IN SILICO PREDICTION OF ΔΔG AS A GUIDE TO IMPROVING STABILITY AND AVIDITY OF ANTIBODIES

Assessing BIOVIA Discovery Studio Predictions of the Change in Free Energy

A frequently cited goal of both pharmaceutical and diagnostics development is to reduce the time required to develop an effective antibody therapeutic or assay. Recent publications have shown that computational estimates of the effects of mutations ***in silico*** are capable of enriching mutations libraries toward those that improve binding or stability in the target environment.

K. Butenhof, Dassault Systèmes BIOVIA. Based on the article "AB-Bind: Antibody binding mutational database for computational affinity predictions" by Sarah Sirin ***et al***. ***Protein Science*** 2016, 25, 393-409.

INTRODUCTION

Antibodies now have a critical role in both therapeutics and diagnostics areas of medicine and research. The market is expected to grow at a CAGR of 8% or more for several years and reach 90B USD by 2017 (Figure 1). There are many challenges to bringing a new active pharmaceutical ingredient to market. Of these challenges the need to reduce total cost of development and decrease the time to market are critical. Combining ***in silico*** predictions with experiment is widely seen as presenting opportunities to achieve these objectives.

In this Case Study we report on an independent study[2] that compares the relative ΔΔG (binding) prediction capability of BIOVIA Discovery Studio with experiment. The antibody library used is particularly interesting because it includes a high percentage of non-alanine mutations, and a high percentage of multiple simultaneous mutations. The dataset focuses on antibody-antigen interactions, but interactions with Fc domains and nanobodies are also included.

Highlights:

- Antibody database composition
  - 32 complexes including mostly high resolution x-ray crystal structures (27) but also including several proteins constructed with homology modeling (5). Experimental determination of binding employed one or several of the following methods.
- The ΔΔG predictions were obtained with unmodified academic and commercial software as noted.
- Results were compared with experimentally determined changes in binding energy obtained from multiple methods.

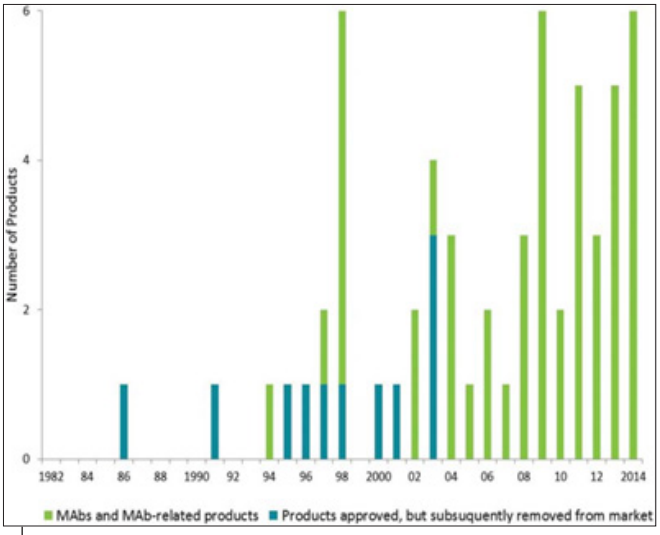

Figure 1: The number of monoclonal antibody products first approved for commercial sale in the US or Europe each year since 1982 is shown. D.M. Ecker, et al.,MAbs. 2015 Jan-Feb; 7(1): 9–14., "The therapeutic monoclonal antibody market."

METHOD

1.Composition of the antibody database (AB_DATABASE)

- Antibody complexes consisted of antibodies interacting with protein antigens. The majority of the structures were from high resolution x-ray structures. All variants (mutations) present were represented by one, but frequently multiple experimental antibody-antigen binding energies.

  i) Basis set includes 32 complexes each with from 7 to 246 variants for a total of 1,001 variants (Fab, Fc, antibody-like and nanobodies)

  ii) Categorized by interaction type (polar, non-polar, aromatic, non-aromatic, charged, neutral)

- Listing of pdb codes for AB database.

  i) **Antigen-antibody pairs** (PDB ID)1MHP, 1MLC, 1VFB, 3HFM, 1DQJ, 1BJ1, 1CZ8, 3BDY, 3BE1, 1N8Z, 3BN9, HM_3BN9, 2NY7, 3NGB, 2NYY, HM_2NYY, 2NZ9, HM_2NZ9, 2JEL, 3NPS, hu225:HM_1YY9, Epithin:HM_3BN9, CR1:HM_2NYY, AR2:HM_2NZ9

ii) **Effector-Antibody pairs** (PDB ID) 1T83, 3WJJ, 1JRH

iii) **Monobody-Antigen** (PDB ID) 3K2M

iv) **Protein1-Protein2** (PDB ID) TGFbeta3:HM_1KTZ, 1KTZ, 1AK4

v) **Effector-Effector** (PDB ID) 1DVF
- Note: homology model names are written as <name>:HM_pdbID where HM=homology model.

32 base antibodies with between 7 and 246 variants present for each.

- Each variant consisted of between 1 and 16 simultaneous mutations.
- 635 single-point mutations, of which 403 were alanine. 466 variants consisting of multiple mutations for which 242 included at least one alanine. Of these 92 contained multiple alanine mutations.
- 119 variants that consisted of multiple mutations were composed of single-point mutations with known (individual) ΔΔG values.
- The location of the mutations relative to the binding interface was determined and results were reported as a total and for each category.

2. Experimental Measurements of change in binding energy

- Multiple methods were used to calculate the binding energy (ΔΔG) for each variant in this study: Surface Plasmon Resonance (SPR), AlphaScreen, ELISA, Kinetic Exclusion Assay (KinExA), Phage-ELISA, Yeast surface display analyzed with flow Cytometry, two different enzymatic methods[1] with the change in binding energy over all mutations ranging from +8 kcal.mol to -3 kcal/mol.

3.Calculation of change in binding

- The change in binding energy for each variant was calculated with multiple methods: baSA-buried accessible surface area, dDFIRE , DFIRE , STATIUM , Rosetta , FoldX , and BIOVIA Discovery Studio[7,8] .
- The results were compared to experimental using AUC (Area under Curve) from Radio Operator Characteristic (ROC) curves.

RESULTS

The calculated values are not reported in the study or supplements. However, the binding data was categorized into four groups based on experimental change in binding energy: | exp ΔΔG| > 0 (all data), | exp ΔΔG| < 0.5 (smallest change), | exp ΔΔG| > 0.5, and | exp ΔΔG| > 1.0 (largest change). In the paper, ROC curves were primarily used to determine the ability of the calculation methods to predict if a variant would be weaker or stronger binding than the parent compound. All prediction methods performed better on average for larger absolute value experimental ΔΔG values. The Pearson correlation coefficients (R-values) are also instructive to compare the different computational methods in this study. Although the actual computed ΔΔG values are not reported, supplement 1 (pro2829-sup-0001-suppinfo01) does contain the overall R values. These values are shown in Table 2. The top two correlation coefficients are 0.34 for FoldX and 0.45 for BIOVIA Discovery Studio.

A variety of subdivisions of the dataset were analyzed to understand the dependence of the observed performance for groupings such as amino acid type, experimental method used, crystal resolution, proximity of the mutation(s) to the center of the complex interface, among others.

This paper reports that most methods performed better when the mutation was in the "Core" (defined as > 25% exposed unbound and <25 % exposed bound) of the binding region, relative to mutations at the "Rim" (<25% ΔASA exposed bound). For example, focusing on the ability to correctly predict Stronger versus Weaker binding for a variant expressed as a 95% confidence interval for Core mutations versus Rim mutations yields the following. For BIOVIA Discovery Studio, the Core mutation confidence level was 0.84 →1.00, and for the Rim variants it was 0.49 → 0.87. For the next best performing method, FOLDX, the Core confidence level was 0.76-0.96 and the Rim confidence was 0.43 → 0.79. Overall, BIOVIA Discovery Studio demonstrated better than a 10-fold enrichment in the top 1% of binders from this data set (11.4 %).

Method	All data	| exp ΔΔG| < 0.5	| exp ΔΔG| > 0.5	| exp ΔΔG| > 1.0
bASA	0.63	nr	0.67	0.68
dDFIRE	0.59	0.58	0.62	0.66
DFIRE	0.65	0.60	0.73	0.78
Discovery Studio	0.73	0.64	0.82	0.88
FoldX	0.70	0.66	0.81	0.87
Rosetta	0.61	0.53	0.65	0.70
STATIUM	0.64	0.60	0.74	0.81

Table 1: Values shown are the AUC for ROC plots for each method showing performance in classifying mutations as either improved vs. weakened binders.

bASA	dDFIRE	DFIRE	STATIUM	Rosetta	FoldX	Disovery Studio
0.22	0.19	0.31	0.32	0.16	0.34	0.45

Table 2: Pearson correlation coefficients for all ΔΔG (computed vs. experiment) improved versus weakened relative to the parent complex.

DISCUSSION

There is a broad range of methods that in principle could be used to predict ***in silico*** binding energies, or solvation stability changes due to mutation. This study only evaluated methods that could scan moderate to large numbers of mutations in a time frame that was deemed to be acceptable to current pharmaceutical discovery standards. For this reason very compute-intensive methods such as FEP and SMD were not considered. There have been several other comparisons of binding energy predictions due to mutations. This study, however, is notable for the concentration of non-alanine mutations (55%) and a large number of multiple simultaneous mutations, rather than containing only single-point mutations. Due to synergies resulting from multiple simultaneous mutations relative to the energy change obtained from point mutations, it is frequently critical to be able to predict changes due to multiple mutations. When predicting ΔΔG(mutation) or ΔΔG(stability) it is common to assume that the mutation does not affect more than the local environment of the antibody. This is due to the vast difference in computation resources that would be required to identify such changes. Similarly, the potential influence of solvent composition including ions is generally not considered. This study makes the further assumption that the ΔΔG(mutation) is not affected by pH (for example pH differences between the experimental conditions at which the ΔΔG was measured and that for which the crystal was isolated).

In this study of the 1,102 mutants listed in supplement 2,494 variants included the mutation of at least one ionizable residue. In these variants a total of 992 individual residues are either mutated to or from an ionizable residue (ERDK). This opens the possibility that the ionization state of these residues and proximate residues might change in the bound complex.

Unlike most commercial methods, BIOVIA Discovery Studio ΔΔG(mutation) and ΔΔG(stability) methods include the ability to determine the pK shifts of the input protein residues on-the-fly and from this predict the changes of binding (or stability) as a function of pH and changing local environment (ΔΔG vs pH). We have shown that binding and protein stability predictions are improved by explicit consideration of the ionizable residues pK, the pH of the system and the environmental changes that occur in the bound versus free environments. We expect that methods such as those implemented in BIOVIA Discovery Studio that scan protonation states would have further improved the industry-leading predictions of BIOVIA Discovery Studio in this study. In addition, the option of including pH-dependence in our method allows monitoring of the change in net charge and prediction of pH regions with minimum solubility for each mutant. The method also extends applicability to new areas such as improving antibody half-life, tuning binding in cellular compartments or tissues with different pH than serum (e.g., endosomes or cancer tissues) or to guide the design of "pH switching" of stability or binding affinity (e.g., for the purposes of biotechnology).

Figure 2: Alpha1-beta1 (VLA-1) integrin with (1MHP) showing interactions with AQC2 Fab

CONCLUSIONS

The predicted change in affinity for a protein complex was tested versus experiment for over 1,000 mutation variants of 32 bound complexes. Many of these variants were the result of multiple simultaneous mutations and involved a diverse range of amino acid residue types (not just single-point mutations or mutations to alanine). The results show that the predictions made by BIOVIA Discovery Studio are sufficient to be able to direct development towards mutations that enrich binding affinity. This coupled ability to explore the change in energy with mutation can be combined with other capability in BIOVIA Discovery Studio to predict and then reduce the risks associated with high viscosity, or aggregation, or decreasing folding stability. BIOVIA Discovery Studio provides a unique solution to reduce development and formulation time, enabling organizations to bring biotherapeutic or diagnostic products to market faster.

REFERENCES:

1. **S. Sirin, J.R. Apgar, E. M. Bennett,** and **A.E. Keating**; Protein Sci. **2016**, 25:393-409 AB-Bind: Antibody binding mutational database for computational affinity predictions
2. **Zhou H, Zhou Y (2002)** Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 11:2714–2726.
3. **Yang Y, Zhou Y (2008)** Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins Struct Funct Bioinf 72:793–803.
4. **DeBartolo J, Taipale M, Keating AE (2014)** Genome-wide prediction and validation of peptides that bind human prosurvival Bcl-2 proteins. PLoS Comput Biol 10: e1003693.
5. **Wedemeyer WJ, Baker D (2003)** Efficient minimization of angle-dependent potentials for polypeptides in internal coordinates. Proteins Struct Funct Bioinf 53:262–272.
6. **Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005)** The FoldX web server: an online force field. Nucleic Acids Res 33:W382–W388
7. Discovery Studio Modeling Environment, Release 4.0. San Diego: Accelrys Software Inc., 2013 [Now Dassault Systèmes BIOVIA]
8. **Spassov VZ, Yan L (2013)** pH-selective mutagenesis of protein–protein interfaces: In silico design of therapeutic antibodies with prolonged half-life. Proteins Struct Funct Bioinf 81:704–714.
9. **V. Spassov and L. Yan**, Proteins **2013**; 81:704–714., "pH-Selective mutagenesis of protein–protein interfaces: In silico design of therapeutic antibodies with prolonged half-life"
10. **V. Spassov. and L. Yan,** A pH-dependent Approach to the Effect of Mutations on Protein Stability and Computational Tool for Combinatorial Scanning Mutagenesis (submitted)

Our **3D**EXPERIENCE® platform powers our brand applications, serving 12 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3DEXPERIENCE®** Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 190,000 customers of all sizes in all industries in more than 140 countries. For more information, visit **www.3ds.com**.

3DEXPERIENCE®

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