Published in final edited form as:

Proteins. 2010 August 15; 78(11): 2490–2505. doi:10.1002/prot.22757.

Antibodies as a model system for comparative model refinement

Benjamin D. Sellers^{1,2}, Jerome P. Nilmeier², and Matthew P. Jacobson^{1,*}

¹ Department of Pharmaceutical Chemistry, University of California, San Francisco, California

² Graduate Group in Biophysics, University of California, San Francisco, California

Abstract

Predicting the conformations of loops is a critical aspect of protein comparative (“homology”) modeling. Despite considerable advances in developing loop prediction algorithms, refining loops in homology models remains challenging. In this work, we use antibodies as a model system to investigate strategies for more robustly predicting loop conformations when the protein model contains errors in the conformations of side chains and protein backbone surrounding the loop in question. Specifically, our test system consists of partial models of antibodies in which the “scaffold” (i.e., the portion other than the complementarity determining region, CDR, loops) retains native backbone conformation, while the CDR loops are predicted using a combination of knowledge-based modeling (H1, H2, L1, L2, and L3) and *ab initio* loop prediction (H3). H3 is the most variable of the CDRs. Using a previously published method, a test set of 10 shorter H3 loops (5–7 residues) are predicted to an average backbone (N-C α -C-O) RMSD of 2.7 Å while 11 longer loops (8–9 residues) are predicted to 5.1 Å, thus recapitulating the difficulties in refining loops in models. By contrast, in control calculations predicting the same loops in crystal structures, the same method reconstructs the loops to an average of 0.5 Å and 1.4 Å for the shorter and longer loops, respectively. We modify the loop prediction method to improve the ability to sample near-native loop conformations in the models, primarily by reducing the sensitivity of the sampling to the loop surroundings, and allowing the other CDR loops to optimize with the H3 loop. The new method improves the average accuracy significantly to 1.3 Å RMSD and 3.1 Å RMSD for the shorter and longer loops, respectively. Finally, we present results predicting 8–10 residue loops within complete comparative models of five non-antibody proteins. While anecdotal, these mixed, full-model results suggest our approach is a promising step towards more accurately predicting loops in homology models. Furthermore, while significant challenges remain, our method is a potentially useful tool for predicting antibody structures based upon a known Fv scaffold.

Keywords

loop prediction; homology modeling; comparative; refinement; all-atom; physics-based force field

Introduction

Reliably accurate models of proteins would be useful to biological and therapeutic studies that investigate protein function at the atomic level. Though tens of thousands of experimental protein structures exist¹, millions of protein sequences, many with unknown function, have been discovered². To address this large gap between numbers of sequences and structures, comparative (or homology) models have been utilized as surrogates for experimental structures in a variety of successful biological studies. Examples include

* Correspondence to Department of Pharmaceutical Chemistry, University of California, San Francisco, Box 2240, San Francisco, CA 94158-2517 (matt.jacobson@ucsf.edu) .

inhibitor discovery³⁻⁷, enzymatic function prediction⁸, and protein-protein docking⁹⁻¹¹. While homology models have been used successfully in many of these applications, in general comparative models are not as useful as crystal structures in applications requiring atomic-level accuracy. For example, McGovern and Shoichet¹² compared docking results with crystal structures and homology models and found that, in general, homology modeled receptors produced worse results. A general method for producing high-accuracy comparative models would extend their usefulness in many applications. In our view, a general method has yet to be developed.

The Critical Assessment of Techniques for Protein Structure Prediction (CASP7)¹³ showed only modest progress in the development of high-accuracy modeling methods. For the template-based (comparative-modeling) category, an important metric for success is whether predicted protein structures are more accurate than the starting homolog template protein, that is, whether the model can be *refined* closer to the native structure. Though some submitted models were closer to the native structure than the best template, no single method improved upon the optimal template on average¹⁴.

Errors in comparative models can be attributed to: 1) errors resulting from limitations of the modeling tools, including inadequate sampling of protein conformations and inaccuracies in the energy or scoring function used and 2) errors in sequence alignments. In this work, we focus exclusively on the first of these challenges.

Loop refinement is an important aspect of protein refinement. Since the overall protein fold is generally conserved between proteins with >30% sequence identity¹⁵, loop regions often show the greatest structural diversity among homologous proteins. Many researchers have validated loop prediction methods by first removing loops from high resolution crystal structures, and then assessing the ability to reconstruct the conformation *de novo*¹⁶⁻²⁵. However, predicting loops in comparative models is a more difficult problem, as we discuss in detail below. Some loops are also inherently flexible and can be found in various conformations in different crystal environments or with different binding partners²⁶.

In comparative models, where small and large errors exist throughout the protein, loop prediction is more difficult because portions of the protein near the loop may be modeled incorrectly, which can make it difficult both to sample a near-native conformation, and to identify it as a near-native conformation using an appropriate scoring function. Three types of errors can surround loops: 1) the positions of the side chain atoms of surrounding residues may be incorrect, 2) the positions of backbone atoms of surrounding residues may be incorrect, or 3) the positions of loop stem backbone atoms, residues adjacent to the loop, may be incorrect. We choose to separate out the issue of incorrect loop stems from that of other incorrect surrounding residues as they create different sampling problems (i.e. incorrect stems modulate the loop geometry and do not necessarily cause steric clashes). Others have developed methods to address incorrectly modeled loop stems²⁷.

In previous work²⁸, we evaluated loop refinement in a simple, model system that contains exclusively side chains errors in residues surrounding loops, error type 1 above. Using this model system, we evaluated a method that refines loops while simultaneously optimizing rotamers of surrounding residues. The method produced median backbone RMSDs <1.3 Å in 6-12 residue loops in our 80 perturbed models.

In this work, we aimed to evaluate loop predictions in a model system that exhibits backbone errors in residues surrounding loops, error type 2 above, in addition to side chain errors. We chose a test set of H3 loops in models of antibody variable fragments (Fv) as our model system. The large and growing importance of antibodies as therapeutics and biochemical tools has motivated interest in comparative modeling of antibodies^{10,29-36}.

Antibodies, through the processes of recombination and somatic hypermutation, display enormous variety in sequence. The overall fold however is very well conserved, and the structural diversity is largely confined to the 6 hypervariable loops comprising the complementarity determining region which is responsible for antigen recognition. Thus, generating accurate comparative models of antibodies is largely a loop prediction problem. In addition, there are large numbers of antibody structures for evaluating modeling methods.

Of the 6 hypervariable loops, the conformations of 5 can often be predicted with good accuracy using well-known, sequence-structure rules³⁷⁻⁴⁰. That is, the backbone conformation of CDR loops L1, L2, L3, H1, and H2 in non-synthetic antibodies generally fall into clusters with a variation of ~ 1 Å RMSD and can be predicted using knowledge-based methods, the so-called canonical loop rules³⁷⁻⁴⁰. The remaining loop, H3, is the most diverse in both sequence and structure, and is the most critical loop for antigen binding. Though several attempts have been made to classify the H3 loop in a similar manner⁴¹⁻⁴⁶, the large structural variability makes the problem challenging, and we view *ab initio* loop prediction as the most promising and general approach. There is some precedent for this strategy. Cardoza et al modeled the H3 loop in a homology model of antibody E52 and found that errors in the surrounding residues prevented identification of a native-like loop⁴⁷. Fine et al.⁴⁸ and Brucoleri et al.⁴⁹ modeled H3 through random sampling followed by scoring with the CHARMM50 force field, while Whitelegg et al.⁵¹ and Vlijmen and Karplus¹⁸ utilized a combination of knowledge-based and *ab initio* methods to sample H3 loop conformations. Sivasubramanian et al.³⁵ used knowledge-based information to generate a specialized backbone fragment library for antibodies. There have also been a number of purely knowledge-based approaches to modeling H3 which will not be reviewed here.

In this paper, we focus on refining the conformation of H3 in the context of a model in which the other loops contain appreciable errors in backbone (mostly < 2 Å RMSD) and side chain conformations. From a non-redundant library of 49 high-resolution x-ray crystal structures, we constructed a test set of 21 Fv models with 5–9 residue H3 loops. Initial, unrefined models were constructed by grafting canonical CDR loop-templates from chains with $< 60\%$ sequence identity from the target onto the native framework, the conserved secondary structure in the Fv which does not include the CDR loops. By grafting CDR loop templates on the native framework, we purposely simplify the problem by eliminating certain factors that can also affect H3 loop prediction, including variability in the heavy and light-chain domain orientation and errors in the H3 loop-stem positions. The H3 loop was then modeled in a non-optimized starting conformation as typical of homology modeling.

In this work, we first show that our previously published¹⁶ *ab initio* loop sampling combined with a physics-based force field can accurately reproduce 5–9 residue H3 loop conformations in x-ray crystal structures of antibody Fv domains on average. The same algorithm applied to comparative models of the same antibodies performs much worse. We alter the loop sampling to include optimization of the backbone and side chain atoms in the residues surrounding H3. The number of degrees of freedom sampled is much larger, which makes the problem challenging, but by doing so we predict near-native loops within many of these antibody models. This approach complements other studies in which the backbone atoms surrounding loops have been optimized⁵²⁻⁵⁴, and studies in which the effects of the environment on loop prediction have been investigated^{20,47}. In addition, since H3 is often directly involved in antigen binding, we discuss the effects of antibody binding partners (i.e. induced fit) on H3 loop prediction accuracy.

Our goal is to apply lessons learned from the simplified antibody test system considered here to the more general problem of loop refinement in comparative models within a variety

of protein families. As a first step, we have applied the method described here to loop refinement in 5 (non-antibody) comparative models. Though challenges remain for our method (e.g., accounting for loop stem variability), we show that it improves results over the starting model and our previous method in some cases. Although these results are anecdotal, they suggest that our approach is a step towards a general refinement method.

Methods

Non-redundant library of antibodies

Our first goal was to create a non-redundant set of high-resolution antibody crystal structures in the Protein Data Bank¹. We searched the PDB for antibody Fv and Fab structures. Single chain (scFv), homo-dimer, and single-domain (e.g. camelid) antibodies were removed leaving an initial set of 459 structures. We clustered these antibody structures using Pisces⁵⁵⁻⁵⁶ with a cutoff of 80% sequence identity, 2.2 Å resolution, and 0.3 R-value. The method clusters each protein chain separately. The result was a non-redundant list containing unpaired heavy and light chains. To generate complete antibody variable fragments, for a given light chain in the non-redundant list, the associated heavy chain was included and vice versa. The final data set included 49 PDB structures. The 21 antibodies in the non-redundant antibody library that have H3 loops of length 5 to 9 residues were used as the test cases in this study.

We constructed multiple sequence alignments of the heavy and light chain variable domains from this library using CLUSTAL⁵⁷. For structures that contained more residues than found in variable fragments (Fv), we truncated the sequences eight residues after the C-terminal end of CDR3. This corresponds roughly to the end of the final beta strand in an Fv.

Force-field and implicit solvent model

As in previous work, we calculate energy using the Optimized Potential for Liquid Simulations all atom (OPLS-AA) force field⁵⁸⁻⁶⁰, the Surface Generalized Born model⁶¹ of polar solvation, an estimator for the nonpolar component of the solvation free energy developed by Gallicchio et al.⁶², and a number of correction terms as detailed in Ghosh et al.⁶¹ and Jacobson et al.⁶⁰. All results are reported for the structures with the lowest energy.

To assess the effects of antigens on loop conformation, we performed control calculations in which we included antigens that are present in the crystal structure (see Results). Small-molecule antigens were parameterized for the 2005 OPLS-AA force-field⁵⁸⁻⁶⁰ using the *hetgrp_ffgen* program (Schrödinger, Inc.).

Construction of antibody Fv models

Antibody Fv comparative models were constructed as follows (see Figure 1):

1. Variable light and heavy chains are modeled separately. As described in detail below, all CDR loops were then removed and each CDR loop excluding H3 was selected from a library of antibody CDR loops and grafted onto the conserved framework. The H3 loop is left truncated initially.
2. The heavy and light chain models are then combined into a single model. Because the models are constructed using the native scaffold residues, the heavy and light chains are already in the native orientation relative to each other. We simply concatenate two PDB files of heavy and light chain models into a single file.
3. All side chains with steric clashes are optimized using the side chain optimization function⁶⁰ in the Protein Local Optimization Program (PLOP)⁶³⁻⁶⁵. Steric clashes are defined as any residue containing atoms with overlap factor <0.7 with any other

atom. Overlap factor is defined as the distance between two atom centers divided by the sum of the two atomic radii. The side chain prediction is executed with a minimum overlap factor of 0.3 and repeated 4 times to enhance sampling.

4. The missing H3 loop is constructed *ab initio* and deliberately placed in a non-optimized starting conformation. To do so, we ran a single execution of the PLOP loop predict command on the H3 loop with the following parameters. The amount of sampling was reduced using the `nconf_min=1` parameter which instructs PLOP to generate a minimum of 1 loop conformation (default is `nconf_min=2N` where N is the number of residues in the loop). Because it can be difficult to generate a loop when the surrounding residues are modeled incorrectly (data not shown), we allow some overlap between the H3 loop and the surroundings by reducing the overlap factor, `ofac=0.5` (default is 0.7). The surrounding residue atoms are held fixed during this initial loop generation.

The result is a complete Fv model, with native-like scaffold residues, template-modeled canonical loops for the non-H3 loops, and an initial, non-optimized H3 loop. Further details of the canonical loop modeling follow.

Definition of CDR loop residues

In all cases, CDR loops are defined as in Chothia and Lesk37. Specifically, the first residue of the loop is the residue following a CTR or CAR motif and the last residue of the H3 loop is the residue before the conserved WG motif. All side chains and the 2N dihedral angles (where N is the H3 loop length) of the backbone connecting these two residues are predicted *ab initio*. Peptide bonds preceding Pro are allowed to assume *cis* or *trans* conformations. Bond lengths and angles are allowed to vary during energy minimization. The endpoint residues for the 21 test cases are listed in Table S4.

Modeling of canonical CDR loops

Though previous studies have identified H3 structural classes⁴¹⁻⁴⁵, H3 is more variable in sequence and structure than the other CDR loops. As described below, we use *ab initio* loop prediction for H3, which we feel is a more general solution. When possible, the H1, H2, L1, L2, and L3 loops in our models were constructed using canonical loop rules, a set of mappings between loop length and key positions in the Fv chain sequence to a set of loop backbone coordinates.

We constructed a library of canonical loops from which templates are chosen for the non-H3 CDRs as follows. Each CDR loop, excluding H3, in the non-redundant antibody library (49 structures) was classified by canonical class as defined in Martin et al⁴⁰. A loop was labeled as a class representative if >75% of the residues matched the sequence rule.

In the target sequence, the CDR loops were identified and assigned to a canonical class. If a loop did not match at least 75% of the residues in any sequence rule, the loop was labeled “non-canonical” and modeled as described in the next section. For those assigned to a canonical class, we chose the best loop-template from the library given the following criteria: the sequence-identity between the template chain and the target chain has the highest sequence identity of all possible class templates while <60% similar in sequence to the target. The 60% cutoff was used to reduce bias towards the native CDR conformation.

Each CDR in the target structure was removed and replaced using the loop-template as follows. The loop-template structure is first structurally aligned to the scaffold-template by aligning the C α coordinates from: a) the three residues preceding the N-terminus and following the C-terminus of the CDR loop, and b) the three residues centered on each

cysteine involved in a conserved disulfide bridge. Initial results (data not shown) showed that aligning to either all the residues in the conserved framework or just the stem residues frequently produced poor structural alignment of the loop end-points. The loop-templates were then grafted onto the native framework using the multiple-template homology modeling feature in PLOP. This function produces a complete model based on 1) a multiple sequence alignment between the target and template sequences, 2) a set of structurally aligned templates, and 3) specification of which template should be used at each target residue position. The backbone conformation of residues within gaps, insertions, and template transitions are minimally refined using the loop prediction feature in PLOP and all non-conserved side chains are optimized and energy minimized. CDR accuracies in the initial comparative model are shown in Table S1 and explained in the results section. Canonical class assignments and template choices for all non-H3 loops are shown in Table S2.

Modeling of non-canonical CDR loops

Of the 105 non-H3 loops in our test set, 19 were not modeled by a canonical class template (see Table S2). Ten of these are non-canonical loops while 9 more are canonical loops with no template in our database that met our sequence identity cutoff. In either case, a non-H3 CDR loop of length N is modeled from a loop template with the same length N with the highest chain sequence identity $<60\%$. If no loop template exists in the non-redundant library of 49 antibodies (see above) that meet these criteria, then loops with $N-1$ residues are checked. If no loop templates are found, then loops of length $N+1$ are searched. The database is searched with alternating shorter and longer loop lengths ($N-1$, $N+1$, $N-2$, $N+2$, etc.) until a template is found. If a loop of different size is used, then the loop is modeled with an appropriate number of gaps or insertions in the sequence using the homology modeling command in PLOP66. For example, if a target L1 loop of length 10 cannot be modeled by a canonical loop-template and no other 10-residue loops exist from a chain $<60\%$ sequence identity with the target, then it may be modeled by a 9 residue L1 loop and the extra residue will be modeled as part of the homology modeling process in PLOP.

Prediction of the H3 loop

Initial models with non-optimized H3 loops were used to test various loop prediction methods.

Hierarchical Loop Prediction – HLP—The HLP protocol (as named here and previously²⁸) was first described in Jacobson et al.¹⁶. This protocol samples only the loop and not the surrounding residues.

HLP with Surrounding Side chain optimization - HLP-SS—The HLP-SS protocol was described previously²⁸ as a method for predicting loops when the surrounding side chains are incorrect. In summary, the method augments the HLP method in that 1) backbone sampling is increased through an additional “initial” stage in which surrounding side chains are excluded from the backbone screening process, and 2) in all stages, the surrounding side chains are optimized simultaneously with the loop side chains.

HLP with Surrounding Side chain and Backbone optimization - HLP-SSB—The new method described here modifies HLP-SS in two ways (see Figure 2).

1. Sampling is increased through
  - a. *Additional initial stages.* Three initial stages were added with reduced overlap factors. The five lowest energy loops in each initial stage are used as starting points in the refinement stages. The three additional initial

stages in HLP-SSB lead to a total of 30 round-one refinement stages compared to 15 in HLP-SS.

- b.** *Reduced steric screening.* The criterion we used to reduce steric screening is based on the “overlap factor”, defined as the ratio of the distance between two atom centers to the sum of the atomic radii. Overlap factors <1.0 indicate some overlap between atoms. As described previously in detail¹⁶, sampled loop conformations with large steric clashes are eliminated, based on the minimum-allowed overlap factor. In the three new initial stages, the overlap factor is reduced to sample more broadly. The overlap factor of the first refinement stage was reduced as well because mild steric clashes were found to exist in native-like conformations following the initial stages. In addition, in the third and sixth initial stages in HLP-SSB, loop backbone samples are not eliminated when they clash with the surrounding side chains in order to allow for native-like samples that may be initially occluded by incorrectly modeled side chains. This is implemented with the “sidefrz=no” option in PLOP. Specific parameters are shown in Figure 2.
- 2.** We optimize the backbone atoms of the non-H3 CDR loop residues before each refinement stage. Specifically, all non-H3 CDR backbone and side chain atoms are energy minimized as an integral part of the loop prediction protocol. The surrounding loops are minimized before each refinement stage using the Protein Local Optimization Program⁶³⁻⁶⁵. By alternating between sampling loops and minimizing the surroundings three times (i.e. predict H3, minimize CDRs, predict H3, minimize CDRs, predict H3), the protocol presented here enables native-like H3 loops to energy minimize simultaneously with the surrounding residues. We found that this iterative and alternating approach was necessary for good results (data not shown); simply optimizing the non-H3 loops prior to H3 loop prediction, for example, made results worse. Likewise, optimizing H3 along with the surrounding CDRs did not allow these loops to “open up” and thus move away from a near-native loop with initial steric clashes (data not shown).

Protonation states of titratable residues

All titratable residues are assigned standard protonation states at pH 7 (e.g. histidine has a single proton, at the δ nitrogen). One exception to this procedure is the protonation of Glu or Asp at position 35^H if residue 50^H is also a Glu or Asp. These framework residues form dyads between their carboxylic acid side chains, with 35^H more buried. We have verified this protonation state change through calculations of the pK_a's of these residues (data not shown) using the Multi-Conformer Continuum Electrostatics (MCCE) program^{64,65}. As a result, Glu35^H is protonated (i.e., neutral) in PDB 2cgr and 1yqv for all calculations.

Measurement of accuracy

In antibodies, accuracy was measured by root mean squared deviation (RMSD) with the conserved scaffold residues (i.e. all non-CDR residues) aligned. Backbone RMSD is measured over N, C α , C, and O atoms. Side chain RMSD of surrounding residues as found in Table S1 are calculated over all side chain heavy atoms. In the non-antibody comparative models, all residues, including the loop, in the starting model were first aligned to the native crystal structure using the MatchMaker function in Chimera⁶⁷ with default settings.

Because we strove to create a non-redundant test set of high-resolution antibodies, we are limited in the number of test cases for each loop length. To derive meaningful statistics, we report average and median RMSDs over several loop lengths. This is not rigorously valid

because long loop RMSDs can vary more greatly and dominate statistics over shorter loop results. To minimize this effect, we have broken our test set into shorter loops (5-7 residues) and longer loops (8-9 residues) report statistics for the two groups separately.

Reported P-values, calculated using Excel (Microsoft) using a 1-tailed, pairwise Student's t-Test, are used to measure statistical significance between two groups of RMSDs (ex. when comparing the increase in accuracy of HLP-SSB over the starting loop).

Generation of non-antibody comparative model test set

A selection of protein target structures were selected from CASP7, CASP8, and the CASP Model Refinement experiment (CASPR) 68. The results are anecdotal, but demonstrate that our method can improve loop refinement results in at least some cases. Within the set of CASP refinement targets, 8-10 residue loops were chosen manually which i) deviated from the native crystal structure and ii) when possible, were not involved with crystal or other molecular contacts. Test cases were removed when the native crystal structure had any one of the following attributes: pH<5.5, pH>8.5 or resolution >2.2 Å. Cases with crystal structures with no specified pH were included (TMR3, TMR5, TR289). Starting models were first aligned to the native crystal structure using the MatchMaker function in Chimera67 with default settings. Since there are no CDR loops in these non-antibody cases, during the backbone refinement stages in HLP-SSB, the residues within 7.5 Å of the starting loop are minimized.

Results

Antibody Fv test set

Attributes of our test set are shown in Table I. Twelve of the twenty-one Fv test cases contain antigens, seven with small molecules and five with protein antigens. H3 loops range from five to nine residues in length. The average sequence identity of the Fv's (combined across heavy and light chains) between all test cases is 54% with minimum and maximum identities of 40% and 78% respectively (see supplemental Table S3). 91% of the test cases share a sequence identity <65% so the test set is reasonably non-redundant.

Reconstructing H3 loops in antibody Fv crystal structures

We first establish baseline results by reconstructing H3 loops in crystal structures, i.e., removing the H3 loop and seeing how well we can predict its conformation when the remainder of the protein is in the native configuration. Past experience suggests these results should be much more accurate than in models of the same antibodies. Effects due to crystal packing and induced fit with antigens can also be investigated. While these control calculations are not representative of refinement scenarios found in comparative models, where crystal environmental factors are unknown, they are critical to determining how meaningful our results are when comparing predictions to the assumed "correct" answer, the crystal structure.

Using a previously published method¹⁶, HLP (see Methods), which does not optimize the surroundings of the loop, the average and median backbone RMSD for the H3 loop predictions are 0.5 Å and 0.5 Å, respectively for the short loop set and 1.4 Å and 0.8 Å, respectively for the longer loop set (Table I, column 4). In this first control experiment, we explicitly represented the effects of crystal packing from adjacent chains in the crystal and included antigen if present. Crystallographic waters were removed, however. The accuracy of these results are similar to previous results generated across a much larger loop prediction test set^{16,28}. Of the shorter loops, PDB 2cgr has the largest predicted backbone RMSD of 1.5 Å. This relatively poor result occurs because near-native conformations are not sampled;

the energy of the native loop is lower in energy than that of any sampled conformations. Upon inspection, the 7 residue H3 loop in 2cgr is sandwiched between the small-molecule antigen and a crystal contact. HLP uses stringent steric screening (see Methods), causing near-native backbone conformations to be eliminated in this case. A similar situation occurs in the longer loop case, 2fbj, which HLP predicts to 6.3 Å. The native loop, which has the lowest energy, is not sampled again due to steric screening in a tight pocket.

To assess the effects of crystal packing, we then predicted H3 loops using HLP without representing crystal symmetry (but still including antigens). In Table I, we omit RMSDs from cases that do not contain an antigen. The average and median backbone RMSDs are 1.1 Å and 0.4 Å respectively for the shorter loops and 1.0 Å and 0.8 Å respectively for the longer loops, suggesting that crystal packing does not play a major role in determining H3 structure in this test set. The increase in average RMSD for the shorter loops is primarily due to a single case, 2cgr, for which the backbone RMSD increases to 3.7 Å. The energy of the native loop is again lower than that of any of the sampled conformations for this case. This is not likely an effect of removing crystal energetic contacts, but rather due to HLP failing to sample a near-native backbone conformation in a tight conformational space. Among the long loop test cases, 1r3j increases to 2.9 Å RMSD and is also a sampling problem. In case 2fbj, sampling of near-native conformations increases when crystal packing is removed.

To assess the effects of antigens on H3 loop conformation, we performed our calculations using HLP without including antigens (or crystal packing). We expected the RMSDs to increase when the antigen was left out. Large increases in RMSD can be associated with induced-fit effects whereas small increases suggest the H3 loop is stable in the bound conformation even without antigen present. As seen in columns 5 and 6 in Table I, average and median backbone RMSDs of predictions without including antigens are 0.8 Å and 0.8 Å respectively for short loops and 2.9 Å and 1.6 Å for longer loops. While the median RMSD for short loops increases only slightly, both average and median RMSDs for long loops increase dramatically. The prediction in 1ct8 increases to 4.6 Å RMSD; this incorrect conformation has a predicted energy <1 kcal/mol less than that of the native loop which may indicate an antigen effect or an energy function problem (e.g. the pH of the crystal structure is pH 4.0 while we assume pH 7.0; see Methods). The prediction for 1iqd increases to 5.9 Å and 1r3j increases to 5.3 Å; both predictions are ~10 kcal/mol lower than the energy of the native loop which may indicate protein antigen effects. H3 loops in both 1iqd and 1r3j crystal structures form a salt-bridge and hydrogen bond with the antigen. Case 1wt5 is a sampling problem; the native loop is ~9 kcal/mol less than the prediction (8.0 Å RMSD). Interestingly, we see 3 cases with a significant (>1.0 Å) decrease in RMSD. 2cgr and 1uz9 have small molecule antigens, and these decreases may indicate some strain in the crystal structure, as measured by the force field we use, which is relieved when the antigen is removed. There also could be a forcefield parameterization issue for the small molecules, which could adversely affect predictions when the ligand is included. In summary, the longer loop predictions appear to be more affected by crystal packing and antigen effects. There are a number of reasons this may be the case. Longer loops may be more flexible and adopt multiple conformations in the unbound state or in solution. Longer loops may also simply extend further and thus make more crystal and antigen contacts.

As a further control to assess the effects of increasing the degrees of freedom sampled to include side chains surrounding H3, we predicted loops in native crystal structures using our previously published protocol²⁸, HLP-SS. We expected the accuracy to degrade slightly from the previous control where the surrounding residues are held fixed. Average and median RMSDs remain similar at 0.8 Å and 0.7 Å respectively for the short loops and increase to 3.0 Å and 2.3 Å respectively for the longer loops (Table I, column 7). This increase may be in part due to the broader sampling space explored by HLP-SS. The

increase may also be related to crystal packing and induced fit effects. Though crystal packing and antigens were removed in the previous control, the surrounding residues were held fixed in the native configuration. In this control, surrounding side chains are allowed to relax to the new environment, one without crystal packing or antigens.

As a final control, we applied our new method, HLP-SSB, which additionally optimizes the CDR residue side chains and backbone, on the native crystal structures. Average and median RMSDs remain similar at 1.0 Å and 0.7 Å respectively for the short loops and increase to 3.1 Å and 2.9 Å respectively for the longer loops (Table I, column 8). Again this increase in RMSD for the longer loops shows further relaxation of the H3 loop and the surrounding residues. This control represents a baseline to which we compare H3 predictions in antibody models. As in the case when predicting H3 in comparative models, information about antigens and crystal packing have been removed. But since the CDR residues start in the native conformation, this control represents the best we could expect to do when refining H3 in comparative models of the antibodies in our test set using HLP-SSB.

Models of antibody Fv's

Antibody Fv models were constructed by grafting CDR loop templates from non-redundant PDB's onto the native framework (non-CDR residues). H3 was constructed *ab initio* but was not initially optimized. For full details, see Methods. By using the native framework, the H3 stem residues and the heavy-light chain domain orientation are in their native conformations and will not impact our calculations. Thus, only backbone and side chains atoms in the modeled CDR loops surrounding H3 will affect our predictions.

The non-optimized H3 loops of the starting models have average and median backbone RMSDs (Table I, column 8) of 3.0 Å and 2.6 Å, respectively, for short loops and 6.4 Å and 5.8 Å, respectively, for longer loops. The initial loops range from near-native (1.1 Å) to grossly incorrect (9.3 Å). The non-H3 CDR loops that were modeled using canonical loop rules have average and median backbone RMSDs of 1.3 Å and 1.2 Å, respectively, ranging between 0.9 Å and 2.3 Å (Table S1, column 3). These accuracies were expected, as most of the loops fall into canonical clusters that vary ~1 Å when aligned locally⁴⁰. (Our RMSDs are calculated as global alignments; see Methods.) Only a subset of the non-H3 CDR residues interact with the H3 loop so the backbone and side chain accuracies of all residues within 7.5 Å of H3 are shown in Table S1. In the starting structures, average and median backbone RMSDs of the surrounding residues are 0.8 Å and 0.8 Å each and range from 0.3 Å and 1.7. These numbers are smaller than the CDR RMSDs because the surrounding residues include some of the conserved framework whose backbone atoms are in the native configuration. The side chains of surrounding residues, however, have larger deviations with average and median RMSDs of 2.4 Å and 2.5 Å, respectively (Table S1, column 1).

Predicting H3 in models of antibody Fv's

We first predicted H3 loops in the antibody models using a previously described method (HLP) that was not specifically designed to address the challenges of refining loops in homology models. As shown in column 10, Table I, the average and median backbone RMSDs are 2.7 Å and 2.6 Å respectively for shorter loops and 5.1 Å and 5.0 Å, respectively for longer loops. In short loops, the average does not change significantly ($p=0.48$) while in longer loops there is a slight decrease in average RMSD ($p<0.03$). This decrease may be in part due to some of the long loops starting with a very large backbone RMSD. HLP fails to predict any loops in three cases, 1ub6, 1r3j, and 1wt5. As mentioned above, HLP has stringent screening (i.e. eliminating conformations with even moderate steric clashes with the surroundings) which can screen out all backbone samples. In seven cases, the RMSD increases, relative to the initial model, using HLP. Thus by optimizing only the H3 loop

itself, this method sometimes increases the backbone RMSD relative to the starting model. These results are consistent with the well-accepted view that refining loops in homology models is very difficult.

We then predicted H3 loops in the antibody models using our previously published method, HLP-SS28. That method was optimized to refine loop conformations when side chains surrounding the loop are in incorrect conformations, but did not explicitly address inaccuracies in the backbone of the model surrounding the predicted loop. Here, the backbone of the surrounding residues contains relatively small structural errors (0.8 Å average RMSD, Table S1), so we were interested in whether HLP-SS would succeed at predicting native-like H3 loops. In fact, HLP-SS is capable of refining the H3 conformations to lower RMSD in almost all (19/21) cases, in contrast to HLP. Average and median backbone RMSDs are 1.3 Å and 1.0 Å, respectively for shorter loops and 4.2 Å and 4.3 Å, respectively for longer loops (Table I). As we found in previous work, when the loop surroundings have errors, the HLP-SS protocol, which explicitly addresses this challenge, is a clear benefit.

However, for most cases the results with HLP-SS remained much worse than the control calculations using the crystal structures. For example, in the long loop cases, HLP-SS, which predicts a 3.0 Å average backbone RMSD in crystal structures, predicts an average 4.2 Å RMSD in our models ($p=0.03$). Analysis of the results suggests that HLP-SS fails when applied to some of the antibody models primarily due to reduced sampling of near-native conformations. In some cases, such as 1a7q which is discussed in detail below (Figure 3), near-native conformations were not sampled at all. In other cases, near-native conformations were sampled sparsely and/or non-native conformations had lower energies. We attributed these challenges to the inaccurate backbone conformations of the loops surrounding H3.

To address this challenge, we have developed a new protocol called HLP-SSB (see Methods). Results using HLP-SSB show similar accuracies compared to HLP-SS for shorter loops with average and median backbone RMSD of 1.3 Å and 1.2 Å, respectively (Table I). By contrast, in longer loops, accuracy increases significantly ($p=0.03$) over HLP-SS to 3.1 Å average and 2.9 Å median RMSD. Through increased sampling and additional minimization of all non-H3 CDRs before each loop refinement stage, more native-like loops are sampled and refined gradually throughout the hierarchical protocol. In the shorter loops, accuracy degrades slightly (maximum RMSD increase is 0.6 Å) for 8/10 cases which may be expected due to the increased sampling space. In contrast, only 3/11 longer loop cases show an increase in RMSD, two of which show a increase of only 0.1 Å. Five of 11 longer loop cases improve by >1 Å RMSD. Thus HLP-SSB increases accuracy over the starting loop ($p=0.00005$), over HLP ($p=0.007$) and over HLP-SS ($p=0.03$). HLP-SSB also provides similar accuracies for longer loops in both crystal structures (omitting crystal packing and antigens) and our models.

Our new method HLP-SSB requires more computing power; however, most of the additional calculations can be run in parallel. For example, in case 1A7Q which contains an 8-residue loop, the total CPU-hours (i.e. hypothetically run on a single CPU) used by HLP, HLP-SS, and HLP-SSB are 5.1, 13.8, and 31.8 CPU-hours, respectively. However, in this example, the approximate total execution times when run in parallel are 0.7, 1.8, and 2.3 hours, respectively. In practice, when a ~30 node cluster is available, our new method is able to increase sampling dramatically with a relatively small increase in total time.

Somewhat paradoxically, on average, the conformations of the side chains surrounding H3 show no significant improvement using HLP-SSB (this is also the case with HLP-SS) while the backbone accuracy increases only slightly (see Table S1). Average side chain RMSD for

residues within 7.5 Å of H3 for the starting model, HLP-SS, and HLP-SSB are 2.4 Å, 2.6 Å and 2.5 Å, respectively. Evidently it is important to treat these side chains as flexible to allow the loop prediction to succeed, but we suspect that the side chain accuracy will not improve significantly until the backbone conformations are even more accurate. Average backbone RMSD for residues within 7.5 Å of H3 decrease only slightly over the starting model from 0.8 to 0.7 Å. Average backbone RMSD for residues in the non-H3 CDRs also improve only slightly using HLP-SSB over the starting conformation, from 1.3 Å to 1.1 Å. These changes are small. However, the non-H3 CDR results in combination with the much improved H3 results on longer loops show that our new protocol is improving the entire antigen-combining site as a whole.

Selected cases are examined in more detail below.

1a7q

In PDB 1a7q, the sampled regions of the energy landscape change dramatically using HLP-SSB (Figure 3). The modifications of the protocol relative to HLP-SS allow near-native conformations to be sampled; only by optimizing the backbone as well as side chains of surrounding residues using HLP-SSB are we able to make a near-native prediction. With a starting backbone RMSD of 0.4 Å in the surrounding residues of our model, there is little deviation from the crystal structure and yet this is enough to prevent HLP and HLP-SS from sampling near-native conformations. In Figure 4, it is clear that HLP-SSB improves the H3 loop as well as a portion of the surrounding loops in direct contact with H3. Trp52^H, Phe27^H, and Gln1^H do not interact with H3 directly but are in incorrect conformations and dominate the surrounding side chain RMSD shown in Table S1, thus hiding the improvement in the surrounding residues.

2pcp

HLP-SSB predicts an H3 loop conformation with backbone RMSD of 4.8 Å. This failure is not due to sampling; near-native conformations are sampled but have higher energies, funneling down to a basin with roughly 1.5 Å backbone RMSD (Figure 5). There may be several factors contributing to the selection, based on the molecular mechanics force field energy (Methods), of this highly non-native loop conformation, including the presence of an antigen in the crystal structure. However, we believe the key factor is that the x-ray crystal structure was obtained at non-physiological pH, 4.669.

Since our protocols assume titratable residues are in standard protonation states at pH 7 (see Methods), we were interested in whether we could better reproduce the crystal structure H3 loop in 2pcp by using protonation states more applicable to pH 4.6 in our comparative model. At this pH, histidine side chains will almost certainly be doubly protonated. More problematic are Asp and Glu side chains. With somewhat upshifted pK_a's, these could be partially or fully protonated as well. Recognizing that it would not be possible to do so in a prospective application, we calculated the pK_a's of all titratable residues in 2pcp by running MCCE64·65 on the crystal structure. Of the residues in and around H3, Glu39^L was predicted to be protonated (i.e., neutral) and His98^L was predicted to be doubly protonated (i.e., positively charged). By setting these new protonation states and re-running HLP-SSB, our predictions improve from 4.8 Å to 1.7 Å (see Figure 6). Our original predictions may in fact be more representative of an unbound antibody at pH 7 than the crystal structure but further experiments would be necessary to verify that assertion.

Predicting loops in non-antibody comparative models

Finally, we were interested in seeing whether improvements in our method could be beneficial within full comparative models. Because our method optimizes the backbone in

the surrounding residues through a simple minimization, we did not assume that our method would be effective in all complete comparative models. Nonetheless, in five anecdotal test cases derived from the CASP refinement experiments (see Methods, Table II, and Figure S1), we show improvement using HLP-SSB over the initial starting structure in 4/5 cases. In addition, HLP-SSB improves predictions over HLP-SS in 4/5 cases. Unlike the antibody test cases, these initial homology models were generated using programs different than our own, thus suggesting our improvements may be applicable over a variety of different starting conditions. These results are not sufficient to conclude HLP-SSB is generally a better refinement method in full models. However, these anecdotal cases are encouraging and suggest HLP-SSB is a step towards a general method.

Discussion

Our approach to the difficult problem of refining loops in protein homology models has been to interpolate between loop “prediction” in crystal structures and loop refinement in homology models, using a series of a model systems. In a previous paper²⁸, we perturbed crystal structures such that the backbone of one loop as well as all side chains in the protein are initially placed in non-native conformations. Here, we constructed a more challenging model system consisting of partial models of antibodies in which the “scaffold” portion (i.e., the portion other than the CDR loops) retains native backbone conformation, while the CDR loops are predicted using a combination of knowledge-based modeling (H1, H2, L1, L2, and L3) and *ab initio* loop prediction (H3). The H3 loop prediction is challenging because loops surrounding it have incorrect side chain and backbone conformations. However, the “stems” of the H3 loop, i.e., the place where they are attached to the scaffold, are in the native conformation, eliminating for now that contribution to errors in loop prediction. This is a deliberate simplification to guide stepwise improvements to loop prediction when surrounding portions of the protein are incorrect. However, we note that in biomedical applications, new antibodies are frequently developed using existing scaffolds (known not to have immunogenic problems for example), so the modeling protocol developed here may also have practical relevance.

The difficulty of refining loops in comparative models can be attributed to 1) difficulties in sampling and 2) difficulties in accurately calculating the free energies of native and non-native loop conformations. In this work, we focus almost exclusively on the first problem, sampling. Although improvements in our energy function (and better treatment of conformational entropy) would most likely increase accuracy, in our view adequate sampling of native-like conformations must come first. Furthermore, the fact that we can accurately predict 19/21 H3 loops to <1.6 Å RMSD in crystal structures suggests that in comparative models, near-native loops should score well using our energy function. However, we recognize that sampling and scoring are inter-related in that, as sampling increases, the number of non-native conformations increases and the burden on the scoring function to identify native-like conformations increases.

Sampling difficulties can further be broken into two categories; errors in the loop environment can prevent 1) sampling of near-native conformations of the loop backbone and 2) the more subtle sampling of key energetic contacts between the loop in a near-native conformation and the surrounding residues. The first problem occurs when residues surrounding the loop occlude near-native backbone samples. This problem is clear in retrospective studies as no near-native samples are evaluated. This may frequently occur in comparative models where side chains in the surroundings are placed in incorrect conformations due to differences at those positions between the target and template sequences. The second sampling difficulty is more subtle. It occurs when near-native loop backbone conformations are sampled, but errors in the loop environment prevent

energetically important contacts from forming. Failure to form these contacts can lead to near-native loop conformations having high energy (note that this is not, per se, a problem with the energy function).

In this work, we modify previously published energy-based, *ab initio* loop prediction methods to enhance sampling in both of these problem areas. Backbone sampling is enhanced in the initial rounds by 1) accepting more backbone conformations for further refinement, and 2) relaxing our loop screening process through reduction of the overlap factor parameter (see Methods). The more subtle difficulties of forming energetically-important contacts on near-native loops are then addressed, in part, through iterative energy-based optimization of the residues surrounding the predicted loop.

The results on the antibody model systems illustrate the challenges of loop refinement in comparative models and suggest that the new HLP-SSB protocol, which addresses some of the key sampling challenges, is a promising step forward. Consistent with previous results, *ab initio* H3 loop prediction starting from crystal structures of the antibodies generates very accurate results using our previous HLP method, even when the effects of antigen and crystal packing are neglected (0.8 and 1.6 Å median global backbone RMSD for shorter and longer loops respectively). Applying HLP to the antibody model systems, median backbone RMSD is reduced in only 11/21 cases over the starting models and leads to worse results in some cases. Additionally, the method fails to generate any loops due to steric clashes in 3/21 cases. The enhancements we have made to the loop prediction methods, resulting in the HLP-SSB protocol, give much more satisfactory results, improving the initial H3 loop conformations in 20/21 cases, to an average backbone RMSD of 1.3 Å in shorter loops and 3.1 Å in longer loops. We also show that the relatively poor performance for test case 2pcp, with an HLP-SSB predicted H3 backbone RMSD of 4.8 Å, is likely due in part to the non-physiological pH of the crystal structure, 4.6.

Because there is little change in the RMSD of the surrounding residues and CDRs, it is difficult to assess how the modeling of L1, L2, L3, H1 and H2 affect our *ab initio* modeling of H3. In some cases (see Table S2) we are unable to model a CDR using a canonical template. For example, in an extreme case, 1iqd, we could not model L1, L2, and L3 using canonical loop templates and this may have affected our HLP-SSB prediction (6.3 Å) which is worse than our HLP-SSB control prediction in the crystal structure (4.7 Å). However, the accuracy of the non-H3 CDR loops are roughly average for the whole test set. While modeling errors in the non-H3 CDRs undoubtedly affect our H3 predictions, they are representative of errors commonly found in full comparative model refinement scenarios (which is why we chose antibodies as a test system).

Additional sampling difficulties will occur as we move towards refining loops in full comparative models. As discussed above, we currently ignore sampling problems due to errors in the loop stem residues. In addition, in the case of antibodies, we currently ignore sampling problems due to errors in the domain orientation of the variable heavy (V_H) and variable light (V_L) chains; the H3 loop can form contacts with both domains. In preliminary results for this work, we found that the relative orientations of the heavy and light chains can affect our H3 predictions (data not shown). Because of this, choosing an appropriate scaffold template that has a similar V_H - V_L orientation as the target's domain orientation (or accurately predicting *de novo* the target's domain orientation) will be imperative when modeling antibodies that are truly novel. We have developed such a predictive method which will be described in a separate paper.

In some ways, the collection of antibody structures is not a perfect model system. For example, our results suggest, but do not prove, that longer H3 loops (8 residues or greater)

are somewhat flexible and can adopt different conformations due to ligand binding⁷⁰⁻⁷² or crystal packing⁷³⁻⁷⁵. However, we believe the benefits outweigh these difficulties. By carrying out a large number of control calculations on the native crystal structures, we were able to deconvolute these problems to some degree and most importantly draw logical and measured conclusions about our improvements.

HLP-SSB does not represent a complete solution to the problem of refining loops in homology models. One major limitation is that it does not explicitly address errors in the loop stems, and it is unlikely to succeed when there are very large errors in the backbone of surrounding portions of the protein. However, we presented anecdotal results suggesting that the method can be useful in at least some fraction of comparative models. We are hopeful that these successes can be generalized as we make further steps towards developing a loop prediction method which increasingly optimizes more of its surroundings.

Extending predictions to longer H3 loops will be required for many antibody applications. However, in this work, we chose to focus on short to moderate length loop (5-9 residues) as a starting point. Predicting loops >9 residues in length is very difficult in comparative models. As we see here, longer loops may be more sensitive to crystal packing and antigen effects and are generally more flexible and prone to sampling failures. The latter issue is particularly a problem as the number of degrees of freedom sampled in the surroundings of the loop is increased, as we have done here. That said, there have been advances in predicting longer loop lengths within crystal structures using the Protein Local Optimization Program^{16-19,28}, as well as with other methods^{20-76,77}. We believe accurately predicting long loops in comparative models is possible, although it may require substantial computational expense, at least when using all-atom molecular mechanics energy functions, as we have done here. In the future, we aim to combine lessons learned in predicting long loops with lessons from the present work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Molecular graphics images were produced using the UCSF Chimera⁶⁷ package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). Our test set of antibody Fv models can be downloaded here: <http://jacobsonlab.org>.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28(1):235–242. [PubMed: 10592235]
2. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M. The Universal Protein Resource (UniProt). *Nucleic Acids Research.* 2005; 33:D154. [PubMed: 15608167]
3. Sali A. 100,000 protein structures for the biologist. *Nat Struct Biol.* 1998; 5(12):1029–1032. [PubMed: 9846869]
4. Honma T, Hayashi K, Aoyama T, Hashimoto N, Machida T, Fukasawa K, Iwama T, Ikeura C, Ikuta M, Suzuki-Takahashi I, Iwasawa Y, Hayama T, Nishimura S, Morishima H. Structure-based generation of a new class of potent Cdk4 inhibitors: new de novo design strategy and library design. *J Med Chem.* 2001; 44(26):4615–4627. [PubMed: 11741479]
5. Schapira M, Raaka BM, Samuels HH, Abagyan R. In silico discovery of novel retinoic acid receptor agonist structures. *BMC Struct Biol.* 2001; 1:1. [PubMed: 11405897]

6. Enyedy IJ, Ling Y, Nacro K, Tomita Y, Wu X, Cao Y, Guo R, Li B, Zhu X, Huang Y, Long YQ, Roller PP, Yang D, Wang S. Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem.* 2001; 44(25):4313–4324. [PubMed: 11728179]
7. Enyedy IJ, Lee SL, Kuo AH, Dickson RB, Lin CY, Wang S. Structure-based approach for the discovery of bis-benzamidines as novel inhibitors of matriptase. *J Med Chem.* 2001; 44(9):1349–1355. [PubMed: 11311057]
8. Song L, Kalyanaraman C, Fedorov AA, Fedorov EV, Glasner ME, Brown S, Imker HJ, Babbitt PC, Almo SC, Jacobson MP, Gerlt JA. Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat Chem Biol.* 2007; 3(8):486–491. [PubMed: 17603539]
9. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D. Protein-protein docking predictions for the CAPRI experiment. *Proteins.* 2003; 52(1):118–122. [PubMed: 12784377]
10. Sivasubramanian A, Maynard JA, Gray JJ. Modeling the structure of mAb 14B7 bound to the anthrax protective antigen. *Proteins.* 2008; 70(1):218–230. [PubMed: 17671962]
11. Sivasubramanian A, Chao G, Pressler HM, Wittrup KD, Gray JJ. Structural Model of the mAb 806-EGFR Complex Using Computational Docking followed by Computational and Experimental Mutagenesis. *Structure.* 2006; 14(3):401–414. [PubMed: 16531225]
12. McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J Med Chem.* 2003; 46(14):2895–2907. [PubMed: 12825931]
13. Moult J, Fidelis K, Kryshtafovych A, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction-Round VII. *Proteins.* 2007; 69(Suppl 8):3–9. [PubMed: 17918729]
14. Kopp J, Bordoli L, Battey JN, Kiefer F, Schwede T. Assessment of CASP7 predictions for template-based modeling targets. *Proteins.* 2007; 69(Suppl 8):38–56. [PubMed: 17894352]
15. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 1986; 5(4):823–826. [PubMed: 3709526]
16. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. *Proteins.* 2004; 55(2):351–367. [PubMed: 15048827]
17. Moult J, James MNG. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins: Structure, Function and Genetics.* 1986; 1:146–163.
18. van Vlijmen HWT, Karplus M. PDB-based protein loop prediction: parameters for selection and methods for optimization. *Journal of Molecular Biology.* 1997; 267(4):975–1001. [PubMed: 9135125]
19. Zhu K, Pincus DL, Zhao S, Friesner RA. Long loop prediction using the protein local optimization program. *Proteins.* 2006; 65(2):438–452. [PubMed: 16927380]
20. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Sci.* 2000; 9(9):1753. [PubMed: 11045621]
21. Deane CM, Blundell TL. CODA: a combined algorithm for predicting the structurally variable regions of protein models. *Protein Sci.* 2001; 10(3):599–612. [PubMed: 11344328]
22. Groban ES, Narayanan A, Jacobson MP. Conformational Changes in Protein Loops and Helices Induced by Post-Translational Phosphorylation. *PLoS Computational Biology.* 2006; 2(4)
23. Wang C, Bradley P, Baker D. Protein–Protein Docking with Backbone Flexibility. *Journal of Molecular Biology.* 2007; 373(2):503–519. [PubMed: 17825317]
24. Spassov VZ, Flook PK, Yan L. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. *Protein Engineering, Design and Selection.* 2008; 21(2):91–100.
25. Cui M, Mezei M, Osman R. Prediction of protein loop structures using a local move Monte Carlo approach and a grid-based force field. *Protein Engineering, Design and Selection.* 2008; 21(12): 729–735.
26. Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC. Structural insights into the evolution of an antibody combining site. *Science.* 1997; 276(5319):1665–1669. [PubMed: 9180069]
27. Monnigmann M, Floudas CA. Protein Loop Structure Prediction With Flexible Stem Geometries. *Proteins: Structure, Function, and Bioinformatics.* 2005; 61:748–762.

28. Sellers BD, Zhu K, Zhao S, Friesner RA, Jacobson MP. Toward better refinement of comparative models: Predicting loops in inexact environments. *Proteins*. 2008; 72(3):959–971. [PubMed: 18300241]
29. Zheng L, Manetsch R, Woggon WD, Baumann U, Reymond JL. Mechanistic study of proton transfer and hysteresis in catalytic antibody 16E7 by site-directed mutagenesis and homology modeling. *Bioorganic & Medicinal Chemistry*. 2005; 13(4):1021–1029. [PubMed: 15670909]
30. Renault L, Essono S, Juin M, Boquet D, Grassi J, Bourne Y, Marchot P. (28) Structural insights into AChE inhibition by monoclonal antibodies. *Chemico-Biological Interactions*. 2005; 157:397–400. [PubMed: 16429535]
31. Staelens S, Desmet J, Ngo TH, Vauterin S, Pareyn I, Barbeaux P, Van Rompaey I, Stassen JM, Deckmyn H, Vanhoorelbeke K. Humanization by variable domain resurfacing and grafting on a human IgG4, using a new approach for determination of non-human like surface accessible framework residues based on homology modelling of variable domains. *Molecular Immunology*. 2006; 43(8):1243–1257. [PubMed: 16118019]
32. Li B, Wang H, Zhang D, Qian W, Hou S, Shi S, Zhao L, Kou G, Cao Z, Dai J. Construction and characterization of a high-affinity humanized SM5-1 monoclonal antibody. *Biochemical and Biophysical Research Communications*. 2007; 357(4):951–956. [PubMed: 17451647]
33. McKinney BA, Kallewaard NL, Crowe JE Jr, Meiler J. Using the natural evolution of a rotavirus-specific human monoclonal antibody to predict the complex topography of a viral antigenic site. *Immunome Res*. 2007; 3(8)
34. Hou S, Li B, Wang L, Qian W, Zhang D, Hong X, Wang H, Guo Y. Humanization of an Anti-CD34 Monoclonal Antibody by Complementarity-determining Region Grafting Based on Computer-assisted Molecular Modelling. *Journal of Biochemistry*. 2008; 144(1):115. [PubMed: 18424812]
35. Sivasubramanian A, Sircar A, Chaudhury S, Gray JJ. Toward high-resolution homology modeling of antibody F (v) regions and application to antibody-antigen docking. *Proteins*. 2008; 74(2):497. [PubMed: 19062174]
36. Marcatili P, Rosi A, Tramontano A. PIGS: automatic prediction of antibody structures. *Bioinformatics*. 2008; 24(17):1953. [PubMed: 18641403]
37. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*. 1987; 196(4):901–917. [PubMed: 3681981]
38. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol*. 1997; 273(4):927–948. [PubMed: 9367782]
39. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al. Conformations of immunoglobulin hypervariable regions. *Nature*. 1989; 342(6252):877–883. [PubMed: 2687698]
40. Martin AC, Thornton JM. Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. *J Mol Biol*. 1996; 263(5):800–815. [PubMed: 8947577]
41. Shirai H, Kidera A, Nakamura H. Structural classification of CDR-H3 in antibodies. *FEBS Letters*. 1996; 399(1-2):1–8. [PubMed: 8980108]
42. Oliva B, Bates PA, Querol E, Avilés FX, Sternberg MJE. Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction. *Journal of Molecular Biology*. 1998; 279(5):1193–1210. [PubMed: 9642095]
43. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Antibody structure, prediction and redesign. *Biophysical Chemistry*. 1997; 68(1-3):9–16. [PubMed: 9468606]
44. Koliashnikov OV, Kiral MO, Grigorenko VG, Egorov AM. Antibody cdr h3 modeling rules: extension for the case of absence of arg h94 and asp h101. *JOURNAL OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY*. 2006; 4(2):415. [PubMed: 16819792]
45. Shirai H, Kidera A, Nakamura H. H3-rules: identification of CDR-H3 structures in antibodies. *FEBS Letters*. 1999; 455(1-2):188–197. [PubMed: 10428499]
46. Kuroda D, Shirai H, Kobori M, Nakamura H. Structural classification of CDR-H3 revisited: A lesson in antibody modeling. *Proteins*. 2008

47. Cardozo T, Totrov M, Abagyan R. Homology Modeling by the ICM Method. *Proteins: Struct Funct Genet.* 1995; 23:403–411. [PubMed: 8710833]
48. Fine RM, Wang H, Shenkin PS, Yarmush DL, Levinthal C. Predicting Antibody Hypervariable Loop Conformations 11: Minimization and Molecular Dynamics Studies of MCPC603 From Many Randomly Generated Loop Conformations. *Proteins.* 1986; 1:342–362. [PubMed: 3449860]
49. Bruccoleri RE, Haber E, Novotný J. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature.* 1988; 335(6190):564–568. [PubMed: 3419534]
50. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry.* 1983; 4(2):187–217.
51. Whitelegg NRJ, Rees AR. WAM: an improved algorithm for modelling antibodies on the WEB. *Protein Eng.* 2000; 13(12):819–824. [PubMed: 11239080]
52. Rosenbach D, Rosenfeld R. Simultaneous modeling of multiple loops in proteins. *Protein Science.* 1995; 4(3):496. [PubMed: 7540907]
53. Mehler EL, Hassan SA, Kortagere S, Weinstein H. Ab Initio Computational Modeling of Loops in G-Protein-Coupled Receptors: Lessons from the Crystal Structure of Rhodopsin. *Proteins: Structure, Function, and Bioinformatics.* 2006; 64:673–690.
54. Rapp CS, Friesner RA. Prediction of loop geometries using a generalized born model of solvation effects. *Proteins Structure Function and Genetics.* 1999; 35(2):173–183.
55. Wang G, Dunbrack RL Jr. PISCES: a protein sequence culling server. *Bioinformatics.* 2003; 19(12):1589–1591. [PubMed: 12912846]
56. Wang G, Dunbrack RL Jr. PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res.* 2005; 33:W94–98. (Web Server issue). [PubMed: 15980589]
57. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22(22):4673–4680. [PubMed: 7984417]
58. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Q ReV Biophys.* 1993; 26:49. [PubMed: 8210313]
59. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B.* 2001; 105(28):6474–6487.
60. Jacobson MP, Kaminski GA, Friesner RA, Rapp CS. Force field validation using protein side chain prediction. *J Phys Chem B.* 2002; 106(44):11673–11680.
61. Ghosh, A.; Rapp, CS.; Friesner, RA. Generalized Born model based on a surface integral formulation. 1998. p. 10983-10990.
62. Gallicchio, E.; Zhang, LY.; Levy, RM. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. 2002. p. 517-529.
63. Zhu K, Shirts MR, Friesner RA, Jacobson MP. Multiscale Optimization of a Truncated Newton Minimization Algorithm and Application to Proteins and Protein-Ligand Complexes. *J Chem Theory Comput.* 2007; 3(2):640–648.
64. Alexov EG, Gunner MR. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophysical Journal.* 1997; 72(5):2075–2093. [PubMed: 9129810]
65. Georgescu RE, Alexov EG, Gunner MR. Combining Conformational Flexibility and Continuum Electrostatics for Calculating pKas in Proteins. *Biophysical Journal.* 2002; 83(4):1731–1748. [PubMed: 12324397]
66. Kenyon V, Chorny I, Carvajal WJ, Holman TR, Jacobson MP. Novel human lipoxxygenase inhibitors discovered using virtual screening with homology models. *J Med Chem.* 2006; 49(4): 1356–1363. [PubMed: 16480270]
67. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry.* 2004; 25(13):1605–1612. [PubMed: 15264254]

68. <http://predictioncenter.gc.ucdavis.edu/caspR/>
69. Lim K, Owens SM, Arnold L, Sacchettini JC, Linthicum DS. Crystal Structure of Monoclonal 6B5 Fab Complexed with Phencyclidine. *J Biol Chem*. 1998; 273(44):28576–28582. [PubMed: 9786848]
70. James LC, Roversi P, Tawfik DS. Antibody Multispecificity Mediated by Conformational Diversity. *Science*. 2003; 299(5611):1362–1367. [PubMed: 12610298]
71. Mundorff EC, Hanson MA, Varvak A, Ulrich H, Schultz PG, Stevens RC. Conformational Effects in Biological Catalysis: An Antibody-Catalyzed Oxy-Cope Rearrangement. *Biochemistry*. 2000; 39(4):627–632. [PubMed: 10651626]
72. Charbonnier J, Carpenter E, Gigant B, Golinelli-Pimpaneau B, Eshhar Z, Green BS, Knossow M. Crystal structure of the complex of a catalytic antibody Fab fragment with a transition state analog: structural similarities in esterase-like catalytic antibodies. *Proceedings of the National Academy of Sciences*. 1995; 92(25):11721–11725.
73. Jennifer A, Kowalski KLJWK. NMR solution structure of the isolated Apo Pin1 WW domain: Comparison to the x-ray crystal structures of Pin1. *Biopolymers*. 2002; 63(2):111–121. [PubMed: 11786999]
74. Baldwin ET, Weber IT, Charles RS, Xuan J, Appella E, Yamada M, Matsushima K, Edwards BFP, Clore GM, Gronenborn AM. Crystal structure of interleukin 8: symbiosis of NMR and crystallography. *Proceedings of the National Academy of Sciences*. 1991; 88(2):502–506.
75. Gudrun Lange-Savage HBALK-HBAPJKJSMFRH. Structure of HOE/BAY 793 Complexed to Human Immunodeficiency Virus (HIV-1) Protease in Two Different Crystal Forms Structure/Function Relationship and Influence of Crystal Packing. *European Journal of Biochemistry*. 1997; 248(2):313–322. [PubMed: 9346283]
76. Spassov VZ, Flook PK, Yan L. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. *Protein Engineering Design and Selection*. 2008; 21(2):91.
77. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins*. 2004; 55(3):656–677. [PubMed: 15103629]

Figure 1.
Flow chart depicting the process of creating the initial antibody Fv model.

Figure 2.

Differences between the previously published HLP-SS (left) and the new method HLP-SSB refinement protocol (right). Boxes with “Loop predict” depict several parallel executions (the number of which is denoted to the right of the box) of the loop prediction command in the Protein Local Optimization Program. Each loop prediction produces a set of all-atom, sampled loop conformations ranked by MM-GBSA energy. New in HLP-SS are three additional initial stages with reduced overlap factors in addition to two energy minimization stages of the backbone and side chains atoms in surrounding residues (the non-H3 CDRs in the case of antibodies). See Methods for details.

Figure 3.

Energy versus backbone RMSD for all loop conformations with energies <35 kcal/mol greater than the conformation with the lowest predicted energy, using three protocols for the 8 residue H3 loop in the antibody with PDB code 1a7q. Each dot represents a loop conformation at a local energy minimum generated during our loop prediction protocol. Top: HLP, Middle: HLP-SS, Bottom: HLP-SSB. Note that the absolute energies are not comparable between graphs due to different numbers of atoms optimized in the three protocols.

Figure 4.

In PDB 1a7q, comparison of predictions using HLP, HLP-SS, and HLP-SSB to the crystal structure (CPK colors). The predicted H3 loop is gold in all three panels. Top: HLP prediction of H3 (no optimization of surrounding residues). Middle: HLP-SS prediction (surrounding side chains optimized). Bottom: HLP-SSB prediction (surrounding side chains and backbone energy optimized). Surrounding residues that are not refined are purple while surrounding residues that are refined are light blue. Residues Trp52^H, Phe27^H and Gln1^H are the largest contributors to the RMSD of surrounding side chains using HLP-SSB (see text).

Figure 5.

Energy vs backbone RMSD plots for conformations <30 kcal/mol greater than the lowest energy conformation of H3 using HLP-SSB in 2pcp at two pH values. Each dot represents a loop conformation at a local energy minimum generated during our loop prediction protocol. Top: Conformations generated with standard protonation states at pH 7.0. Bottom: Conformations generated with predicted protonation states at pH 4.6 (the pH of the associated crystal structure).

Figure 6.

Comparison of H3 predictions using HLP-SSB at different pH values in 2pcp. Gray: crystal structure. Orange: prediction of H3 (4.8 Å backbone RMSD) using HLP-SSB with standard protonation states at pH 7. Blue: prediction of H3 (1.7 Å backbone RMSD) using HLP-SSB with predicted protonated forms of Glu34^L and His93^L consistent with the pH of the crystal structure, 4.6.

Table 1

Comparison of predicted H3 loop N-C α -C-O RMSD in Å in crystal structures and models of 21 antibody variable fragments using various protocols. Crystal packing: '+' indicates other chains in the crystal unit cell are explicitly included in the prediction, as a control (see Methods). Antigen: '+' indicates the binding partner (antigen), if it exists in the crystal structure (column 2), is included. Method: the protocol used, either HLP16, HLP-SS28, or HLP-SSB as introduced here. Column 1: PDB code. Column 2: 'P' indicates the antibody has a protein antigen present in the crystal structure. 'S' indicates a small molecule antigen. '-' indicates no antigen present in crystal structure. Column 3: the H3 loop length. The remaining columns show backbone RMSD for the predicted loops using various protocols. A '-' in column 5 indicates no antigen is present in the crystal structure so the RMSD is not included.

Crystal Structure Controls													Comparative Models			
Crystal packing			+	-	-	-	-	-	-	-	-	-	-	-		
Antigen			+	+	-	-	-	-	-	-	-	-	-	-		
Method	HLP	HLP	HLP	HLP	HLP	HLP-SS	HLP-SSB	Initial	HLP	HLP-SS	HLP-SSB	HLP-SSB	HLP-SSB	HLP-SSB		
PDB code	Antigen	H3 Length														
1cr9	-	5	0.5	-	0.5	1.0	0.6	3.7	4.2	1.0	1.2	1.0	1.2	1.2		
1mex	S	5	0.3	0.4	0.5	0.9	1.5	2.4	0.6	0.5	1.1	0.5	1.1	1.1		
1ngz	-	5	0.3	-	0.3	0.3	0.3	2.4	1.2	1.2	1.3	1.2	1.3	1.3		
1uac	P	5	0.3	0.3	1.5	0.7	0.7	1.0	0.6	0.8	0.9	0.8	0.9	0.9		
1ub6	-	6	0.7	-	0.6	0.6	0.5	5.6	*	1.3	1.3	1.3	1.3	1.3		
1flr	S	7	0.4	0.4	1.3	0.6	0.9	1.1	1.7	1.0	1.6	1.0	1.6	1.6		
1kcv	-	7	0.6	-	1.0	0.6	2.0	5.4	2.6	0.7	0.8	0.7	0.8	0.8		
1mju	-	7	0.3	-	0.2	0.5	0.5	2.8	4.4	1.0	1.2	1.0	1.2	1.2		
1yqv	P	7	0.6	0.6	1.0	1.7	1.9	2.3	4.9	1.8	2.0	1.8	2.0	2.0		
2cgr	S	7	1.5	3.7	1.2	1.2	0.7	3.2	4.3	4.2	2.1	4.2	2.1	2.1		
Avg			0.5	1.1	0.8	0.8	1.0	3.0	2.7	1.3	1.3	1.3	1.3	1.3		
Median			0.5	0.4	0.8	0.7	0.7	2.6	2.6	1.0	1.2	1.0	1.2	1.2		
1a7q	-	8	0.8	-	1.1	1.7	0.8	6.9	4.7	4.3	0.5	4.3	0.5	0.5		
1uj3	P	8	0.2	0.2	0.4	1.6	1.9	6.7	4.7	1.9	1.0	4.7	1.9	1.0		
1uz8	S	8	1.1	1.3	0.3	2.3	4.5	5.8	3.2	3.2	2.0	3.2	3.2	2.0		

Crystal Structure Controls										Comparative Models			
Crystal packing													
Antigen													
Method													
PDB code	Antigen	H3 Length	HLP	HLP	HLP	HLP-SS	HLP-SSB	Initial	HLP	HLP-SS	HLP-SSB	HLP-SSB	HLP-SSB
2pcp	S	8	0.9	0.9	1.3	2.2	2.0	9.3	6.2	4.7	4.8 (1.7) [‡]		
1ct8	S	9	0.4	0.4	4.6	6.3	4.0	5.3	5.9	3.3	2.9		
liqd	P	9	0.8	0.8	5.9	5.1	4.7	8.0	5.0	5.2	6.3		
1mqk	-	9	2.8	-	2.8	2.8	2.9	3.5	6.2	3.0	2.0		
1r3j	P	9	0.3	2.9	5.3	4.2	4.2	5.5	*	7.7	3.6		
lum5	S	9	1.0	0.6	1.6	4.6	6.8	5.5	5.4	5.1	4.0		
1wt5	-	9	0.3	-	8.0	0.7	1.5	5.3	*	2.6	2.7		
2fbj	-	9	6.3	-	1.1	1.1	1.0	8.3	4.5	5.0	4.5		
Avg			1.4	1.0	2.9	3.0	3.1	6.4	5.1	4.2	3.1 (2.8) [‡]		
Median			0.8	0.8	1.6	2.3	2.9	5.8	5.0	4.3	2.9 (2.7) [‡]		

* No RMSD is listed because all generated backbone conformations were screened out due to steric clashes.

[‡] RMSD in parenthesis is achieved when using non-standard protonation states for certain side chains in 2pcp, to reflect the pH of the crystal structure, 4.6.

Table II

Comparison of predicted loop N-Cα-C-O RMSD in Å in full homology models using HLP16, HLP-SS28, and HLP-SSB. Column 1: the CASP model target designation (TMR: CASPR; TR288 and TR308: CASP7; TR289: CASP8). Column 2: PDB identifier of the native crystal structure. Columns 3 and 4: the residue identifiers as specified in the PDB entry for the start and end of the loop. Columns 5-8: The loop backbone RMSD of the starting model, HLP prediction, HLP-SS prediction, HLP-SSB prediction, respectively. The starting model was aligned to the crystal structure using Chimera. See Methods for details.

Model	Native PDB	Loop start	Loop end	Start	HLP	HLPSS	HLPSSB
TMR3	1vla	A:96	A:103	3.2	1.7	2.0	2.9
TMR5	1vlg	A:94	A:103	3.9	8.9	5.3	2.3
TR288	2gzv	A:26	A:34	2.1	3.1	3.8	3.3
TR308	2h57	A:39	A:46	4.7	4.4	4.9	3.9
TR389	2vsw	A:111	A:119	2.5	1.6	2.0	1.5