JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

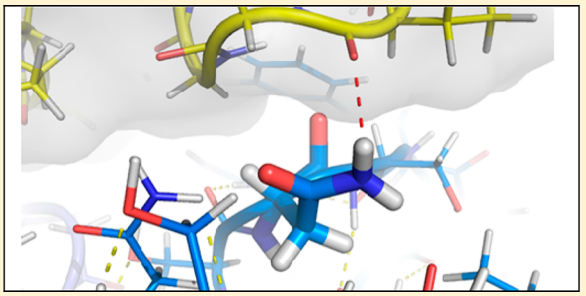
Engineering Protein Therapeutics: Predictive Performances of a Structure-Based Virtual Affinity Maturation Protocol

Michael Oberlin,[†] Romano Kroemer,[†] Vincent Mikol,[†] Hervé Minoux,[†] Erdogan Tastan,[†] and Nicolas Baurin*,[†]

[†]SANOFI R&D, Centre de Recherche de Vitry/Alfortville, LGCR/SDI, 13 quai Jules Guesde - BP 14 - 94403 Vitry-sur-Seine Cedex, France

Supporting Information

ABSTRACT: The implementation of a structure based virtual affinity maturation protocol and evaluation of its predictivity are presented. The *in silico* protocol is based on conformational sampling of the interface residues (using the Dead End Elimination/A* algorithm), followed by the estimation of the change of free energy of binding due to a point mutation, applying MM/PBSA calculations. Several implementations of the protocol have been evaluated for 173 mutations in 7 different protein complexes for which experimental data were available: the use of the Boltzamnn averaged predictor based on the free energy of binding ($\Delta\Delta G_*$) combined with the one based on its polar component only ($\Delta\Delta E_{pol*}$) led to the proposal of a subset of mutations out of which 45% would have successfully enhanced the binding. When focusing on those mutations that are less likely to be introduced by natural in vivo maturation methods (99 mutations with at least two base changes in the codon), the success rate is increased to 63%. In another evaluation, focusing on 56 alanine scanning mutations, the *in silico* protocol was able to detect 89% of the hot-spots.



INTRODUCTION

With more than thirty approvals for therapeutic use in the United States and Europe,[1,2] five for the year 2011 alone,[3] and over 200 mAb candidates in the clinical pipeline,[4] protein therapeutics are the object of increasing interest in pharmaceutical research. Currently the most prominent biotherapeutics are antibodies or antibody-like molecules. Such molecules – once suitably engineered – possess undeniable qualities: excellent pharmacokinetics, up to subnanomolar affinity, and low immunogenicity. Monoclonal antibodies have been produced by immunization of mice, construction of hybridomas, and selection of single clones expressing the desired antibody. More recently, directed evolution techniques such as phage display and related in vitro library display methods have been commonly used.[5] But before they can become clinical candidates, one needs to optimize some of the antibodies' properties in order to maximize their chances of success during clinical development. In the field of protein engineering, such an optimization consists in finding the right mutations to improve the property of interest. However, the preparation and characterization of all the possible mutants would imply high costs and have a significant impact on timelines, due to the combinatorial nature of the task. Therefore computational protein engineering is often used to steer in vitro experiments or at least to limit the number of variants to be generated and tested subsequently in vitro or in vivo.

The study presented in this paper focuses on improving the affinity of an antibody to its antigen. There are several reasons for trying to improve the affinity of an antibody. Increasing the affinity might lead to an antibody not only with improved in vitro but also with improved in vivo biological activity.[6] Also, an improved affinity may translate into a less frequent dose regimen for the patient and consequently a cheaper cost of good. Only a reduced set of amino acids needs to be considered when engineering an antibody with the aim of improving the affinity to its antigen. The Fv fragment, a beta-sheet secondary structure, bears the 6 hypervariable loops (Complementary Determining Regions [CDR[7]], about 60 out of 230 Fv residues) where variation in sequence and lengths allows the immune system to sample an almost infinite diversity in order to generate an antibody interface with a high-affinity to a specific antigen. Framework positions that are not in direct contact with the antigen are also known to play a role in the antibody–antigen interaction.[8] Even though the number of amino acids on the antibody is reduced, the total number of solutions to consider is almost infinite ($\sim 20^{60}$, $>10^{78}$). In order to narrow down the total number of solutions, a 3D model of the antibody–antigen is therefore mandatory. Protein–protein docking can be envisaged, but it remains an extremely challenging task to just accurately identify the antibody amino acids involved in the interaction.[9] In any case, the availability of the 3D atomic coordinates of the complex can be considered crucial for predicting affinity enhancing mutations.

Received: March 16, 2012
Published: July 12, 2012

dx.doi.org/10.1021/ci3001474 | J. Chem. Inf. Model. 2012, 52, 2204–2214

Fortunately these coordinates can often be obtained swiftly by crystallography.[10]

Having established the prerequisites, the *in silico* affinity maturation is essentially a two-step process. The first step is to generate each mutation at the interface. It requires a conformational sampling method, and the most efficient one - if it converges - was reported to be the DEE/A* protocol[11−14] (Dead End Elimination algorithm, followed by the A* algorithm). The second step is the calculation of the free energy of binding between the mutated antibody and the antigen. Usually, a large number of conformations have to be evaluated, this is why MM/PBSA[15] (molecular mechanics Poisson−Boltzmann surface area) is widely chosen as a good compromise between desired accuracy and time.

The aforementioned approach has been already tested by Lippow on antilysosyme antibodies, where the exclusive use of the polar component of the free energy of binding to predict favorable mutations was evaluated as well.[16] Another study published by Clark[17] reported a comparison of 4 methods (electrostatic optimization, side chain repacking, flexible backbone, and DEE/A* +MMPBSA) of *in silico* affinity maturation, but the study covered only one system, namely antiVLA1.

The high variability of properties when going from one system to another makes it difficult to get a general idea of the performance of a method if the data set is not large enough. Therefore we created our own, larger, data set containing several systems with published experimental data. Subsequently we developed our own implementation of the DEE/A*-MM/PBSA protocol and tested it on this data set. This data set could be used as an external benchmark as is recurrent in other areas such as the field of high throughput docking of small molecules.[18,19]

Furthermore, we evaluated two variations of the protocol: the first was the use of the polar part of the free energy of binding alone, in order to see whether it is sufficiently predictive of the effect of a point mutation on affinity. This alternative was suggested by Lippow,[16] and it was reported as giving better results on the tested antilysosymes than using the total free energy of binding. The second protocol variation that was evaluated in the present study concerns the inclusion of all energies calculated for multiple conformations of the residue under investigation, via a Boltzmann-weighted average.[20]

To complete this more global evaluation of the computational affinity maturation protocol, we investigated the performance of this protocol on two types of mutation subsets. The first subset contains mutations with a lower mutation probability. The more DNA base changes a mutation requires, the lower is the probability for this mutation to have been already explored previously by experimental methods, such as phage-display or *in vivo* maturation.[21] Therefore this subset included only those mutations. The second subset concerns the systematic mutation of each residue of the interface into alanine, in order to identify the so-called interaction hot-spots. We evaluated the predictivity of this approach, also referred to as computational alanine scanning, which could be used instead of *in vitro* alanine scanning to steer *in vitro* maturation accordingly, focusing in particular on those amino acids for which the affinity could be more easily improved.[22,23]

RESULTS

The *in silico* protocol described in the Materials and Methods section was applied to 7 protein systems: an anti-VLA1 antibody, the antilysozymes D44.1, D1.3, and HyHEL-63 antibodies, an anti-EGFR antibody (cetuximab), an anti-HER2 (trastuzumab) variant bH1, and the Barnase-Barstar complex. Experimentally, the effect of a mutation can be considered as positive when the measured impact on binding is favorable ($\Delta\Delta G_{binding} < 0$ kcal/mol). Conversely, the effect of a mutation can be considered as negative when the measured impact on binding is unfavorable ($\Delta\Delta G_{binding} > 0$ kcal/mol). In the present study two criteria were applied for the classification of mutations using computed values: a mutation was predicted positive if both the calculated impact on binding ($\Delta\Delta G_{binding} < cutoff$) and antibody stability were favorable. For details the reader is referred to the Materials and Methods section. The initial *cutoff* used to predict positive mutations was 0 kcal/mol. Mutations for which no conformation was found by DEE/A* sampling or for which the sampling algorithm did not converge were considered as negative.

General Performances. The results for the 6 variations of the computational protocol are summarized in Table1. For this evaluation, 173 mutations of 7 protein complexes with available experimental data have been used. The 6 variations differ in the way the free energy of binding was calculated, which is subsequently used to predict if a mutation will have a beneficial or detrimental impact on binding. For this reason, the 6 energies will be referred to as "predictors" in the remainder of the article. A detailed listing of all predictions for all systems and all predictors is given in the Supporting Information.

For the systems studied by Lippow[16] (D44.1, D1.3, and cetuximab), 36 mutations are reported, of which 14 were shown experimentally to improve the binding between the antibody and the antigen (positive mutations). Using $\Delta\Delta G_*$ as a predictor (*calculated* $\Delta\Delta G_{binding}$, averaged as described in eq 4), 10 true positives (*TP*) and 6 false positives (*FP*) were found. The sensitivity and specificity values for this predictor were 0.71 and 0.73, respectively. The best performances were obtained by the *top* $\Delta\Delta E_{pol}$ predictor, with a success rate of 73% . The success rates (*SR*) for other predictors are as follows $SR_{\Delta\Delta G*}$ = 63%, $SR_{\Delta\Delta G**}$ = 60%, $SR_{\Delta\Delta Epol*}$ = 69%, $SR_{\Delta\Delta Gtop}$ = 64%, and $SR_{\Delta\Delta G* \& \Delta\Delta Epol*}$ = 71%.

For the anti-VLA1 system, 67 mutations were available,[17] of which 10 were found to be positive. 70% of all positive mutations were found with the $\Delta\Delta G_*$ predictor (sensitivity = 0.7), and the specificity reached 0.67. The success rate of the double criteria predictor ($\Delta\Delta G_*$ & $\Delta\Delta E_{pol}$) was improved when compared to the $\Delta\Delta G_*$ *predictor* ($SR_{\Delta\Delta G* \& \Delta\Delta Epol*}$ = 35% and $SR_{\Delta\Delta G*}$ = 27% respectively). This is due to a significantly lower number of false positives (13 and 19, respectively).

For HHEL-63 and HER2 complexes, only mutations to Ala were available (alanine scanning). The $\Delta\Delta G_*$ predictor identifies 7 of the 11 positive mutations, *i.e.* hot-spots of interaction, and 36 out of of the 45 negatives (Sensitivity = 0.64 and specificity = 0.8).

The Barnase-Barstar is a system that is well-known for its optimized interface. As a consequence, the mutations found in the literature are mainly unfavorable (1 positive and 13 negative mutations). The $\Delta\Delta G_*$ predictor has a specificity of 0.69, and the positive mutation was predicted by none of the predictors.

The global performance on all 7 systems is as follows ("*All*" part of the Table 1): The sensitivity varies in the [0.56−0.69] range, while the specificity varies in the [0.72−0.79] range. More importantly, the success rate (% of positive mutations in the selected subset) varies between 37% and 45%. This corresponds to a 2-fold enrichment when compared to the 21% that would be obtained from a random selection (cf. the random picking column in Table 1). The best performances were obtained with the double criteria predictor

Table 1. Evaluation of 6 Predictors Regarding the Impact of Mutations on the $\Delta\Delta G_{binding}$[a]

	system	exp. (threshold =0 kcal/mol)	$\Delta\Delta G_{*}$	$\Delta\Delta G_{**}$	$\Delta\Delta E_{pol*}$	*top* $\Delta\Delta G$	*top* $\Delta\Delta E_{pol}$	$\Delta\Delta G$* & $\Delta\Delta E_{pol*}$	random picking
TP	D44.1	6	5	4	5	3	4	5	
FP			4	5	5	3	4	4	
TN		5	1	0	0	2	1	1	
FN			1	2	1	3	2	1	
sensitivity			**0.83**	**0.67**	**0.83**	**0.50**	**0.67**	**0.83**	
specificity			**0.20**	**0.00**	**0.00**	**0.40**	**0.20**	**0.20**	
success rate			**0.56**	**0.44**	**0.50**	**0.50**	**0.50**	**0.56**	**0.55**
TP	D1.3	3	1	1	2	1	2	1	
FP			2	1	0	1	0	0	
TN		17	15	16	17	16	17	17	
FN			2	2	1	2	1	2	
sensitivity			**0.33**	**0.33**	**0.67**	**0.33**	**0.67**	**0.33**	
specificity			**0.88**	**0.94**	**1.00**	**0.94**	**1.00**	**1.00**	
success rate			**0.33**	**0.50**	**1.00**	**0.50**	**1.00**	**1.00**	**0.15**
TP	Cetuximab	5	4	4	4	3	5	4	
FP			0	0	0	0	0	0	
TN		0	0	0	0	0	0	0	
FN			1	1	1	2	0	1	
sensitivity			**0.80**	**0.80**	**0.80**	**0.60**	**1.00**	**0.80**	
specificity			**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	
success rate			**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
TP	anti-VLA1	10	7	7	7	8	7	7	
FP			19	17	15	18	17	13	
TN		57	38	40	42	39	40	44	
FN			3	3	3	2	3	3	
sensitivity			**0.70**	**0.70**	**0.70**	**0.80**	**0.70**	**0.70**	
specificity			**0.67**	**0.70**	**0.74**	**0.68**	**0.70**	**0.77**	
success rate			**0.27**	**0.29**	**0.32**	**0.31**	**0.29**	**0.35**	**0.15**
TP	HyHEL-63	0	0	0	0	0	0	0	
FP			1	1	2	1	3	1	
TN		21	20	20	19	20	18	20	
FN			0	0	0	0	0	0	
sensitivity			0	0	0	0	0	0	
specificity			**0.95**	**0.95**	**0.90**	**0.95**	**0.86**	**0.95**	
success rate			**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
TP	anti-HER2	11	7	7	7	5	5	7	
FP			8	7	11	6	10	7	
TN		24	16	17	13	18	14	17	
FN			4	4	4	6	6	4	
sensitivity			**0.64**	**0.64**	**0.64**	**0.45**	**0.45**	**0.64**	
specificity			**0.67**	**0.71**	**0.54**	**0.75**	**0.58**	**0.71**	
success rate			**0.47**	**0.50**	**0.39**	**0.45**	**0.33**	**0.50**	0.31
TP	Barnase-Barstar	1	0	0	0	0	0	0	
FP			4	4	5	4	5	4	
TN		13	9	9	8	9	8	9	
FN			1	1	1	1	1	1	
sensitivity			**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	
specificity			**0.69**	**0.69**	**0.62**	**0.69**	**0.62**	**0.69**	
success rate			**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.07**
TP	*All*	36	24	23	25	20	23	24	
FP			38	35	38	33	39	29	
TN		137	99	102	99	104	98	108	
FN			12	13	11	16	13	12	
sensitivity			**0.67**	**0.64**	**0.69**	**0.56**	**0.64**	**0.67**	
specificity			**0.72**	**0.74**	**0.72**	**0.76**	**0.72**	**0.79**	
success rate			**0.39**	**0.40**	**0.40**	**0.38**	**0.37**	**0.45**	**0.21**

[a]The evaluation was performed for each protein system separately, as well as for all the mutations across all the systems (*All*). The predictors flagged by an * and ** have been Boltzmann-averaged according to eq 4 and eq 3 respectively. top $\Delta\Delta E_{pol}$ and *top* $\Delta\Delta G$ are nonaveraged predictors based on the best conformation. *TP*: true positives, *FP*: false positives, *TN*: true negatives, *FN*: false negatives. The random picking success rate is the expected success rate of a random selection to predict *TP* out of *TP*+*TN*. See the Material andn Methods section for more details. Numbers that are not in bold refer to the number of *TP, FP, TN,* or *FN*.

dx.doi.org/10.1021/ci3001474 | *J. Chem. Inf. Model.* 2012, 52, 2204−2214

($\Delta\Delta G_* < 0$ *kcal/mol* & $\Delta\Delta E_{pol*} < 0$ *kcal/mol*) with a sensitivity of 0.67 and a success rate of 45%.

What Is the Impact of the Predicted $\Delta\Delta G_{binding}$ *Threshold on Predictive Performances?* To address this question, we compared different thresholds governing the decision that a mutation is predicted positive. Several cutoff values (−1, −0.5, 0, 0.5, and 1 kcal/mol) were tested for the $\Delta\Delta G_*$ and $\Delta\Delta E_{pol*}$ predictors (Table 2). As expected, a lower cutoff led to a lower

Table 2. Evaluation of $\Delta\Delta G_*$ and $\Delta\Delta E_{pol}$ Predictors at Different Cutoff Values[a]

	cutoff (kcal/mol)	exp. (threshold = 0 kcal/mol)	$\Delta\Delta G_*$	$\Delta\Delta E_{pol*}$	$\Delta\Delta G_*$ & $\Delta\Delta E_{pol*}$
TP	-1	36	15	15	12
FP			21	25	18
TN		137	116	112	119
FN			21	21	24
sensitivity			**0.42**	**0.42**	**0.33**
specificity			**0.85**	**0.82**	**0.87**
success rate			**0.42**	**0.38**	**0.40**
TP	-0.5	36	17	20	16
FP			29	29	22
TN		137	108	108	115
FN			19	16	20
sensitivity			**0.47**	**0.56**	**0.44**
specificity			**0.79**	**0.79**	**0.84**
success rate			**0.37**	**0.41**	**0.42**
TP	0	36	24	25	24
FP			38	38	29
TN		137	99	99	108
FN			12	11	12
sensitivity			**0.67**	**0.69**	**0.67**
specificity			**0.72**	**0.72**	**0.79**
success rate			**0.39**	**0.40**	**0.45**
TP	0.5	36	26	26	25
FP			43	47	38
TN		137	94	90	99
FN			10	10	11
sensitivity			**0.72**	**0.72**	**0.69**
specificity			**0.69**	**0.66**	**0.72**
success rate			**0.38**	**0.36**	**0.40**
TP	1	36	26	29	26
FP			44	54	40
TN		137	93	83	97
FN			10	7	10
sensitivity			**0.72**	**0.81**	**0.72**
specificity			**0.68**	**0.61**	**0.71**
success rate			**0.37**	**0.35**	**0.39**

[a]The whole data set was taken into account. Cutoff refers to the threshold used to select positive mutations by an *in silico* predictor. The cutoff values listed apply to the results shown in the respective cells. Definitions of $\Delta\Delta G_*$ and $\Delta\Delta E_{pol*}$ are described in the caption of Table 1. Numbers that are not in bold refer to the number of *TP*, *FP*, *TN*, or *FN*. See the Material and Methods section for more details.

Sensitivity (*i.e.* less positives detected) and a higher specificity (*i.e.* more negatives found). The effect of the cutoff on the performances of the double criterion ($\Delta\Delta G_*$ & $\Delta\Delta E_{pol*}$) is illustrated in Figure 1. Interestingly, the success rates (*SR*) remain stable for each predictor across the different cutoff values.

What Is the Performance of the Predictors on the Mutations Less Likely To Be Sampled by Directed Evolution or Experimental Methods? To address this question, a more focused data set was compiled by selecting only those mutations with the property $\Delta codon$ above 1. $\Delta codon$ is the minimal number of base changes at the DNA level required to generate a given mutation from the wild-type. Also, in order to focus the analysis on antibody engineering applications, the Barnase-Barstar system was not taken into account. As shown in Table 3,

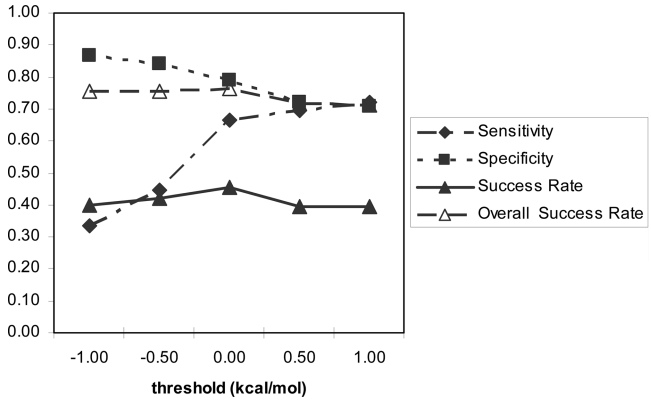


Figure 1. Effect of the cutoff on various parameters when using the double criterion ($\Delta\Delta G_*$ & $\Delta\Delta E_{pol*} < cutoff$) predictor. The success rate is the ratio of *TP* among the predicted positives, whereas the overall success rate is the ratio of true predictions and all predictors, *i.e.* $(TP+TN)/(TP+TN+FP+FN)$.

Table 3. Comparison of 3 Predictors for the Whole Data Set or the "$\Delta Codon > 1$" Focused Data Set[a]

	data set	exp.	$\Delta\Delta G_*$	$\Delta\Delta E_{pol*}$	$\Delta\Delta G_*$ & $\Delta\Delta E_{pol*}$
TP	large	35	24	25	24
FP			34	33	25
TN		124	90	91	99
FN			11	10	11
sensitivity			**0.69**	**0.71**	**0.69**
specificity			**0.73**	**0.73**	**0.80**
success rate			0.41	0.43	0.49
TP	focused	22	15	15	15
FP			16	14	9
TN		77	61	63	68
FN			7	7	7
sensitivity			**0.68**	**0.68**	**0.68**
specificity			**0.79**	**0.82**	**0.88**
sensitivity°			**0.43**	**0.43**	**0.43**
specificity°			**0.49**	**0.51**	**0.55**
success rate			**0.48**	**0.52**	**0.63**

[a]The Barnase-Barstar complex was not taken into account. $\Delta\Delta G_*$ and $\Delta\Delta E_{pol*}$ are defined as outlined previously. The criteria to predict positive mutations on the remaining data set were $\Delta\Delta G_* < 0$, $\Delta\Delta E_{pol*} < 0$, or ($\Delta\Delta G_* < 0$ & $\Delta\Delta E_{pol*} < 0$).

the predictive performances were significantly better on this subset. 63% of predicted positives (using the double criterion $\Delta\Delta G_* < 0$ & $\Delta\Delta E_{pol*} < 0$) are indeed affinity enhancing mutations, as compared to 49% for the whole data set.

Computational Alanine Scanning. To evaluate the performances of the predictors in predicting hot-spots by computational alanine scanning, the data set was reduced to the 2 systems (anti-HER2 and HyHEL-63) for which only alanine mutations were available (Table 4). To be consistent with the principle of hot-spot detection, where the hot-spot is defined

dx.doi.org/10.1021/ci3001474 | J. Chem. Inf. Model. 2012, 52, 2204−2214

Table 4. Results for Computational Alanine Scanning Applied to HER2 and HyHEL-63[a]

	cutoff (kcal/mol)	exp. (threshold =1 kcal/mol)	$\Delta\Delta G^*$	$\Delta\Delta E_{pol*}$	$\Delta\Delta G_*$ & $\Delta\Delta E_{pol*}$
TP	-1	28	2	4	2
FP			2	3	2
TN		28	26	25	26
FN			26	24	26
sensitivity			**0.07**	**0.14**	**0.07**
specificity			**0.93**	**0.89**	**0.93**
success rate			**0.50**	**0.51**	**0.50**
TP	-0.5	28	6	7	4
FP			2	3	2
TN		28	26	25	26
FN			22	21	24
sensitivity			**0.21**	**0.25**	**0.14**
specificity			**0.93**	**0.89**	**0.93**
success rate			**0.54**	**0.54**	**0.52**
TP	0	28	13	15	12
FP			3	5	3
TN		28	25	23	15
FN			13	13	16
sensitivity			**0.50**	**0.54**	**0.43**
specificity			**0.89**	**0.82**	**0.83**
success rate			**0.66**	**0.64**	**0.48**
TP	0.5	28	16	19	16
FP			3	6	3
TN		28	25	22	25
FN			12	9	12
sensitivity			**0.57**	**0.68**	**0.57**
specificity			**0.89**	**0.79**	**0.89**
success rate			**0.68**	**0.71**	**0.68**
TP	1	28	17	22	17
FP			3	6	3
TN		28	25	22	15
FN			11	6	11
sensitivity			**0.61**	**0.79**	**0.61**
specificity			**0.89**	**0.79**	**0.83**
success rate			**0.69**	**0.79**	**0.58**

[a] $\Delta\Delta G_*$ and $\Delta\Delta E_{pol*}$ are defined as outlined previously. The success rate applies to the accurate prediction of negatives: $SR = TN/(TN+FN)$.

as an amino acid with a critical contribution to the overall binding, negative mutations were defined as mutations with an experimental $\Delta\Delta G_{binding}$ of at least 1 kcal/mol. The sensitivity and specificity were calculated as usual, but the success rate does not apply anymore to the positives and was calculated as follows: $SR = TN/(TN+FN)$. In Table 5, a comparison of the hot-spots predicted *in silico* versus the experimental hot-spots is shown.

DISCUSSION

General Performances. The $\Delta\Delta G_*$ can be considered as the most relevant predictor since it combines all energetic contributions (polar and apolar) and is a Boltzmann average over the best conformations from the sampling procedure. Taking into account the 173 mutations of the 7 systems, the sensitivity of the basic criterion $\Delta\Delta G_{top} < 0$ *kcal/mol* (including polar and apolar contributions of the lowest energy conformation) is 56%, and its success rate is 38% (Table 1). Sensitivity and success rate are complementary. Sensitivity measures the ratio of all positives, *i.e.* mutations leading to an improved binding, detected by the predictor. Success rate measures the proportion of positives in the selection made based on the predictor. This parameter is of particular importance since it reflects the expected chances of success when proposing particular mutations for further cloning and characterization. By comparison, the success rate corresponding to a random selection is 21%. In the published study of D44.1, D1.3, and cetuximab success rates between 50% and 67% are reported with $\Delta\Delta G_{top}$ or the $\Delta\Delta G_{elec}$ as predictors.[16] In the present study the success rate was 64% using $\Delta\Delta G_{top}$ for these three systems. Clarks et al. reported success rates reaching 48% on antiVLA1,[17] compared to 44% in our study on the same system.

Choice of a Predictor. The $\Delta\Delta G_{top}$ can be considered as a reference to make a prediction since it is based on the lowest energy conformation and it takes into account all the energetic contributions. As reported in the literature,[16] the polar part of the free energy of interaction was identified as the best predictor to propose mutations that enhance the affinity of antibodies. In our study, the comparison between the predictions based on $\Delta\Delta G_{top}$ and on its polar part alone ($\Delta\Delta E_{pol\ top}$) indicates that sensitivity and success rate are on average better with $\Delta\Delta E_{pol\ top} < 0$ *kcal/mol*, when applied on the 3 systems studied by Lippow et al. However, this descriptor is less successful for the other complexes. Despite a better Sensitivity (64% instead of 56%) on the whole data set, the success rate remains on the same level than the predictions based on $\Delta\Delta G_{top}$. This is due to an increased number of false positives produced at the same time than some additional true positives.

The aforementioned predictors are based on the best conformation found by sampling. However, up to 30 conformations are produced for each mutation, and we suggest to average all these energies via a Boltzmann average. This averaging scheme should be more realistic, given that it better reflects the dynamic nature of the recognition where a distribution of conformations over time contributes to binding. The results obtained on the whole data set yield similar success rates for both types of descriptors but a better sensitivity for Boltzmann-averaged predictors: more true positives were found with $\Delta\Delta G_* < 0$ *kcal/mol* compared to $\Delta\Delta G_{top} < 0$ *kcal/mol*

Table 5. Residues of HER2 Identified As Hot-Spots by the *in Vitro* Alanine Scanning (in Black if $\Delta\Delta G_{binding} > 1$ kcal/mol, in Grey if $\Delta\Delta G_{binding} > 0.5$ kcal/mol) or by *in Silico* Alanine Scanning (in Black if $\Delta\Delta G_* > 1$ kcal/mol)[a]

Light Chain	27 Q	28 D	29 I	30 P	30a R	30b S	30c I	30d S	31 G	32 Y	50 W	51 G	52 S	53 Y	91 H	92 Y	93 T	94 T
In-vitro alascan	■						▒				■			▒	▒			■
In-silico alascan	■		■		■		■			■	■			■	■	■		■

Heavy Chain	30 K	31 D	32 T	33 Y	50 R	52 Y	53 T	54 N	56 Y	58 R	95 W	96 G	97 G	98 D	99 G	100 F	100a Y
In-vitro alascan				■	■				■	■	■				■	■	■
In-silico alascan				■	■				■	■	■					■	■

[a] The remaining residues (white cells) are considered optimizable.

dx.doi.org/10.1021/ci3001474 | J. Chem. Inf. Model. 2012, 52, 2204−2214

(sensitivity of 0.67 and 0.56, respectively). Regarding the number of affinity-enhancing mutations, Boltzmann averaging increased both the number of true positives and false positives, resulting in an unchanged success rate. The Boltzmann averaging procedure used here included weighting the energies of the antibodies, antigens, and corresponding complexes all on their own (eq 4), before calculating the differences to estimate the $\Delta G_{binding}$ (eq 4). We also attempted to weight directly the $\Delta G_{binding}$ by the total free energy of the complex (eq 3), but the resulting $\Delta\Delta G_{**}$ predictor did not yield a significantly better *SR*, sensitivity or specificity (Table 1).

As with $\Delta\Delta G_{top}$ and $\Delta\Delta G_{*}$, Boltzmann averaging was applied to the polar part only of the free energy of interaction. The total free energy of the system was used to generate the weighting coefficient of each isolated binding partner or of the complex. The resulting $\Delta\Delta E_{pol*}$ predictor performed better than the $\Delta\Delta E_{pol}$ based on the best conformations. With a success rate of 40% and a Sensitivity of 70%, the prediction yields a similar quality than the $\Delta\Delta G_{*} < 0\ kcal/mol$ criterion.

The predictions based on $\Delta\Delta E_{pol*} < 0\ kcal/mol$ or on $\Delta\Delta G_{*} < 0\ kcal/mol$ showed a big overlap of true positives lists but significant differences regarding the false positives. The combination of both was thought to lead to a better performance on false positives. With this more stringent criterion, a mutation will be predicted as enhancing the affinity if it improves both the total energy and its polar component. The result was a decrease of the number of false positive by 24%, leading to an increase of the success rate (from 40 to 45%) and of the specificity (from 72 to 79%). In summary this double criterion was found to be best predictor for 173 mutations of the 7 systems (Table 1) studied here.

What Is the Impact of the $\Delta\Delta G$ Threshold ? Intuitively, we used the threshold of *0 kcal/mol* for predictors to select good mutations, *i.e.* mutations thought to improve the affinity to the antigen. As can be seen in Figure 1, the *SR* for the double criterion ($\Delta\Delta E_{pol*} < threshold$ & $\Delta\Delta G_{*} < threshold$) does not vary much (from 39 to 45%) if the threshold varies between −1 and +1 kcal/mol. The main difference between different cutoffs is due to the total number of positives identified. One could contemplate using the threshold as a criterion to adjust the final number of proposed mutations to the capacities of the lab where the corresponding experiments are performed. Overall, the 0 kcal/mol threshold seems to be a reasonable compromise to be used by default.

What Are the Performances of the Predictors on the Mutations Less Likely to Be Sampled by Directed Evolution or Experimental Methods? An antibody subject to *in silico* affinity maturation has been produced by either *in vivo* or *in vitro* methods. *In vivo*, a phase of error prone proliferation of B cells provokes somatic hypermutations of the variable part of antibodies gene. This process allows an affinity maturation of immunoglobulin by accumulating variations of the initial antigen specific antibody. The mechanism is complex and not yet fully understood, but the consequence is the transversion or transition of DNA bases.[24,25] Statistically, it leads most frequently to amino acid mutations due to single base changes in their codon.[21] The other approach is the *in vitro* production of antigen targeted antibody, which is based on the expression and evaluation of a library of antigens. This library is normally produced by error prone PCR methods, leading to a preponderant sampling of single codon mutations. Whether somatic mutation occurs in immunization approaches or whether error prone PCR was used to generate a library, the low proportion of mutations based on double or triple codon mutation is a feature that could be exploited further. If we assume that mutations corresponding to single mutations are more likely to have been already optimized, it is of interest to measure the performances of the *in silico* affinity maturation on mutations that are based on multiple codon changes.

We therefore excluded Barnase-Barstar from the data set. It is an *in vivo* system that was not produced by the immune system or *in vitro* error-prone systems. The remaining systems comprised 159 mutations, of which 35 were positives. As shown in Table 3, the additional exclusion of mutations with single codon base changes limited the data set to 99 mutations, of which 22 were positives. The measured success rate on this focused data set, when using the best performing predictor ($\Delta\Delta E_{pol*} < 0\ kcal/mol$ & $\Delta\Delta G_{*} < 0\ kcal/mol$), is 63%. One might think that the slightly better performance for this reduced set is due to a higher proportion of positives, which would make the task of predicting positives easier, but that is not the case. The ratio of positives mutations in the reduced set is identical to the one in the 159 mutations data set, *i.e.* 22%.

Interestingly, all the most enhancing mutations ($\Delta\Delta G_{exp}$ lower than −0.5 kcal/mol) are based on at least two codon base changes. All together, this suggests a pragmatic approach to successfully identify the best mutations, *i.e.* to propose mutations with at least 2 codon base changes ($\Delta Codon > 1$), for which the binding is improved according to the affinity maturation procol ($\Delta\Delta E_{pol*} < 0\ kcal/mol$ & $\Delta\Delta G_{*} < 0\ kcal/mol$).

Computational Alanine Scanning. Experimentally, alanine scanning analysis is performed to identify amino acids that are crucial for binding. A recent study made use of experimentally determined hot-spots to successfully focus affinity maturation[26] on optimizable positions, *i.e.* non hot-spot amino acids. When the crystallographic structure of the complex is available, the alanine scanning analysis can be performed by *in silico* techniques.[22,27,28] We have tested the ability of our protocol to successfully predict the hot-spots of some protein−protein interfaces for which alanine scanning results were available.

We used HyHEL-63 and anti-HER2 systems to evaluate the ability of our protocol to predict hot-spots of interaction by alanine scanning (see Table 4). For alanine scanning on these two systems, the double criterion $\Delta\Delta G_{*}$ & $\Delta\Delta E_{pol*}$ turned out not to be the best choice (only 15 out of 28 hot-spots found with a threshold of 1 kcal/mol). It appears that the addition of $\Delta\Delta E_{pol*}$ to the $\Delta\Delta G_{*}$ predictor decreases the success rate, whereas it increases the success rate when predicting favorable mutations. This is not surprising since an enhancing mutation has to be favorable in both its polar and apolar components ($\Delta\Delta E_{pol} \leq 0\ kcal/mol\ and\ \Delta\Delta E_{apol} \leq 0\ kcal/mol$). For a mutation X→ALA the loss in binding energy can be due to a change in either the polar or apolar energy. In the latter case, adding a polar term to the calculation will at best not change the result and at worst will just add "noise" to the calculation.

The $\Delta\Delta G_{*}$ predictor gave better results, especially with a threshold of 1 kcal/mol: 25 out of 28 hot-spots were found (specificity = 89%), at the cost of 11 additional predictions found to be false negatives. Using $\Delta\Delta E_{pol*} < 1$ kcal/mol, less hot-spots (22 true negatives) were found but only 6 false negatives were predicted. When inspecting the mutations responsible for the seemingly better success rate of $\Delta\Delta E_{pol*}$ (79% vs 69% for $\Delta\Delta G_{*}$), we noted that a majority are large hydrophobic residues. All of the 3 missed True Negatives, but also 3 of the 5 avoided False Negatives are either a

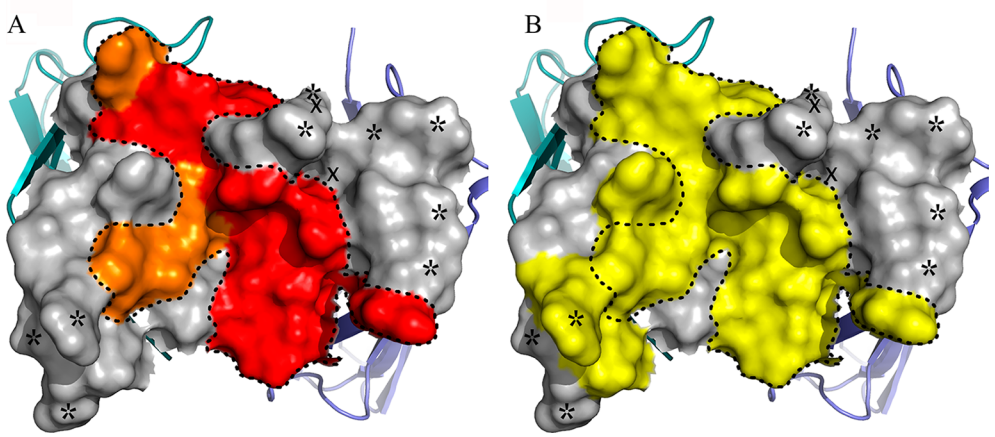


Figure 2. Comparison of (A) in vitro and (B) in silico alanine scaning analysis. A and B pictures are representations of the antibody−antigen interface of anti-HER2 bH1, viewed from the antigen side. In the two cases the light chain is on the left, and the heavy chain is on the right. Residues that were analyzed by alanine scanning are represented as surface. Residues predicted as hot-spots by the *in vitro* study are surrounded by a dashed line in the two representations. Residues selected by *in vitro* analysis to be randomized are symbolized by * and **X** (corresponding to positions where an enhancing mutation was found or not, respectively). **A**: residues found as hot-spot by in vitro alanine scanning were colored in orange (0.5 kcal/mol < $\Delta\Delta G$ <1 kcal/mol) or red ($\Delta\Delta G$ > 1 kcal/mol). **B**: residues predicted as hot-spots by in silico alanine scanning are colored in yellow.

Tryptophane or a Tyrosine. This is not surprising since the interactions made by these large hydrophobic residues can only be described partially by $\Delta\Delta E_{pol*}$. In other words, in this particular case the improvement in the prediction is due to the composition of the data set in terms of hydrophobic versus hydrophilic residues in combination with an energy term that works better for one type of residues than the other type.

Good performances at predicting true positives (*i.e.*, amino acids that could improve the overall binding affinity upon mutation) can be very interesting since such positions of the interface are optimizable. We were interested to compare the predictive performances of our implementation with other computational alanine scanning protocols. Li et al. tested 4 methods on a large set of alanine mutations, and the averaged success rates (calculated as $TN+TP/(TN+TP+FN+FP)$) range from 66 to 72%. Another protocol based on linear scaling quantum mechanical methods was reported[23] to reach 75% of averaged success rate. In our case, when taking into account the 56 mutations of HyHEL-63 and anti-HER2, the averaged success rate is 69% by using $\Delta\Delta G_*$ and reach 79% with $\Delta\Delta E_{pol*}$. Therefore, the protocol that we implemented based on DEE/A* and MM/PB-SA allows to predict hot-spots with a performance as high as other recently published protocols of computational alanine scanning.

As a retrospective exercise, we compared the outcome of an experimental alanine scanning to our *in silico* alanine scanning implementation in the particular context of affinity maturation. Bostrom et al.[26] used experimental alanine scanning to decipher bH1 bispecificity against both HER2 and VEGF. In particular, they deduced the functional paratope from alanine scanning data on 35 positions. Subsequently, by considering that the hot-spots of interaction should be excluded from further maturation, they applied residue randomization as a last round of affinity maturation (Table 5). Nine enhancing mutations were successfully identified. Applying our *in silico* alanine scanning protocol with the $\Delta\Delta G < 1$ *kcal/mol* criterion to those 35 positions, we predicted 17 hot-spots. These 17 predicted hot-spots included 13 of the 14 experimentally determined true hot-spots. Interestingly, the 18 remaining positions considered as optimizable contain 8 of the 9 enhancing mutations (Figure 2). This demonstrates that *in silico* alanine scanning of a 3D high-resolution antibody−antigen complex can be used for focusing the randomization on key positions. We therefore propose that the use of such an *in silico* alanine scanning protocol to steer the *in vitro* affinity maturation is a complementary approach to the design of specific variants containing a limited number of mutations.

CONCLUSIONS AND OUTLOOK

Several parameters of a structure-based virtual affinity maturation protocol have been explored. Beyond the fine-tuning of key parameters for an optimal predictivity, this study shed some light on how such protocols can be used in a variety of ways to steer the affinity maturation of protein therapeutics.

The best predictor to find mutations that enhance the affinity of the antibody to the antigen was found to be the double criteria predictor ($\Delta\Delta E_{pol*} < 0$ *kcal/mol and* $\Delta\Delta G_* < 0$ *kcal/mol*), with a 45% success rate measured on 7 different systems totalling 173 point mutations. The number of mutations proposed by *in silico* affinity maturation can be further reduced with no impact on the success rate in different ways, as follows. For instance, if the initial list of mutations cannot be handled experimentally, lowering slightly the threshold up to −1 kcal/mol is an option. Also, by considering only mutations involving more than 1 codon base change vs the wild-type sequence, one can reduce the list of mutations. More importantly, this criterion can be explicitly considered in the selection strategy since these particular mutations are less likely to have been sampled by the experimental methods. The measured success rate on this particular subset of mutations is 63%.

In the particular case where alanine mutations only are considered, the structure-based virtual affinity maturation protocol presented in this study was highly predictive to identify the hot-spots of interaction (*in silico* alanine scanning). The potential usefulness of such an in silico alanine scanning protocol has been demonstrated by applying it to the trastuzumab-HER2 complex, where *in vitro* randomization would have been focused on 19 optimizable positions out of 35 positions in contact with the antigen.

The two types of evaluation we have performed illustrate how a structure-based virtual affinity maturation protocol can be applied to support two complementary protein design

strategies. The first strategy consists in identifying specific mutations and allows to propose a limited set of mutants for further cloning, expression, and characterization. The second strategy consists in identifying specific positions where experimental randomization can be performed.

The structure-based virtual affinity protocol could also be used to design antibodies with improved pharmacokinetics (PK). J. Chaparro-Riggers et al.[29] have shown that the introduction of histidines at specific CDR positions maintains the affinity to the antigen at physiological pH but alters the affinity at the endosomal acidic pH. The reduced binding in the endosome leads to increased dissociation of the antibody−antigen complex, and consequently the antibody escapes target-mediated degradation *via* the neonatal Fc receptor (FcRn) mediated recycling pathway. Provided a high resolution antibody−antigen structure is available, the *in silico* protocol could be used to identify the key positions where the introduction of histidines should not reduce the affinity to the antigen under normal pH conditions.

In terms of applications, the scope of the approach goes beyond the optimization of monoclonal antibodies, and any protein for therapeutic use could benefit from this type of approach. All the antibody-derived formats can be included here, as well as more recent formats based on other folds (fibronectin, lipocalin, ankyrin ...) and also peptides. The application of the *in silico* protocol to enhance the cross-reactivity to the antigen in other species is another potential application. In this case, the protocol could guide the design in order to obtain equipotent protein binding to both the human antigen and the antigen in a species relevant for toxicity studies, *e.g.* cynomologus.

In further studies, several parts of the *in silico* protocol could be improved. The data set we have collected will help to develop those improvements. One can cite the ability to exhaustively sample and score all the possible combinations of mutations (double, triple ...), and the ability to accurately model and score insertions within the loops is an even more challenging task. In terms of execution time, the use of a Generalized Born model instead of Poisson−Boltzmann for modeling solvation could considerably speed up the calculations. Of course the cost in terms of predictivity needs to be measured. The explicit modeling of crystallographic waters warrants also further investigations since it is known that some of these water molecules mediate key interactions.[30]

The current flowchart of the *in silico* protocol could also be extended by applying more CPU-intensive methods based for example on molecular dynamics simulations of the protein−protein interface. In particular, such simulations would allow to quantify the entropic part of the binding event with the aim to refine the estimation of $\Delta G_{binding}$.[28] Also, the explicit modeling of the backbone flexibility, which is one of the main limitations of the presented approach, could be explored.

MATERIALS AND METHODS

Data Set and Protein Setup. The crystallographic structures of 6 antibody/antigen complexes and of 1 non-immune-related protein/protein complex were used: the anti-lysosymes D44.1 (pdb code: 1MLC,[31] resolution of 2.5 Å), D1.3 (pdb code: 1VFB,[30] resolution of 1.8 Å), and HyHEL-63 (pdb code: 1DQJ,[32] resolution of 2.0 Å) antibodies, the anti-EGFR cetuximab antibody (pdb code: 1YY9,[33] resolution of 2.61 Å), the anti-HER2 (trastuzumab) variant bH1 (pdb code: 3BE1,[26] resolution of 2.9 Å), the anti-VLA1 antibody in complex with the I domain of the integrin VLA1 (pdb code: 1MHP,[34] resolution of 2.8 Å), and a nonimmune-related complex between barnase and barstar (pdb code: 1BRS,[35] resolution of 2.0 Å). The following protocol was applied to each of the protein complexes, using the MOE software.[36] Water molecules were deleted. Missing residues were omitted from the model. This latter operation did not have any influence on the results because none of the missing residues were found at the interface or in its vicinity. Finally, hydrogens were added. The resulting complex structures were minimized with harmonic constraints on all heavy atoms (100 kcal/mol/Å^2) to remove steric clashes. The force field used was AMBER94, with a reaction field model for implicit solvation, a cutoff of 10 Å on nonbonded interactions, a dielectric constant of 4 inside the protein, and 80 for the solvent. The protonate 3D[37] protocol from MOE[36] was used to determine the tautomeric form and protonation state of histidines.

Protocol/Workflow. As illustrated in the Figure 3, the affinity maturation protocol includes different steps, which deal mainly with conformational sampling and calculation of the free energy of binding.

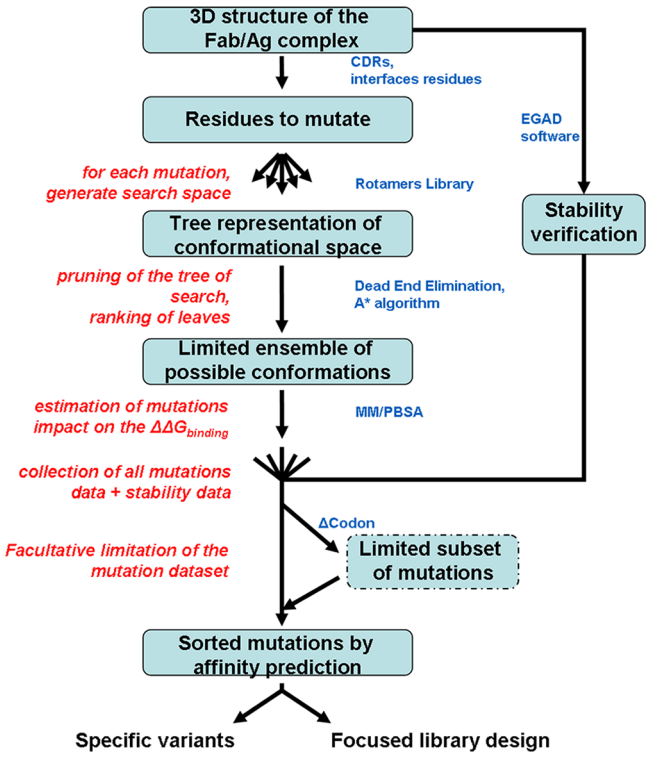


Figure 3. *In silico* affinity maturation workflow. The starting point is a 3D structure of the protein−protein complex. Residues to mutate are selected by CDR definition or by proximity to the partner of interaction (interface residues). Subsequently, each mutation is processed separately. The conformational search space, limited to the side chains of mutated residues and to side chains of neighboring residues, is generated by using a library of rotamers. The search space is organized as a search tree, Dead End Elimination pruning and the A* algorithm are applied.[12−14] The resulting conformations with the lowest total energy (calculated with MM) are then submitted to MM/PBSA calculation in order to estimate the effect of the mutation on $\Delta\Delta G_{binding}$. The stability of all mutations is verified by EGAD. The resulting data set can be restricted by their $\Delta codon$ in order to the lower the number of proposals. The outputs are the list of best mutations that could be used to propose point mutations. Alanine-scanning detection of hot-spots can be used to orient the design of a focused library for *in vitro* affinity maturation.

dx.doi.org/10.1021/ci3001474 | J. Chem. Inf. Model. 2012, 52, 2204−2214

Conformational Search Space. To limit the search space to a reasonable size, we kept the backbone fixed and carried out the search using a limited number of side-chain rotamers. For this task the DeMaeyer library[38] was found to be a good compromise between a thorough sampling of 3D conformations and ensuring the convergence of the conformational sampling by the DEE-A* algorithm. We added hydroxyl and sulfhydryl sampling (every 30°) to the library, in order to allow a conformational sampling with an all-atom model. This step increased the possible number of rotamers from 330 to 671. All neighboring side-chains with at least one heavy atom within a radius of 4 Å of the heavy atoms of the side chain of the mutated residue were also considered as flexible, and their degrees of freedom were sampled together with the mutated residue.

Conformational Sampling Algorithm (DEE/A*). As described in the work of Desmet et al.,[12] the conformational search space can be considered as a search tree, by exploiting the pairwise decomposability of the interactions in the physical model. Therefore, the first step of the analysis consisted in the precalculation of an energy matrix of pairwise interactions that is subsequently used by the DEE/A* algorithm. This first step was carried out with the CHARMM software[32] (version 32.1). The initial structure is initially relaxed by applying a short minimization performed with the CHARMM27 force field[33] (500 steps of ABNR, cutoff 14 Å and with internal and external dielectric constant of respectively 1 and 80), and then interactions between residue pairs are calculated in the gas phase with CHARMM defaults parameters.

One search tree was built for each mutation. The Dead-End Elimination (DEE[12−14]) was used to limit the size of the tree, and the A* algorithm[14] was used to find the conformations with the lowest energies, including the global minimum energy conformation (GMEC). DEE and A* steps are based on the previsouly calculated pairwise energies of interactions as described in the Leach and Leamon work.[14] These algorithms had been implemented in Perl. Only the GMEC and conformations within 10 kcal/mol from the GMEC were kept, with a limit of the 30 lowest energy conformations per mutation.

Computational Estimation of the Free Energy of Binding. For each mutation, the 30 conformations obtained by the procedure described above were then submitted to a more accurate protocol using the AMBER[41] software. Here the effect of a mutation on $\Delta G_{binding}$ is estimated by the difference between the energy of binding of the mutated system ($\Delta G_{binding,mut}$) and the energy of binding of the reference system ($\Delta G_{binding,wt}$)

$$\Delta\Delta G_{binding} = \Delta G_{binding,mut} - \Delta G_{binding,wt} \quad (1)$$

In general, the free energy of binding can be decomposed[28] into an enthalpic component in the gas phase ($\Delta H_{gas\ binding}$), a desolvation term (ΔG_{desolv}) to take solvation effects into account, and an entropic term (ΔS)

$$\Delta G_{binding} = \Delta H^{gas}_{binding} + \Delta G_{desolv} - T\Delta S \quad (2)$$

$\Delta H_{gas\ binding}$ is the sum of the electrostatic (ΔH_{elec}) and van der Waals (ΔH_{vdW}) interactions between partners of the complex, and it includes also an internal energy term (ΔH_{intra}) in order to account for the structural changes induced by binding. For the purpose of calculating $\Delta\Delta G_{binding}$ one can assume that the mutation will not perturb significantly the protein structures; therefore, $\Delta H_{intra} = 0$ and subsequently $\Delta\Delta H_{intra} = 0$. The term ΔG_{desolv} is divided into a nonpolar ($\Delta G_{nonpol\ desolv}$) and an electrostatic component ($\Delta G_{elec\ desolv}$). The first term was calculated as the sum of a cavitation term and a solvent−solute van der Waals interaction term, and the second term was evaluated using the Poisson−Boltzmann equation. Given that the backbone had been kept fixed, and in order to allow for a large number of mutations to be evaluated, the impact of the mutations on translational, conformational, and vibrational entropies was neglected ($\Delta\Delta S = 0$).

The Poisson−Boltzmann continuum electrostatic solvation model[42] and a continuum solvent vdW model[43] have been used to calculate the ΔG_{desolv} term (Poisson−Boltzmann solver of sander) with the AMBER[41] software (version 10). The main parameters were as follows: a dielectric constant of 4 within the protein and 80 in the implicit solvent, ionic strength of 0.145 M, probe sphere of 1.4 Å radius, grid size of 0.5 Å, charges and atoms radii from the Amber94 force field. $\Delta H_{gas\ binding}$ was calculated with the same force field and parameters. Instead of using a single value from Poisson−Boltzmann calculations, the $\Delta G_{elec\ desolv}$ of each conformation was obtained by averaging the energies over three 0.125 Å translations along the xyz diagonal. This later step was performed to remove the dependency of the calculated value on the initial 3D position of the grid.

Boltzmann Averaging of Energies. According to the protocol described above, each mutation results in up to 30 conformations with the corresponding energy values. One strategy to determine $\Delta\Delta G_{binding}$ would be to take the best value for each mutation. However, one could argue that for each mutation one needs to consider different side chain rotamer states. In order to account for this, the 30 energies were combined via a Boltzmann averaging procedure. This procedure was, for example, successfully applied to average energies of interaction.[20,44] In this study Boltzmann averaging was carried out in two different ways:

- Boltzmann averaging of the $\Delta G_{binding}$ values of each complex, implying that for a mutation with multiple conformations i, one can write eq 3

$$\Delta G^{averaged}_{binding} = \sum_i \Delta G^{i}_{binding} \times \frac{e^{-\Delta G^{i}_{total}/k_BT}}{\sum_i e^{-\Delta G^{i}_{total}/k_BT}} \quad (3)$$

- Boltzmann averaging of the energies of the antigen A, of the antibody B, and of the complex AB before the calculation of an averaged $\Delta\Delta G_{binding}$, as shown in eq 4

$$\Delta G^{averaged}_{binding} = G^{averaged}_{AB} - G^{averaged}_{A} - G^{averaged}_{B} \quad (4)$$

where for each of the three terms the Boltzmann averaging scheme is applied individually (eq 5)

$$G^{averaged}_{x} = \sum_i G^{i(x)}_{x} \times \frac{e^{-G^{i(x)}_{total}/k_BT}}{\sum_i e^{-G^{i(x)}_{total}/k_BT}} \quad (5)$$

Computational Estimation of the Protein Stability. The EGAD software[45] was used to predict the effect of the mutation on protein stability. It has been shown that this algorithm is able to estimate the stability of a protein with an average unsigned error of 1 kcal/mol.[45] It was compared to other tools and shown to be the best method despite the incorporation of a fixed backbone paradigm.[46] Two indicators resulting from EGAD analysis were used to predict stability: The first one is the folding energy variation due to the mutation. The second one is a confidence indicator based on the number of steric clashes with the environment induced by the mutation compared to the wild-type structure. If the

surrounding is expected to be disturbed by the introduced mutation it is assumed to be a destabilizing mutation. The following criteria were applied to all of the mutations in this study: if the EGAD estimation of stability change is higher than 3 kcal/mol or if the change in the number of clashes is higher than 2, then the mutation is considered as destabilizing.

Statistical Evaluation of Predictions. We used sensitivity and specificity scores to assess the rate of positives or negatives found. The sensitivity is the total of true positives divided by the sum of true positives and false negatives. The specificity is the total of true negatives divided by the sum of true negatives and false positives. A value of 0.5 corresponds to a random prediction, and a lower value means a prediction worse than random. In order to come up with a measure that is easier to comprehend, we introduced also the definition of success rate (*SR*), as follows

$$SR = \frac{TP}{TP + FP} \tag{6}$$

where *TP* is the number of true positives, and *FP* is the number of false positives. However, in the case of alanine scanning another way of defining the success rate was necessary. The success of an alanine scanning is based on the prediction of unfavorable mutation; therefore, we based the *SR* of alanine scanning analysis on negatives as follows

$$SR = \frac{TN}{TN + FN} \tag{7}$$

where *TN* is the number of true negatives, and *FN* is the number of false negatives.

■ ASSOCIATED CONTENT

Supporting Information

Predictions and experimental data for all the protein systems. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: nicolas.baurin@sanofi.com.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bugelski, P. J.; Martin, P. L. Concordance of preclinical and clinical pharmacology and toxicology of therapeutic monoclonal antibodies and fusion proteins: cell surface targets. *Br. J. Pharmacol.* **2012**, *166* (3), 823−846.

(2) Martin, P. L.; Bugelski, P. J. Concordance of preclinical and clinical pharmacology and toxicology of monoclonal antibodies and fusion proteins: soluble targets. *Br. J. Pharmacol.* **2012**, *166* (3), 806−822.

(3) Honig, P. K. The right medicines–just not enough of them. *Clin. Pharmacol. Ther.* **2012**, *91* (3), 369−372.

(4) Reichert, J. M. Monoclonal antibodies as innovative therapeutics. *Curr. Pharm. Biotechnol.* **2008**, *9* (6), 423−430.

(5) Levin, A. M.; Weiss, G. A. Optimizing the affinity and specificity of proteins with molecular display. *Mol. Biosyst.* **2006**, *2* (1), 49−57.

(6) Zhu, Z.; Hattori, K.; Zhang, H.; Jimenez, X.; Ludwig, D. L.; Dias, S.; Kussie, P.; Koo, H.; Kim, H. J.; Lu, D.; Liu, M.; Tejada, R.; Friedrich, M.; Bohlen, P.; Witte, L.; Rafii, S. Inhibition of human leukemia in an animal model with human antibodies directed against vascular endothelial growth factor receptor 2. Correlation between antibody affinity and biological activity. *Leukemia* **2003**, *17* (3), 604−611.

(7) Wu, T. T.; Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **1970**, *132* (2), 211−250.

(8) Foote, J.; Winter, G. Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **1992**, *224* (2), 487−499.

(9) Chen, R.; Tong, W.; Mintseris, J.; Li, L.; Weng, Z. ZDOCK predictions for the CAPRI challenge. *Proteins* **2003**, *52* (1), 68−73.

(10) Stura, E. A.; Graille, M.; Housden, N. G.; Gore, M. G. Protein L mutants for the crystallization of antibody fragments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58* (Pt10 Pt 1), 1744−1748.

(11) Voigt, C. A.; Gordon, D. B.; Mayo, S. L. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **2000**, *299* (3), 789−803.

(12) Desmet, J.; De Maeyer, M.; Hazes, B.; Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **1992**, *356* (6369), 539−542.

(13) Goldstein, R. F. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J.* **1994**, *66* (5), 1335−1340.

(14) Leach, A. R.; Lemon, A. P. Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. *Proteins* **1998**, *33* (2), 227−239.

(15) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897.

(16) Lippow, S. M.; Wittrup, K. D.; Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **2007**, *25* (10), 1171−1176.

(17) Clark, L. A.; Boriack-Sjodin, P. A.; Eldredge, J.; Fitch, C.; Friedman, B.; Hanf, K. J.; Jarpe, M.; Liparoto, S. F.; Li, Y.; Lugovskoy, A.; Miller, S.; Rushe, M.; Sherman, W.; Simon, K.; Van Vlijmen, H. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci.* **2006**, *15* (5), 949−960.

(18) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789−6801.

(19) Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. FLAP: GRID molecular interaction fields in virtual screening. validation using the DUD data set. *J. Chem. Inf. Model.* **2010**, *50* (8), 1442−1450.

(20) Paulsen, J. L.; Anderson, A. C. Scoring ensembles of docked protein:ligand interactions for virtual lead optimization. *J. Chem. Inf. Model.* **2009**, *49* (12), 2813−2819.

(21) David, M. P.; Asprer, J. J.; Ibana, J. S.; Concepcion, G. P.; Padlan, E. A. A study of the structural correlates of affinity maturation: antibody affinity as a function of chemical interactions, structural plasticity and stability. *Mol. Immunol.* **2007**, *44* (6), 1342−1351.

(22) Li, L.; Zhao, B.; Cui, Z.; Gan, J.; Sakharkar, M. K.; Kangueane, P. Identification of hot spot residues at protein-protein interface. *Bioinformation* **2006**, *1* (4), 121−126.

(23) Diller, D. J.; Humblet, C.; Zhang, X.; Westerhoff, L. M. Computational alanine scanning with linear scaling semiempirical quantum mechanical methods. *Proteins* **2010**, *78* (10), 2329−2337.

(24) Peled, J. U.; Kuang, F. L.; Iglesias-Ussel, M. D.; Roa, S.; Kalis, S. L.; Goodman, M. F.; Scharff, M. D. The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.* **2008**, *26*, 481−511.

(25) Liu, M.; Schatz, D. G. Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* **2009**, *30* (4), 173−181.

(26) Bostrom, J.; Yu, S. F.; Kan, D.; Appleton, B. A.; Lee, C. V.; Billeci, K.; Man, W.; Peale, F.; Ross, S.; Wiesmann, C.; Fuh, G. Variants of the antibody herceptin that interact with HER2 and VEGF at the antigen binding site. *Science* **2009**, *323* (5921), 1610−1614.

(27) Huo, S.; Massova, I.; Kollman, P. A. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* **2002**, *23* (1), 15−27.
(28) Zoete, V.; Irving, M. B.; Michielin, O. MM-GBSA binding free energy decomposition and T cell receptor engineering. *J. Mol. Recognit.* **2010**, *23* (2), 142−152.
(29) Chaparro-Riggers, J.; Liang, H.; Devay, R. M.; Bai, L.; Sutton, J. E.; Chen, W.; Geng, T.; Lindquist, K.; Casas, M. G.; Boustany, L. M.; Brown, C. L.; Chabot, J.; Gomes, B.; Garzone, P.; Rossi, A.; Strop, P.; Shelton, D.; Pons, J.; Rajpal, A. Increasing Serum Half-life and Extending Cholesterol Lowering in Vivo by Engineering Antibody with pH-sensitive Binding to PCSK9. *J. Biol. Chem.* **2012**, *287* (14), 11090−11097.
(30) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (3), 1089−1093.
(31) Braden, B. C.; Souchon, H.; Eisele, J. L.; Bentley, G. A.; Bhat, T. N.; Navaza, J.; Poljak, R. J. Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* **1994**, *243* (4), 767−781.
(32) Li, Y.; Li, H.; Smith-Gill, S. J.; Mariuzza, R. A. Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63(,). *Biochemistry* **2000**, *39* (21), 6296−6309.
(33) Li, S.; Schmitz, K. R.; Jeffrey, P. D.; Wiltzius, J. J.; Kussie, P.; Ferguson, K. M. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **2005**, *7* (4), 301−311.
(34) Karpusas, M.; Ferrant, J.; Weinreb, P. H.; Carmillo, A.; Taylor, F. R.; Garber, E. A. Crystal structure of the alpha1beta1 integrin I domain in complex with an antibody Fab fragment. *J. Mol. Biol.* **2003**, *327* (5), 1031−1041.
(35) Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry* **1994**, *33* (30), 8878−8889.
(36) Chemical Computing Group *MOE (Molecular Operating Environment)*, v2009.10; Montreal, Canada, 2009.
(37) Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75* (1), 187−205.
(38) De Maeyer, M.; Desmet, J.; Lasters, I. All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Folding Des.* **1997**, *2* (1), 53−66.
(39) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.
(40) MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56* (4), 257−265.
(41) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668−1688.
(42) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **2002**, *23* (13), 1244−1253.
(43) Tan, C.; Tan, Y. H.; Luo, R. Implicit nonpolar solvent models. *J. Phys. Chem. B* **2007**, *111* (42), 12263−12274.
(44) Li, Y.; Liu, Z.; Wang, R. Test MM-PB/SA on true conformational ensembles of protein-ligand complexes. *J. Chem. Inf. Model.* **2010**, *50* (9), 1682−1692.
(45) Pokala, N.; Handel, T. M. Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J. Mol. Biol.* **2005**, *347* (1), 203−227.
(46) Potapov, V.; Cohen, M.; Schreiber, G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng., Des. Sel.* **2009**, *22* (9), 553−560.