DASSAULT SYSTEMES

BIOVIA

ANTIBODY STRUCTURE MODELING WITH DISCOVERY STUDIO

APPLICATION BRIEF

This application brief describes the state-of-the-art *in silico* Antibody Structure Modeling workflow implemented in BIOVIA Discovery Studio® software and highlights key reasons why structural models of antibody targets are needed.

INTRODUCTION

Antibodies are increasingly important in medical diagnostics and in the treatment of a broad range of disease states including cancer, inflammation and auto-immune diseases. Through antibody drug conjugates (ADCs), antibodies also enable the targeted delivery of traditional drugs. In contrast to traditional chemotherapeutic agents, ADCs target and attack the cancer cells so that healthy cells are less severely affected.

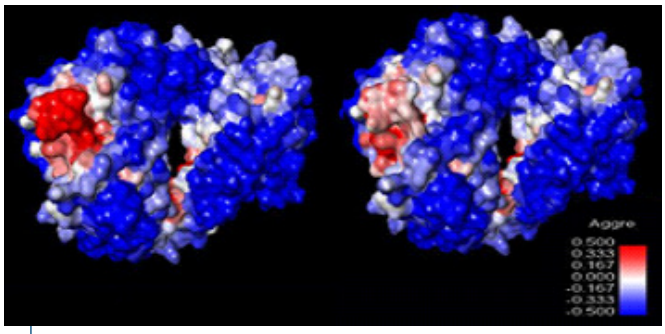

Figure 1: The Aggregation prediction of a target before and after mutations designed to maintain affinity and reduce aggregation (aggregation-prone regions in red).

Many critical properties of antibodies cannot be predicted from sequence alone, or can be predicted more accurately with access to a 3D antibody structure. A few examples of the usefulness of quality 3D structure models are:

- Humanization
- Identify exposed and buried residues to assess likely antigenicity
- Maturation
- Predict changes in stability or antigen binding (avidity) with in-silico mutations (including f(pH) or f(temperature)
- Formulations
- Rank targets by developability index, or identify patches with high predicted aggregation propensity (figure 1).

Improving the understanding of these properties as early as possible in development and formulation will impact speed and total cost to market. When 3D structures are not available from X-RAY or NMR, 3D *in silico* models generated by BIOVIA Discovery Studio are the solution.

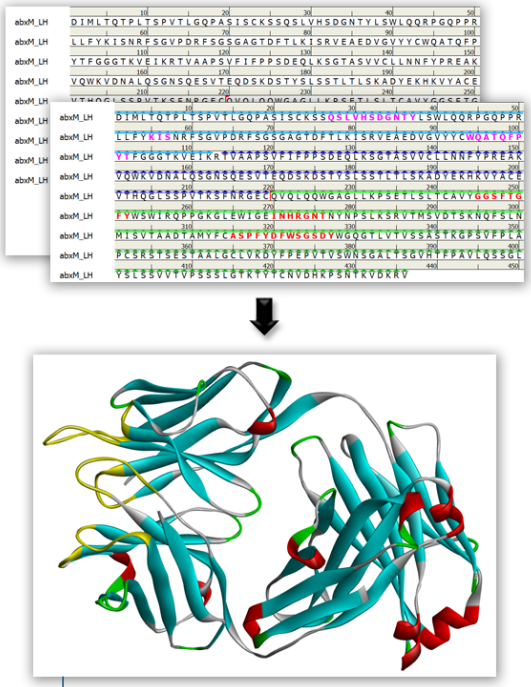

Figure 2: The generation of a structure proceeds from sequence using multiple templates for framework residues and CDR loop residues.

METHOD

BIOVIA Discovery Studio provides all of the capabilities needed for experts and non-experts to generate a quality antibody structure based on sequence input (figure 2). The modeling workflow consists of three principle stages as described below.

1. Identify Framework Templates

- Find quality crystal structures with suitable framework region sequence similarity.
- Results can optionally include complementarity-determining region (CDR) similarity and filter by organism.

2. Generate 3D model based on Framework Templates

- Multiple Framework Templates can be used as input. Conformations are reduced to restraints that are then used to anneal the target conformation using homology modeling software (Modeler).

3. Refine Antibody Loops to predict CDR conformations

- Refines the highly sequence variable CDR loops based on homology to multiple quality CDR loop templates, or if none exist then apply de novo force field-based methods.

RESULTS

This application brief is based on our participation in Antibody Model Assessment II[2] (hereafter referred to as AMA-II) in which conformations for eleven antibody Fv sequences were predicted and evaluated relative to known but unreleased crystal structures (a "blinded" exercise). In the analysis of the predicted structures multiple metrics were evaluated for their ability to distinguish the highest quality model structures. This study showed that backbone root-mean-square deviation (RMSD) is not a sensitive indicator of structure quality. That traditional metric does not capture the critical backbone orientation. The AMA-II [2] study recommends the use of RMSD based on the amino acid carbonyl group atoms C and O;, therefore, this is the metric we use below (RMSDCO). In this study we tested a variety of work flows as described in [1]. Briefly, these consisted of modeling the framework regions (FR) using a single template, or separate optimal templates for the L and H chains, or using multiple templates for the L and H FR. For the CDR loops the CDR length is dependent on the annotation method used, and filtering based on canonical loop type prediction is an option. Subsequent automated structure determinations were applied to all eleven models to more rigorously evaluate the relative merits of these methods. Results are summarized in Figure 3. This shows consistent quality results which on average outperform other currently available methods. Also, it demonstrates that fully automated structure generation can achieve similar accuracy to those generated by expert modelers using modest CPU power and time restraints [1].

DISCUSSION

The antibodies in AMA_II included some quite challenging targets. For example, Ma2-01 (model assessment-II, structure 01) was from Rabbit which has far fewer templates than Human or Mouse. Ma2-05 contains a lambda light chain which are far less common in the crystal database than kappa. Ma2-10 contained a long H3 loop (16 residues using IMGT).

The protocols developed and available through BIOVIA Discovery Studio are designed such that the output provides suggested templates in ranked order based on similarity score and template resolution, allowing easy selection when the protocols are run in a semi-automated workflow. The best results on average were achieved when multiple templates were selected both for generating the framework regions and when determining the optimal conformation of the CDRs using the **Model Antibody Loops** protocol. This is because where the degree of sequence homology differs relative to the template, or different conformations are present for templates with identical sequence, the Modeler solution engine is able to refine the target conformation using a weighted combination of restraints derived from all templates. Figure 3 shows the VL and VH RMSDCO values relative to crystal structure for the three user-submitted models for each target in AMA_II and three automatically generated models. The boxes shown in this figure relate the quality of our predictions to those of the structures submitted by other participants of AMA-II (Figure 3, see caption).

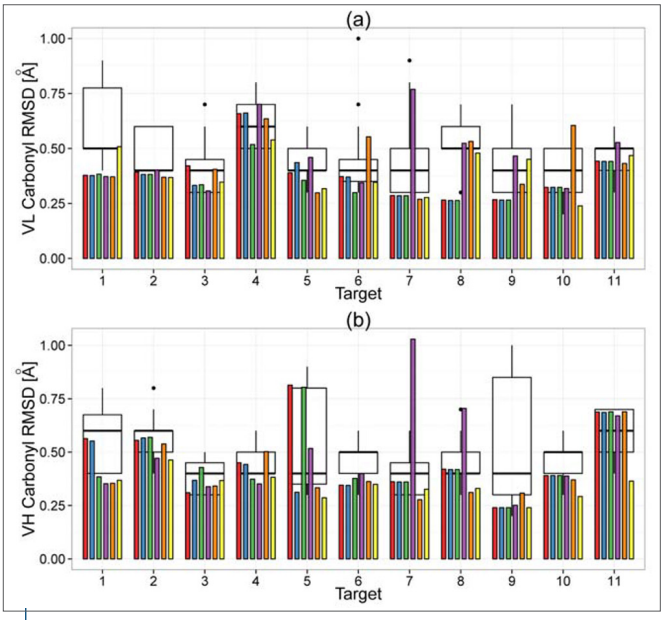

Figure 3: Results for the BIOVIA (Accelrys) models submitted for AMA-II are shown as red (model 1), blue (model 2), and green (model 3) bars. The results for different automated approaches in the post-experiment analysis are shown as purple (single template), orange (chimeric template) and yellow (top five templates) bars. In each panel, the box plots in the background indicate the distribution for the models submitted by AMA-II participants. 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 RMSDCO values that fall within a factor of 1.5 times the interquartile distance of the box boundaries. Any outliers falling in the regions beyond the tails are drawn as black circles. (a) Plots RMSDCO of the model structures compared to the X-ray structure calculated over b-core of the VL region. (b) Plots the same data for the VH region. This figure is extracted from Figure 2 of ref [1] where further discussion can be found.

The framework RMSDCO data demonstrates a benefit from the ability to choose separate heavy and light chain templates when the similarity difference between the alternative single templates is sufficiently large. Figure 3 shows that separate VL and VH framework templates at worst have no impact and at best allow a considerable improvement in RMSDCO relative to a single framework template (target 05, 07 and 08). As shown in figure 4, the CDR loop conformations achieved by **Model Antibody Loops** and BIOVIA Discovery Studio's de novo loop models provide state-of-the-art fidelity to reference crystal structures.

Group	RMSD[Å]								Tilt [°]
	VL	VH	L1	L2	L3	H1	H2	H3	
ACC	0.4	0.5	0.6	0.3	1.2	1.0	0.8	3.1	5.7
CCG	0.6	0.5	1.2	0.6	1.2	0.9	1.0	3.9	5.7
JEF	0.4	0.4	0.7	0.5	1.3	1.1	1.0	2.9	6.1
JOA	0.4	0.5	0.8	0.3	1.3	1.0	0.8	2.5	4.9
MMT	0.6	0.6	0.9	0.7	1.5	1.1	0.9	3.5	6.9
PIG	0.5	0.6	1.1	0.5	1.0	0.9	0.8	3.3	5.9
SCH	0.5	0.5	1.2	0.6	1.1	1.2	1.0	3.4	4.4

Figure 4: Average peptide carbonyl RMSD values for each group participating in AMA-II. Averages are for model 1 from each group, excluding the rabbit target, Ab01. Based on data presented in ref [2]. ACC is BIOVIA (Accelrys). RMSDco values shown are light chain only (VL), heavy chain only (VH), the five canonical CDRs , and CDR H3. Tilt refers to the deviation in the angle between the VL and VH chains relative to the crystal structure.

CONCLUSIONS

Antibody-based diagnostics and therapeutics are already a critically important part of the Life Science commercial economy with worldwide sales increasing six-fold since 2003 [3]. Reducing product time to market, and improving the ability to "fail early" in the Discovery phase are critically important. The traditional workflow following the creation of hybridomas is the in vitro optimization. This optimization seeks to reduce immunogenicity, increase avidity and improve stability by minimizing undesirable post-translational modifications (PTM) and reducing aggregation or viscosity. *In silico* structure predictions generated by the methods described above enable the prediction of key bulk and molecular level properties which could not be predicted by sequence alone. Conformational models allow prediction of the solvent accessibility and charge distribution at the surface of the molecule that are critical ingredients in estimates of the critical properties that are needed to evaluate candidate targets. Such structural models can come from single crystal X-RAY crystallography or NMR spectroscopy but these analytic methods are also expensive and require a significant investment of time. The alternative is modeling based on homology and de novo methods, as presented in this application brief.

The results of the blinded AMA-II study [1][2] demonstrated that antibody models generated by BIOVIA Discovery Studio have superior fidelity to the gold standard high resolution crystal structures based on the RMSDco comparison metric recommended as being the most sensitive by the independent researchers of the ABA-II assessment [2].

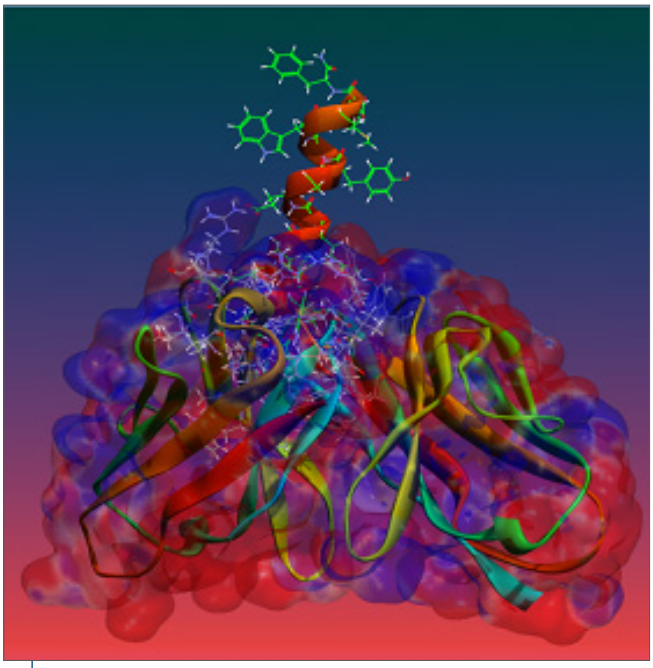

Figure 5: Anti-Gastrin Fv model structure with predicted docked conformation of gastrin predicted and rendered by BIOVIA Discovery Studio

Access the paper here: "Automated Antibody Structure Prediction using BIOVIA Tools: Results and Best Practices"

REFERENCES

1 Marc Fasnacht, Ken Butenhof, Anne Goupil-Lamy, Francisco Hernandez-Guzman, Hongwei Huang, and Lisa Yan, Proteins, 2014, in press, "Automated antibody structure prediction using BIOVIA tools: Results and best practices."

2 Almagro JC, Teplyakov A, Luo J, Sweet W, Kondagantlil S, Hernadez-Guzman F, Stanfield, Gilliland GL. Second antibody modeling assessment (AMA-II). Proteins, in press.

3 Monoclonal Antibodies, 2010, Datamonitor Ltd, Report code: HC0029-002.

DASSAULT SYSTEMES | The 3DEXPERIENCE® Company

Dassault Systèmes Corporate
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

BIOVIA Corporate Americas
BIOVIA
5005 Wateridge Vista Drive,
San Diego, CA
92121
USA

BIOVIA Corporate Europe
BIOVIA
334 Cambridge Science Park,
Cambridge CB4 0WN
England