higher RMSDs than those for VL. They are displayed in Figure 2(b) and Table II. Most submitted models have RMSDs of around 0.5 Å or slightly better, with the exceptions being targets Ab05 and Ab11.

The relatively high RMSD value of the VH domain of the models for target Ab11 highlights the sensitivity of model accuracy to the selection of the template. We selected 2W9D, the template with the highest sequence similarity to target Ab11 (96.7%), but another structure, 1F11, with slightly lower sequence similarity (94.5%), would have been a better choice. The H1 and H2 loops and stem residues in 1F11 adopt very similar conformations to those in the X-ray structure of Ab11, with RMSDs around 0.5 Å, whereas H1 and H2 CRD regions in 2W9D adopt very different conformations from Ab11 (with RMSDs greater than 2 Å). Retrospective analysis shows that 1F11 has a slightly higher sequence identity to the target (87.9% vs. 86.8% for 2W9D) and the residues in CDRs H1 and H2 loops and the corresponding stem regions are more conserved in 1F11. However, it is important to note that selecting the better template (1F11) was very difficult in this case as the sequence-differences are quite subtle. As shown in Figure 2, the top5 templates approach improves the accuracy of the model in this case since the method is less dependent on the choice of a single best template.

For target Ab05, models accAb05m1 and accAb05m3, which have relatively high VH RMSD values, were built using the top5 approach. When selecting the five

Table II
Peptide Carbonyl RMSD Values of the Models Submitted

Name	Peptide Carbonyl RMSD [Å]											Tilt [°]
	VL	VH	L1	L2	L3	H1	H2	H3	C-β	FR	All	
accAb01m1	0.4	0.6	3.6	0.5	5.5	0.7	1.4	6.0	0.7	0.9	1.9	6.6
accAb01m2	0.4	0.6	3.6	0.6	5.5	0.7	1.5	6.2	0.7	0.9	1.9	6.6
accAb01m3	0.4	0.4	4.4	0.3	5.7	1.2	1.9	6.3	0.6	0.7	1.9	6.0
accAb02m1	0.4	0.6	0.6	0.3	1.4	1.1	1.1	4.8	0.7	0.8	1.3	6.0
accAb02m2	0.4	0.6	0.6	0.3	1.1	1.1	1.1	4.8	0.7	0.8	1.3	6.1
accAb02m3	0.4	0.6	0.8	0.3	0.6	1.1	1.1	6.0	0.7	0.9	1.4	6.1
accAb03m1	0.4	0.3	0.3	0.4	0.6	2.9	0.4	2.3	0.6	0.7	1.0	5.8
accAb03m2	0.3	0.4	0.3	0.3	0.7	2.7	1.1	3.5	0.7	0.8	1.1	6.8
accAb03m3	0.3	0.4	0.4	0.3	0.5	6.1	3.8	10.3	0.5	0.7	2.3	3.7
accAb04m1	0.7	0.5	1.1	0.4	0.4	1.0	1.0	1.9	0.9	1.3	1.4	6.9
accAb04m2	0.7	0.4	1.2	0.4	0.4	1.0	1.0	2.1	0.9	1.3	1.3	6.9
accAb04m3	0.5	0.4	1.3	0.3	0.7	0.7	0.9	1.8	0.8	1.1	1.2	7.4
accAb05m1	0.4	0.8	1.0	0.4	3.7	0.4	0.4	3.7	1.0	1.1	1.6	8.4
accAb05m2	0.4	0.3	0.8	0.5	2.5	0.2	0.3	3.7	1.1	1.2	1.5	12.0
accAb05m3	0.4	0.8	0.9	0.3	3.9	0.4	0.5	3.7	0.9	1.1	1.6	8.4
accAb06m1	0.4	0.3	0.4	0.3	0.7	0.7	0.4	4.8	0.6	0.7	1.3	4.9
accAb06m2	0.4	0.3	0.4	0.3	0.7	0.6	0.3	3.9	0.6	0.7	1.1	4.9
accAb06m3	0.3	0.4	0.4	0.3	0.8	0.4	0.3	3.0	0.5	0.6	0.9	4.7
accAb07m1	0.3	0.4	0.8	0.2	0.7	0.3	0.3	3.1	0.6	0.6	0.9	3.3
accAb07m2	0.3	0.4	0.8	0.2	0.7	0.3	0.3	3.2	0.6	0.6	0.9	3.3
accAb07m3	0.3	0.4	0.8	0.2	0.7	0.3	0.3	3.0	0.6	0.6	0.9	3.3
accAb08m1	0.3	0.4	0.7	0.2	0.9	0.5	1.1	2.3	0.4	0.6	0.8	1.7
accAb08m2	0.3	0.4	0.7	0.2	0.9	0.5	1.1	4.7	0.4	0.6	1.1	1.7
accAb08m3	0.3	0.4	0.7	0.2	0.9	0.5	1.1	4.3	0.4	0.6	1.1	1.7
accAb09m1	0.3	0.2	0.3	0.3	0.5	0.4	0.7	3.0	0.3	0.4	0.7	1.9
accAb09m2	0.3	0.2	0.3	0.3	0.5	0.4	0.7	4.2	0.3	0.4	0.9	1.9
accAb09m3	0.3	0.2	0.3	0.3	0.4	0.4	0.7	4.6	0.3	0.4	0.9	1.9
accAb10m1	0.3	0.4	0.6	0.3	1.6	0.4	0.4	1.9	0.7	0.9	1.0	6.6
accAb10m2	0.3	0.4	0.6	0.3	1.6	0.4	0.4	2.2	0.7	0.9	1.0	6.6
accAb10m3	0.3	0.4	0.6	0.3	1.6	0.4	0.4	3.1	0.7	0.9	1.2	6.6
accAb11m1	0.4	0.7	0.5	0.3	0.9	2.6	2.3	3.6	1.1	1.3	1.7	11.9
accAb11m2	0.4	0.7	0.5	0.3	0.9	2.6	2.3	3.4	1.1	1.3	1.6	11.9
accAb11m3	0.4	0.7	0.5	0.5	0.9	2.6	2.3	2.7	1.1	1.3	1.6	11.9
Avg(All)	0.4	0.4	0.9	0.3	1.5	1.1	1.0	3.9	0.7	0.9	1.3	5.9
SD(All)	0.1	0.2	1.0	0.1	1.6	1.2	0.8	1.7	0.2	0.3	0.4	3.0
Avg(M1)	0.4	0.5	0.6	0.3	1.1	1.0	0.8	3.1	0.7	0.9	1.2	5.7
SD(M1)	0.1	0.2	0.3	0.1	1.0	0.9	0.6	1.1	0.2	0.3	0.3	3.1

CDR regions for which the X-ray structure adopts a non-canonical structure are indicated by black bars on the side of the table cells. Double bars indicate CDRs for which the X-ray structure adopts a minor canonical conformation (classification according to Ref. 33, adopted from Ref. 11. Averages are shown for the entire set of models (All), as well as model 1 for each target (M1) (excluding the rabbit structure Ab01, which was considered an outlier because of the low number of rabbit templates in the PDB). The ranking of models was determined manually. Note that the carbonyl RMSDs reported here can be substantially higher than the more commonly reported C-α RMSDs (reported in the Supporting Information Table S1).

templates, we picked only three from the default list with the highest sequence similarity for the overall Fv region (3MLW, 2XWT, 3N9G). The remaining two templates from the default list had comparatively low sequence similarity for the light chains and were replaced by a manual selection of templates with higher light chain similarity (4D9L and 4FQJ). Analyzing the germlines of Ab05 and the selected templates explained why this was not a good choice; Ab05 is composed of germlines IGLV1-40 and IGHV5-51.[¶] All of the top four default overall templates also had VH domains from the VH5 germline family and VL domain from the VL1 germline family, matching the target exactly. By contrast, three of the templates we used in the manual selection, 3N9G, 4D9L, and 4FQJ have VH domains from the VH1 family. The manually selected templates produced models with

[¶] As determined by the IMGT/GapAlign server.[31]

VH RMSDs of 0.8 Å. By contrast, the VH domain of the chimeric model accAb05m2, which was based on 3NA9 from the VH5 germline family, had a VH RMSD of 0.3 Å. As pointed out in Ref. 11, one of the distinguishing features of the VH5 germline is a buried TRP residue in Chothia position 82, which leads to a displacement of B and E β-strands of ∼2 Å compared to VH domains from other germlines, which typically feature a LEU or MET residue in position 82.

In retrospect, given the structural differences between members of the VH5 and VH1 germline family, selecting templates from the same germline family should have trumped selecting by higher light chain similarity. As the results for the automated methods show, this would have led to better models. It is not clear to what extent this generalizes to other germline families.

FV domain. RMSD values for the overall framework region, including both the VL and VH domains, are listed in the C-β and FR columns of Table II. The C-β columns measures the RMSD only over the β-sheet core residues used for superposition, as defined previously in the methods section. The values in the FR column include all residues in the domain, except the CDR regions, and the termini residues. Additionally, the table lists the tilt angle, which compares the relative orientation of the domains between the model and X-ray structure.

Given that the quality of the VL and VH domains is good for most models, large values in the RMSD of the overall framework should be explained by the relative orientation of the domains. The tilt angles of the models are also plotted in Figure 2(c). The data in Table II shows that the FV β-core RMSDs for most models are around 0.7 Å or below, with the exception of models for targets Ab04, Ab05, and Ab11, which have RMSD values around 1 Å. The tilt angle correlates relatively well with the RMSD of the β-core region. Figure 2(c) shows that the models submitted by Accelrys typically have tilt angles near the median for the submitted models. Models submitted for targets Ab08 and Ab09 have very good tilt angles below 2°. On the other hand, the deviation of tilt angles for models for targets Ab05 and Ab11 are comparatively high.

For target Ab11, this can be explained by the use of a sub-optimal interface template. Models for target Ab11 were built using the chimeric template approach, with 2W9D as the interface template. We selected 2W9D, rather than the highest sequence similarity template (2OZ4) as the interface template because it was the top heavy chain template per sequence similarity and was among the top 10 overall templates. This approach of choosing the top heavy or light chain template as the interface template was an approach successfully applied in the first Antibody Modeling Assessment experiment.[10] However, in the case of target Ab11, this was not an optimal approach. This can be seen in Figure 1(a), which shows that any of the top five templates had lower RMSDs than 2W9D. The default setting in our software would have picked 2OZ4 as the interface template, which has a sequence similarity and identity of 94.5% and 86.1% respectively. 2W9D, as the 8th best template by comparison has a sequence similarity and identity of 90.0% and 73.4% respectively, which is significantly lower. Indeed, single and chimeric models using 2OZ4 as interface templates produce models with tilt angles between 6° and 7°.

Similarly, the interface template for chimeric model accAb05m2, 3N9G, was also a lower ranked overall template (ranked 5th with sequence similarity of 91.1%), which was manually chosen over templates with higher overall sequence similarity because all the top templates were ligand-bound. However, 3N9G has a much higher RMSD with respect to the target X-ray structure than the top template [cf. Fig. 1(a)]. Here a chimeric model built on the top template (3MLW, 96.1% sequence similarity) has a tilt angle of around 6.5°. Again, an indication that this manual choice was ill-advised is provided by the germline makeup of 3NG9 (IGLV1, IGHV1), compared to the makeup of the target and the default template 3MLW (IGLV1, IGHV5).

For the multi-template models for target Ab05, the high tilt angles are most likely due to the choice of templates with VH domain from a germline other than VH5, as discussed previously. Interestingly, the two templates with VH domains from the VH5 family, 3MLW and 2XWT, have tilt angles of 4.9° and 6.4°. The templates with VH domains from the VH1 germline family all have tilt angles above 10°, with 3N9G, which was the template selected for the chimeric model accAb05m2 having a tilt angle of 12.0°. Not surprisingly, the models built on these templates have comparatively high tilt angles.

There seem to be some interesting systematic differences between structures and sequences from VH from germline VH1 compared to those of germline VH5 around residue H60, a region in close proximity to L3. However it is not clear from this limited set of data to what extent germline specific properties influence the tilt angle and by what mechanism. A more detailed examination of this issue is warranted.

Automated models

Figure 2 shows the performance of the different framework modeling approaches for automated modeling. The single and chimeric approaches typically perform similarly to the corresponding submitted models, though the chimeric approach in some cases, such as target Ab07, performs better in modeling the VL and VH domains, as expected.

For the automated approach we do not expect much difference in tilt angle between the single and chimeric
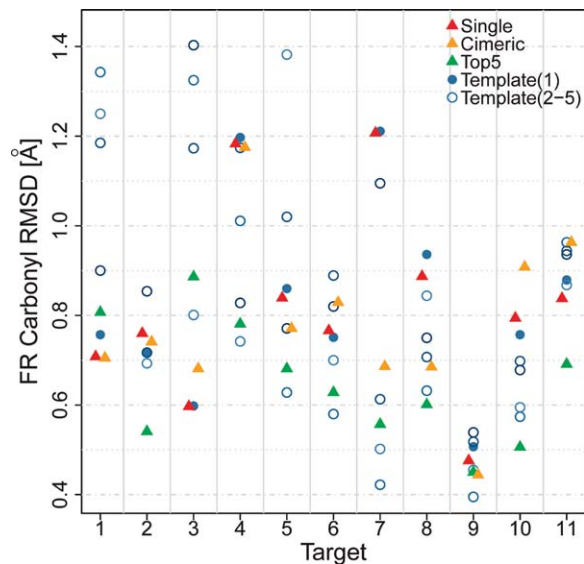
Figure 3

Peptide carbonyl RMSD of the Fv framework region for five most sequence-similar templates and models built using automated methods during post-experiment analysis. The filled blue circles indicate the RMSD for the template with the highest sequence similarity; the open circles indicate the RMSDs for templates with the second to fifth highest sequence similarity. The plot also shows RMSDs for the models built using the single template (red triangles), chimeric template (orange triangles) and top five templates (green triangles) methods. The single and chimeric model RMSDs tend to be close to that of the top template ranked by similarity, whereas the RMSDs for the top five models tends to be close to that of the template with the lowest RMSD, and sometime even below.

template approaches since in both cases, the same structure (highest sequence similarity template), was used to model the overall orientation of the VH and VL domains.[#] There is no clear trend to which method, single or chimeric, performs better on average. Single template models on average have a RMSD of 0.67 Å $\pm$ 0.18Å over the β-core region, compared to 0.65 Å $\pm$ 0.12 Å for the chimeric template models. On the other hand, the top5 template approach seems to outperform the other approaches on average, with 0.53 Å $\pm$ 0.12 Å RMSD for the β-core regions. However due to the small sample size, the difference is not statistically significant at the 5% level.

For the VH and VL domains, with the exception of VL for targets Ab01 and Ab09, the top five template models is always better than the median of all submitted AMA-II models and it is often as good as the best submitted models. The performance of the multiple template approach is even better for the tilt angle; for 9 out of the 11 targets, it produces a model with a tilt angle of less than 6°. This is somewhat surprising. To better

understand this, we looked at the templates used in the automated top5 template approach. Figure 3 shows the framework RMSDs to the target X-ray structure for all five templates involved in model construction, along with the RMSDs of the corresponding top5, chimeric and single template models for each targets. It is clear from the figure that the top template does not always have the best RMSD as already discussed in the section on template selection. Also note that the single template and chimeric models have RMSDs very close to that of the top template,[**] which is expected. Somewhat surprisingly, the RMSD of the model produced by the top5 multiple template approach is almost always near the low end of the RMSD range of the 5 templates used, and sometimes below the lowest template RMSD.

This can most likely be explained by the way MODELER combines information from the different templates: As mentioned in the methods section, MODELER uses an additive distance restraint function that incorporates information from each template, weighted by local sequence similarity. This results in models that locally agree with one or another of the templates, not the average of all templates.[3,32]

For the top5 template approach used here, this effectively ignores outlier regions in the templates: If three or more of the templates agree on the local conformation or some templates have more similar sequences in a region, models that agree with this template conformation will have a better energy than models based on the deviating structures. An illustration of this effect can be seen in Figure 4, which shows a region of the backbone near the N-terminal for target Ab08. In this region, two of the templates, among them the top template 1PKQ, deviate from the conformation of the backbone of the target. However, the other three templates agree with the target structure, which results in the top model built by MODELER being very similar to the target in this region. It is possible that this is a general effect; because the templates in antibody modeling typically have high sequence similarity to the template and by extension to each other, they will typically have similar framework structures, with some local deviation for individual templates. The top model built by MODELER is then one that is based on the most common or most similar local substructures, thus eliminating outliers. As long as the target agrees with these local conformation most of the time, the overall RMSD is likely to be lower than any of the templates with local deviations.

An interesting case is target Ab03, where the top5 template approach produced a model with a tilt angle of about 9°. Closer inspection reveals that three of the five templates in this case were from the same structure,

---

[#]Minor differences are expected though since the core regions in the VH and VL domains will be different for the chimeric and single template approaches.

[**]The small differences are due to the somewhat different IMGT CDR ranges we used in the loop remodeling process during the experiment, compared to the CDR ranges adopted from Ref. 11 for the post-experiment RMSD calculations.
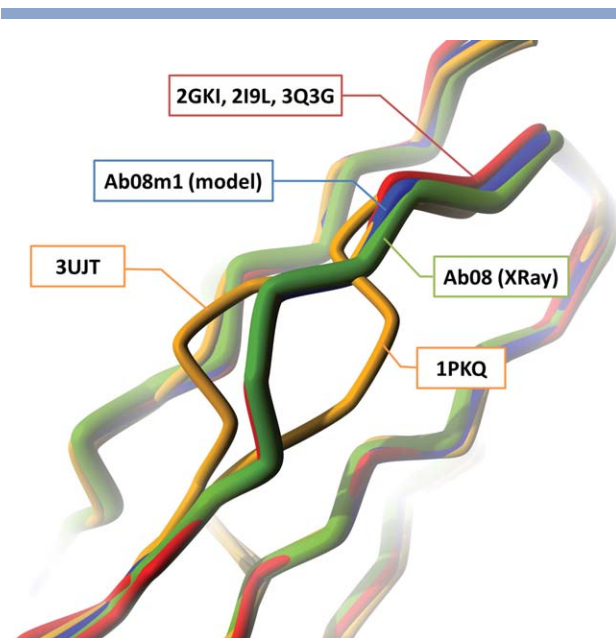
Figure 4

Backbone conformation of N-terminal of VH domain for X-ray structure (green), model accAb08m1 (blue), and templates 1PKG (orange), 2GKI (red), 2I9L (red), 3Q3G (red), 3UJT (orange). The range shown is about residues 3 to 13 based on Chothia numbering. X-ray structure, model and templates 2GKI, 2I9L, and 3Q3G superimpose well in this region, whereas 1PKG and 3UJT adopt a different backbone conformation.

1RZI, which has eight copies of the Fv region in the unit cell, all closely interacting. The automated template selection procedure for the top5 approach currently only removes 100% sequence redundant templates. This caused three Fv regions from 1RZI to be picked up, since they have slight differences in their sequences due to residues not resolved in the X-ray structure in some of the domains. Having multiple copies of a very similar structure would most likely bias the top5 approach towards that structure, which seems to be what happened in this case. Further testing should be done to see if adding a filter to remove copies of Fv domain from the same PDB entry would improve the results.

Given the relationship between RMSD and sequence similarity of templates shown in Figure 1, there seems to be no clear signal in the sequence similarity to pick out a "best" individual template. A multiple template based approach with best local conformations modeled from multiple templates is likely to produce a better model.

We reach the same conclusion when testing using the nine antibodies from the AMA-I experiment (e.g., framework models built using the top five templates had lowest RMSD).

However, current results are based on a limited set of data; a more systematic study on a larger set of data is needed to confirm the validity of this approach.

Similarity versus identity

We also explored the effect of using sequence identity as opposed to sequence similarity when selecting the templates used in building models. For individual targets, such as target Ab11, the choice of template by identity rather than similarity can make a big difference. This is expected given the relatively large variance of template RMSD observed in Figure 1. For templates with a relatively small difference in sequence similarity to the target (cf. Fig. 1) there is relatively large variability in the models built using either approach. However, on average we did not measure a statistically significant difference between choosing templates by sequence identity versus sequence similarity. In most cases both approaches result in the same selection, so the number of data points from this experiment is too small to draw meaningful conclusions. A more systematic study on a larger dataset is required.

CDRs

Results for the predictions of the CDR loops are presented in Table II and Figure 5. In general, predictions for L1, L2, and L3 were accurate—for L2, all predictions were 0.5 Å or less, and for L1 and L3, most predictions were below 1Å. Exceptions with higher RMSDs were the L1 models for targets Ab01, Ab04, and Ab05 and the L3 models for targets Ab01, Ab02, Ab05, and Ab10. This is not surprising as these loops are either non-canonical, or in a minor canonical conformation according to the classification by North et al.,[33] as pointed out by the organizers in their general evaluation.[11] Also, target Ab01 is a rabbit antibody which is hard to model since there are very few templates from the same organism in the PDB.

One case that turned out to be interesting upon further examination was L3 for target Ab05. Our models for Ab05 have an incorrect conformation of the N-terminal of the VL domain. Rather than forming an extended strand as in the target Ab05, the residues assume a curved conformation. This led to an unusual problem in the modeling of the L3 loop. Since the N-terminal is relatively close to the L3 loop, the incorrect position of the N-terminal residues led to incorrect distance restraints during the building of the L3 loop model with MODELER based on 2J6E. In order to satisfy the spurious restraints from the incorrect N-terminal in the model, the L3 is bent away from the correct orientation in the template. This leads to the relatively high RMSD of L3 in our models. If the L3 model is rebuilt based on an input structure for which the first three residues are removed from VH using 2J6E as a template, the resulting L3 model has an RMSD of ~1.8 Å. This is still not particularly accurate, but better than the L3 in the submitted models. While this case is relatively rare, it is not currently well handled automatically and further work is
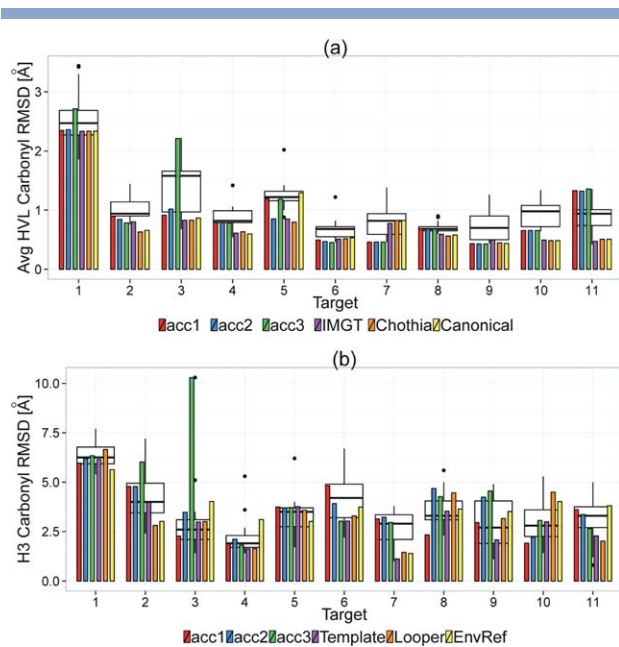
Figure 5

Plots in this figure compare the quality of the CDR loops of models for each target. In each panel, the results for the Accelrys models submitted for AMA-II are shown as red (model 1), blue (model 2), and green (model 3) bars. The box plots in the background indicate the distribution for the models submitted by all groups participating in AMA-II (as reported in Ref. 11[††]). (a) Plots average peptide carbonyl RMSD for non-H3 hypervariable loop regions of the model structures compared to the X-ray structure. RMSDs are calculated by averaging the peptide carbonyl RMSDs for the L1, L2, L3, H1, and H2 CDRs (i.e. the average of the corresponding columns in Table II). The results for models built using different automated approaches for CDR refinement in the post-experiment analysis are shown as purple (IMGT), orange (Chothia) and yellow (Canonical) bars. (b) Plots the H3 backbone carbonyl RMSD after superposition of the VH β-core regions. The results for models built using different automated approaches to model H3 in the post-experiment analysis are shown as purple (Template), orange (Looper) and yellow (Environment Refinement) bars.

needed to improve the method to reduce the chance of spurious restraints.

For the H1 and H2 loops predictions submitted were not as accurate as for L1, L2, and L3 loops. Although these loops seem to be more difficult to predict, as this trend appeared to be true for all participants,[11] most of the predictions were still fairly accurate, with average RMSDs of ∼1 Å.

One exception is target Ab11, for which our predictions for H1 and H2 have RMSDs of 2.6 Å and 2.3 Å RMSD respectively. As discussed above, the origin of the problem stemmed from the selection of the 2W9D template for VH. For H1 and H2, the stems of CDR regions of this template do not superimpose well with the target structure. As mentioned above, 1F11 would have been a better choice than 2W9D because the stem

residues for the H1 and H2 CDRs are better preserved in 1F11 with respect to the target sequence. Though the stem residues are taken into account for CDR template selection, they are not rebuilt during the model building process. Since the stems in this case do not superimpose well, the rebuilt loops in the model are deformed substantially in order to accommodate restraints imposed by the different spacing and orientation of the stem residues. Indeed, a chimeric model built based on 1F11 using the same loop templates has RMSDs for H1 and H2 of 0.6 and 0.5 Å. Interestingly, if the H2 model had been refined based on the IMGT CDR definition, the resulting models also would have been better because the IMGT CDR definition for H2 is longer by two residues at either stem. This would have caused the incorrectly oriented stems to be replaced and a model using IMGT with 2W9D as framework template has an H2 RMSD of 0.7 Å.

The H1 CDR was also a problem in target Ab03. In this target, we identified H1 templates with 100% identity for residues 22-36, which includes H1 and the stem regions on either side, and residue 94, which is part of the canonical loop definition according to Chothia.[20] For models accAb03m1 and accAb03m2 these templates (2NPS_B, 3QOT_H, and 1RZI_B) were used for modeling H1. However the H1 conformations of these templates have RMSDs of 4.1 Å, 3.3 Å, and 2.3 Å respectively. The top template found during retrospective analysis was from 2CMR, with an H1 RMSD of 1.4 Å. However, 2CMR has an ASP at residue 27, compared with GLY in Ab03 and the other templates listed above. Additionally, because H1 in 2CMR is interacting with an antigen and because it has a lower sequence identity, it was not an obvious choice as template. The third model target Ab03, accAb03m3, was submitted without template based refinement of CDRs. The heavy chain CDR regions of 2XTJ, the single template for accAb03m3, show comparatively low similarity to the corresponding CDRs in Ab03 when compared to some of the other available CDR templates. As a consequence the inaccurate predictions for the VH CDRs in accAb03m3 are not very surprising.

We ran an automated post-experiment evaluation of different approaches to CDR loop modeling, comparing loops predicted based on the IMGT[17] versus the Chothia[18–20] CDR definitions. In addition, we considered filtering loop templates by canonical type. However, there was no significant difference between the different approaches, which was surprising. A closer inspection of the data revealed that the templates selected by the IMGT, Chothia, and Canonical filtering approaches were in fact identical in a majority of the cases. Even in cases where there was a difference, only one or two of the three templates were different. Given this, we would not expect any significant difference in the averages between the methods over such a small sample size.

[††]Data used for boxplots in figure is from a prepublication manuscript provided by organizers.

There were cases where individual approaches showed advantages. For example, for the L1 of target Ab04, a hydrogen bond between TYR71 and the backbone nitrogen of THR31 causes a flip in the peptide bond at that position with respect to structures that do not have TYR at position 71. TYR71 is part of the canonical loop definition for Chothia canonical type 2B. In the automated runs, all templates identified using the canonical filtering method had the correct conformation for the peptide bond at position 31, whereas in the runs without canonical filtering, only one of the three templates had the correct peptide bond flip. Consequently, the model from the canonical filtering method had an L1 RMSD of 0.6 Å, whereas the L1 in the IMGT model had an RMSD of 1.1 Å. Interestingly, the loop from the Chothia based model has the correct peptide bond orientation despite the fact that only one of the templates had this conformation, indicating that MODELER managed to select the correct local template in this case, possibly due to the sequence identity at residue 71. The submitted models (accAb04m1, accAb04m2, accAb04m3) were generated without considering canonical types and as a consequence had comparatively high RMSDs ($\sim$1.2 Å). It is worth noting that this important structural error is only apparent from the carbonyl RMSDs. The corresponding C-α RMSDs for L1 for the submitted models for target Ab04 are 0.4 Å (cf. Supporting Information Table S1), which highlights the importance of using carbonyl RMSDs when assessing the models.

Similarly, in some cases it was advantageous to select the IMGT CDR definition, as was the case for loop H2 in target Ab11, as discussed above.

From this we conclude that while in most cases all methods result in the same CDR template selection, there is some anecdotal evidence that considering canonical types, which encode information about key residues outside the actual CDR region, can improve models in some cases. Unfortunately, the sample size in this experiment is too small to draw conclusions with statistical significance.

CDR H3

As expected, predicting the conformation of the H3 loop in the models was the most challenging task. The prediction targets included a diverse set of loop lengths: five 8-residue loops (targets Ab01, Ab03, Ab04, Ab05, Ab07), two 10-residue loops (Ab09, Ab11), three 11-residue loops (Ab02, Ab08, Ab10), and one 14-residue loop (Ab06). The results of our predictions are listed in the H3 column of Table II and in Figure 5(b). For some of the shorter loops (targets Ab03 and Ab04) as well as the 11-residue loop in target Ab10, we made some reasonable predictions around 2 Å. As expected, for longer loops, the predictions were not as accurate.

It is interesting to notice that in most cases where we submitted a template based model in addition to an ab initio model, the template-based model tends to do better. Furthermore, for the automated models generated in the post-experiment analysis, we did not observe any improvement from using the Looper or the environment refinement approach over the purely template based models for H3. This is probably a reflection of the fact that the ab initio loop prediction method is based on physical energy, which to a large degree depends on a very accurate environment for the loop (typically not the case in a homology model).

Model quality

In general, the models submitted for the first stage were high quality. When evaluated with MolProbity,[34,35] the quality scores are generally in line with those of the target X-ray structures (see Supporting Information Table S1 of Ref. 11 for a full list of MolProbity scores).

In particular, the models had on average 96.5% of residues in Ramachandran favored areas, with 0.3% outliers, which indicates good backbone geometry. This is an improvement over the models submitted for AMA-I for which the percentage of residues in Ramachandran favored areas was 91.9%. The average clash score for the models was also good (4.2); it was only marginally higher than that of the X-ray structures at 3.1.[‡‡]

These results indicate that the constrained minimization used as last step in the model building process was clearly beneficial in improving model quality (such a step was not included for AMA-I); it relieves local stress in the models while preserving the overall conformation of the models due to the constraints used in the minimization.

In terms of *cis-trans* isomers, the models submitted exhibited few problems. All of the cis-prolines in the target were modeled with the correct conformation (cf. Table 4 in the general assessment[11]). There were five cases where an incorrect isomer was copied into the model from a template. This was the case for the models for target Ab03, where a cis isomer for GLY104 in VH was copied from template 2XTJ for models 1 and 3. For target Ab04, two of our models incorrectly copied the cis-isomer for HIS8 in VL from template 3MXV, which has a cis-proline at this position. Finally for model 1 in target Ab10, there is an incorrect cis-isomer for GLY100 in VH which seems to have been introduced during the H3-refinement stage.

Stage 2

Supporting Information Table S2 shows the results for the prediction of the H3 CDR loop. The first column

---

[‡‡]Unfortunately, a comparison of the clash scores between our AMA-I and AMA-II models is difficult since the hydrogen parameters used in the current version of MolProbity have been updated since the version that was used to evaluate AMA-I models.[36]

labeled acc-m0 shows the RMSD of the best model from the first stage, whereas the remaining columns show the RMSDs of the models submitted for the second stage. With the exception of targets Ab10 and Ab11, the first model from the second stage is better than the best model from the first stage. This is not surprising since predicting a long loop is easier in its crystal environment than when the prediction is based on a model structure. For the shorter loops our predictions were generally good, with predictions of 1 Å or less for the eight residue loops for targets Ab03, Ab04, and Ab05, and models with less than 2 Å for target Ab07 (also an eight residue loop). However, we did not always choose the best generated loop conformation to be our top model. This was the case for target Ab05, where we produced a very good model with 0.8 Å RMSD, but picked the 2.8 Å conformation as our first model.

As expected, the model quality drops for the longer loops, with predictions in the range of 2 to 5 Å, with some reasonable predictions for the 10- and 11-residues loops for targets Ab08, Ab09, and Ab10.

A further analysis of the whole ensemble of loops generated during prediction reveals that for the longer loops, the problem was often due to insufficient sampling. For a majority of the longer loop targets (Ab02, Ab06, Ab08, Ab09), no acceptable loop conformation (i.e., below 2 Å RMSD) were among the conformations sampled. For targets Ab10 and Ab11 acceptable conformations were generated (0.8 and 1.0 Å), but only for target Ab10 was a reasonable conformation selected for the models submitted.

Because our approach required relatively short computation times (typically less than 30 min), the results were not unexpected. However, this severely restricts the amount of conformations sampled, which can be a major limitation for longer loops. The results of this experiment (and other studies) indicate that in order to achieve more accurate predictions, more extensive sampling is required.

However, such resources might not be available for typical scientists wanting to build models for a large number of sequences, and the approach used here produces a reasonable model even with relatively limited computational resources.

CONCLUSION

Our antibody modeling tools have greatly evolved since the first Antibody Modeling Assessment (AMA-I) experiment in 2009. Based on the evaluation of our models submitted for this blind prediction study, we conclude that our methods are state of the art (see Supporting Information Table S4 and Ref. 11 for comparison to other AMA-II groups) and produce accurate models with RMSDs of the VH and VL framework regions below 1 Å in most cases. Similarly, predictions for the L1 and

L2 CDRs are typically accurate, while predictions for L3, H1, and H2 are generally a bit less accurate, but still around 1 Å on average if the outliers discussed previously (Ab01, Ab05, Ab11) are excluded.

The RMSD values of the models we submitted for AMA-II on average are lower across the board than the corresponding numbers for the models submitted to the first Antibody Modeling Assessment.[§§] The source of the improvement is two-fold: The first is improvements in our methods, including a new curated antibody structure and sequence database, better ways to select framework and CDR templates based on the database, more robust methods to graft CDR loops using MODELER and the use of more consistent workflows.

The second source of improvement is better database coverage. Since AMA-I, the number of antibody X-ray structures in the PDB has greatly increased, making it more likely to find one or several very close templates for a target. (Indeed, the average sequence identity for top templates during AMA-I was about 83%, whereas in AMA-II, it was about 90%).

We did try to address the question of the relative contribution of these two sources by re-predicting AMA-I models using a database consisting of only structures available during that experiment, but using recent methods (unpublished data). The results indicate that improved methodology plays a central role in the improved results.

The analysis presented above suggests that the best approach for building a model for a new antibody is to consider a number of templates with the highest sequence similarity to the target structure as potential templates. If there are a number of templates with very similar sequence similarity, building a model based on multiple templates seems to yield the best result. If a model on a simple or chimeric template is built, the available templates should be inspected to avoid outliers. When building CDRs it also appears beneficial to use multiple templates. In general, it does not appear to make much of a difference which particular loop definition (e.g., Chothia or IMGT) is used, since both seem to result in identical or very similar CDR template selection. However, in cases where the CDR definitions are very different, one should check which definition better captures the local variation of potential templates for a given CDR to ensure variations in stem regions do not lead to inaccurate models. For H3 refinement, ab inito remodeling of the region does not seem to offer much benefit in terms of the RMSD from the real structure. However, one has to note that ab initio remodeling often resolves other structural problems, such as steric clashes, and can therefore be important in preparing a structure for further use. Similarly, further minimization of a structure can help in resolving such issues, but care should be taken to restrain unaffected parts of the

---

[§§]Comparison based on backbone atom RMSD, which was used in the evaluation of AMA-I.[10]

structure as to not introduce inaccuracies due to the minimization.

It also appears that our tools have evolved to the point that manual intervention does not always lead to better models. While manual intervention helped in many cases, there were several instances where a decision by a predictor to select one template over another actually lead to a worse model than using the default template suggested by the automated approach. Of course, domain experts with in-depth knowledge of the targets would most likely always be able to improve models compared to an automated approach. On the other hand, our results and analysis indicate that the automated approaches in most cases are sufficient to create quite accurate models without direct manual intervention.

The AMA-II experiment also pointed to a number of areas we will explore to further improve our tools. The first concerns CDR refinement. As discussed above for the refinement of CDR L3 of target Ab05, there are cases where our current approach can lead to spurious restraints. While this seems to occur rarely, it needs to be addressed by more careful screening of restraints used in CDR loop refinement. A second area that needs improvement is H3 modeling. Our current methods are designed to rapidly build fairly accurate models for relatively short loops (ten or fewer residues). This speed is achieved in part by limiting the amount of sampling performed, which compromises the accuracy of the longer loops (e.g., loops with 12 residues or more). For the framework model construction, the top5 template approach looks promising. In general, selecting the top five sequence-similar templates seems to perform very well. However as some of the cases discussed show, the approach does not work well in all cases and more work is needed to determine more specific rules to guide automated template selection for multiple template models.

More generally, the results of the AMA-II experiment indicate that selecting templates based purely on sequence similarity does not always identify the optimal templates, and that additional criteria might improve the quality of the selected templates. However, the dataset of the current study is not large enough to clearly identify the factors that would allow for a more reliable selection of templates. Studies on larger datasets are required.

ACKNOWLEDGMENTS

REFERENCES

1. Al-Lazikani B, Jung J, Xiang Z, Honig B. Protein structure prediction. Curr Opin Chem Biol 2001;5:51–56.
2. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins 1995;23:ii–iv.
3. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
4. Sircar A, Kim ET, Gray JJ. RosettaAntibody: antibody variable region homology modeling server. Nucleic Acids Res 2009;37:W474–W479.
5. Molecular Operating Environment (MOE). Montreal, QC: Chemical Computing Group Inc. 2013.
6. Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. Antibodies in HIV-1 vaccine development and therapy. Science 2013;341:1199–1204.
7. Sliwkowski MX, Mellman I. Antibody therapeutics in cancer. Science 2013;341:1192–1198.
8. Chen R, Li, L, Weng ZP. ZDOCK: An initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
9. Sircar A, Gray JJ. SnugDock: Paratope structural optimization during antibody-antigen docking compensates for errors in antibody homology models. PLoS Com Biol 2010;6(1):e1000644.
10. Almagro J, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, Butenhof K, Labute P, Thorsteinson N, Kelly K, Teplyakov A, Luo J, Sweet R, Gilliland GL. Antibody modeling assessment. Proteins 2011;79:3050–3066.
11. Teplyakov A, Luo J, Obmolova G, Malia TJ, Sweet R, Stanfield RL, Kodangatti S, Almagro JC, Gilliland GL. Antibody modeling assessment II: structures and models. Proteins 2014;82:1563–1582.
12. Almagro JC, Teplyakov A, Luo J, Sweet W, Kondagantlil S, Hernadez-Guzman F, Stanfield, Gilliland GL. Second antibody modeling assessment (AMA-II). Proteins 2014;82:1553–1562.
13. Bernstein FC, Koetzle TF, Williams GJ, Meyer EE, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535.
14. Research Collaboratory for Structural Bioinformatics PDB. Available from: http://www.rcsb.org/pdb/. Last accessed 20 May, 2014.
15. Eddy SR. Profile hidden markov models. Bioinformatics 1998;14:755–763.
16. Eddy S. HMMER 3.1. Available from: http//:hmmer.org/. Last accessed 20 May, 2014.
17. Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 2003;29:55–77.
18. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 1997;273:927–948.
19. Chothia C, Lesk AM. Canonical structures for the hypervariable regions. J Mol Biol 1987;196:901–917.
20. Morea V, Arthur M, Lesk AM, Tramontano A. Antibody modeling: implications for engineering and design. Methods 2000;20:267–279.
21. Kuroda D, Shirai, Kobori M, Nakamura H. Structural classification of CDR-H3 revisited: a lesson in antibody modeling. Proteins 2008 608–620.
22. Shirai H, Kidera A, Nakamura H. H3-rules: identfcation of CDR-H3 structures in antibodies. FEBS Lett 1999;455:188–197.
23. Shirai H, Kideral A, Nakamura H. Structural classification of CDR-H3 in antibodies. FEBS Lett 1996;399:1–8.
24. Spassov VZ, Flook PK, Yan L. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. Protein Eng Des Sel 2008;21:91–100.
25. Spassov VZ, Yan L, Flook PK. The dominant role of side-chain backbone interactions in structural realization of amino acid-code. ChiRotor: a side-chain prediction algorithm based on side-chain backbone interactions. Protein Sci 2007;16:494–506.
26. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy

minimization and dynamics calculations. J Comp Chem 1983;4: 187–217.

27. Brooks BR, Brooks CL, III, Mackerell AD, 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. CHARMM: the biomolecular simulation program. J Comp Chem 2009;30:1545–1615.

28. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:823–826.

29. Kosloff M, Kolodny R. Sequence-similar, structure-dissimilar protein pairs in the PDB. Proteins 2008;71:891–902.

30. Sela-Culang I, Alon S, Ofran Y. A systematic comparison of free and bound antibodies reveals binding-related conformational changes. J Immunol 2012;189:4890–4899.

31. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobu-lins or antibodies, T cell receptors, MHC, IgSF and MhcSF. Nucleic Acids Res 2010;38:D301–D307.

32. Sánchez R, Šali A. Evaluation of comparative protein structure modeling by MODELLER-3. Proteins 1997;Suppl 1:5–58.

33. North B, Lehmann A, Dunbrack RL, Jr. A new clustering of antibody CDR loop conformations. J Mol Biol 2011;406:228–256.

34. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D 2010;66:12–21.

35. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375–W383.

36. Richardson Lab, Duke University. MolProbity, Main Page. Available from: http://molprobity.biochem.duke.edu/. Last accessed 8 October, 2013.