Supplement to

Structure Refinement and Computational Validation Protocol for Proteins with Large Variable Regions Applied to Model HIV Envelope Spike in its CD4 and 17b Bound State

Muhibur Rasheed, Radhakrishna Bettadapura, Chandrajit Bajaj

Model Evaluation Criteria (Scoring Function)

Evaluating Local Stereochemistry, Tertiary Folds, and Contacts We utilize a suite of tools consisting of Verify3D (Luthy et al., 1992), PROCHECK (MacArthur et al., 1993), ERRAT (Colovos and Yeates, 1993), ProSA (Sippl, 1993), and MolProbity (Davis et al., 2007). Note that it is possible for a structure to get perfect scores in local stereochemistry checkers like PROCHECK but have poor tertiary structural folds and get bad scores in global correlation tools like Verify3D, ERRAT, ModEval etc. On the other hand, there are high resolution structures deposited in the PDB database which contain local errors and have poor PROCHECK score. So, we consider a model as high quality only if it scores well in each of the validation tools. Note that, PDB evaluation suite ADIT, Modeval (Shen, 2006) and Qmean z-score (Benkert et al., 2011) were used for independent validation of the final model, but not used during the scoring and search.

Evaluating Quaternary Contacts/Protein-Protein Interactions When the protein is in complex, then the quaternary structure quality, i.e. the quality of the interface between the proteins must also be considered. We use five scoring terms which can be efficiently computed. The first two terms, clash and severe clash, compute the number of atoms of one protein whose center lies, respectively, too close and inside the VDW volume of any atom of the other (See Figure S1A for an example). Good crystal structures typically have no severe clashes, and fewer than 10 clashes. The third scoring term, interface area, is computed as the part of the molecular surface of one protein which is within 2Å from any point on the molecular surface of another protein (Figure S1B). Interface area can vary a lot depending on the type of the proteins involved and approriate thresholds must be calibrated (see calibration in the Results section and Figure S2). The last two terms are statistical residue-residue contact scores, computed based on contact potentials for each residue-residue types reported by Glaser et al. (Glaser et al., 2001), where positive and negative potentials indicate, respectively, higher and lower probability of the finding such contact on an interface. These scores were applied in (Chowdhury et al., 2013) and successfully ranked native interfaces, especially for antibody-antigen complexes.

Evaluating Quality of Fitting We used two scoring terms, ETR and MIS, to evaluate the quality of fitting of a model to an EM-map. The external-total ratio (ETR) (Pettersen et al., 2004) is defined as $-N_{out}/N$ where N_{out} is the total number of atoms of the model which lies outside a given isocontour of the density map, and N is the total number of atoms. Lower ETR indicates better fitting (see Figure S1C). The mutual information score (MIS) (Shatsky et al., 2008; Vasishtan and Topf, 2011), is given by $MIS = \sum_{x \in B} \sum_{y \in A} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$ where $p(x)$ and $p(y)$ are the percentage of voxels in the volumetric representation of the model B and the density map A that take on intensities equal to x and y respectively and $p(x,y)$ is the percentage of voxels in B with intensity x that are aligned with voxels in A with intensity y. Essentially, MIS is maximized when the model has larger overlap with the map (see Figure S1D).

Overall Score We have primarily two different types of scoring protocols for quality assessment. The first, $s_{internal}$, is a set of scores to evaluate the stereochemistry and tertiary folds using a consensus of several state of the art structure assessment tools. The send set of scores, developed by us, evaluates the ternary interactions and fitting to a density map. The overall score for them are defined as-

$s_{internal} = g_{(\phi-\psi)} + g_{(all)} + z_{Verify3D} + z_{ProSA} + z_{MolProbity} + z_{ERRAT}$

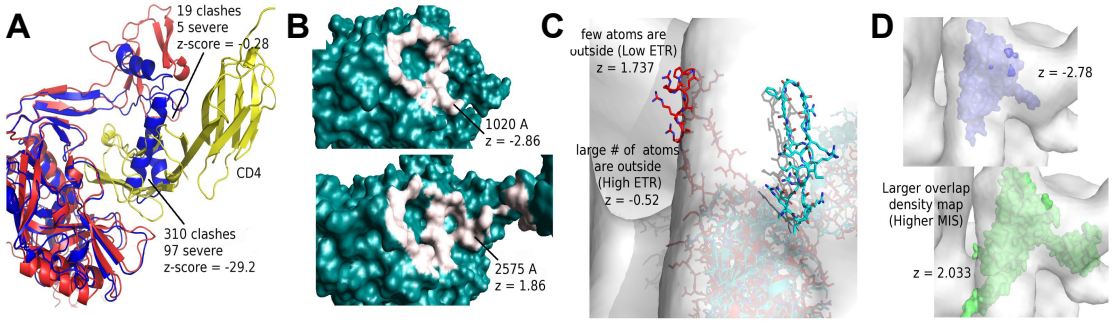
and

$s_{external} = z_{cl} + z_{RCP} + z_{RCN} + z_{IA} + z_{ETR} + z_{MIS}$

where $g_{(\phi-\psi)}$, $g_{(all)}$, $z_{Verify3D}$, z_{ProSA}, $z_{MolProbity}$, z_{ERRAT}, z_{cl}, z_{RCP}, z_{RCN}, z_{IA}, z_{ETR}, z_{MIS} are PROCHECK g-factors for $\phi - \psi$ angles and for all angles, z-score of Verify3D, ProSA, MolProbity and ERRAT's composite scores, and z-scores of clash, positive residue contact, negative residue contact, interface area, ETR and MIS scores respectively. The z-score for a term X is defined as $(s_X - \mu_X)/\sigma_X$ where s_X is the raw score of the model, and μ_X and σ_X are the mean and standard deviation of the raw scores over all structures, or a benchmark/control set of structures. See the next section for details on the calibration of the composite score for modeling gp120.

The sum of the z-scores intuitively means that a model have to be better than average in all aspects to be considered as accurate. Notice that some of the scoring terms are complementary to each other and

provide check and balance. Poor models may get better score in some terms, but not all. For example, a model of gp120 which penetrates CD4 when placed at the correct orientation (derived based on 1GC1), will get high scores for interface area and positive residue contacts, but will get penalized by clashes and negative residue contacts.



E

	Scoring Term	Procheck G-factor (phi-psi) *	Procheck G-factor (all) *	Molprobity clash z-score *	Verify3D compositeZ-score *	Prosall composite z score *	ERRAT Score	Overall Score and Assessment	Resolution
Crystal Structures	1G9M	-2.28	-3.02	-4.75	-0.32	-0.70	-0.49	✓	-11.56	2.2
	1G9N	-3.38	-4.20	-5.73	-0.48	-0.70	-1.03	!	-15.52	2.9
	1GC1	-3.23	-4.61	-4.66	-0.96	-0.95	-0.19	!	-14.60	2.5
	1RZJ	-2.24	-2.90	-4.90	-0.16	-0.70	-0.95	✓	-11.85	2.2
	1YYL	-1.61	-2.13	-2.54	-0.96	-0.74	-0.05	★	-8.03	2.75
	1YYM	-0.94	-1.42	-1.88	-0.64	-0.74	0.67	★	-4.95	2.2
	2I5Y	-1.18	-1.60	-2.02	-0.96	-0.79	1.12	★	-5.43	2.2
	2I60	-1.34	-2.01	-2.73	-0.80	-0.79	0.12	★	-7.55	2.4
	2NXZ	-1.06	-1.30	0.10	-0.32	-0.91	1.12	★	-2.37	2.04
	2NY0	-1.06	-1.60	-0.16	-0.16	-0.87	1.37	★	-2.48	2.2
	2NY3	-0.90	-1.30	0.61	-0.32	-0.91	0.91	★	-1.91	2
	2NY5	-0.98	-1.48	-0.55	-0.48	-0.79	1.50	★	-2.78	2.5
	2NY7	-1.53	-2.07	-0.94	-1.12	-0.99	1.01	★	-5.64	2.3
	2QAD	-2.71	-3.67	-2.47	-0.96	-0.89	-0.95	✓	-11.65	3.3
	3HI1	-2.28	-3.13	-2.93	-1.12	-0.79	-1.31	✓	-11.56	2.9
	3IDY	-2.32	-3.25	-1.93	-1.12	-0.95	-0.72	✓	-10.29	3.2
	3JWD	-2.12	-2.31	-2.07	-0.96	-0.79	-0.88	✓	-9.12	2.61
	3JWO	-2.32	-2.54	-2.81	-0.96	-0.74	-0.82	✓	-10.19	3.51
	3RJQ	-1.65	-2.72	-5.79	-0.80	-0.74	-1.47	!	-13.17	2.6
	3SE8	-0.75	-1.06	-0.44	-0.48	-0.50	1.03	★	-2.20	1.89
Control	Min	-3.38	-4.61	-5.79	-1.12	-0.99	-1.47		-15.52	
	Max	-0.75	-1.06	0.61	-0.16	-0.50	1.50		-1.91	
	Avg	-1.79	-2.42	-2.55	-0.70	-0.79	0.00		-8.14	
	StdDev	0.79	1.01	1.98	0.33	0.12	1.00		4.43	
PSVS ranges	Avg (high-res)	0.00	0.00	0.00	0.00	0.00				
	Avg (med-res)	-0.18	-0.53	-1.39	-0.43	-0.17				
	Avg(low-res)	-1.57	-2.46	-2.45	-0.59	-0.17				
	StdDev (high-res)	1	1	1	1	1				
	StdDev (med-res)	1.14	1.19	1.54	1.28	1.03				
	StdDev (low-res)	1.21	1.42	3.1	1.28	0.99				

Correlation of overall score and the resolution is -0.5927 which corresponds to a tail probablity of 0.006175

F

# of clashes	# severe clashes	rescont pos	rescont neg	interface area	ETR	MIS	Overall Score and Assessment	Co-crystallized with CD4 and/or 17b?
0.299	0.543	-0.32	1.626	-1.35	4.097	-2.03	✓	2.87	Y
0.444	0.394	-0.21	2.372	-1.14	4.202	-3.54	✓	2.52	Y
0.492	0.543	-0.8	2.372	-1.63	4.166	-2.78	✓	2.36	Y
0.299	0.245	-0.34	2.372	-1.44	4.132	-3.3	✓	1.96	Y
0.299	0.394	-0.52	2.506	-0.89	4.131	-2.71	★	3.21	Y
0.347	0.394	-0.55	0.753	-1.02	4.166	-2.62	✓	1.47	Y
0.203	0.097	-0.45	3.252	-0.78	4.202	-2.93	★	3.59	Y
0.347	0.097	-0.52	2.372	-1.12	4.131	-0.34	★	4.96	Y
0.54	0.097	-0.48	2.372	-1.25	3.854	-2.7	✓	2.43	Y
0.251	0.245	-0.55	2.372	-1.3	3.854	-2.8	✓	2.07	Y
0.588	0.692	-0.33	2.372	-1.26	3.993	-3.26	✓	2.80	Y
0.299	0.097	-0.39	2.372	-1.38	3.922	-2.26	✓	2.65	Y
-0.67	-0.05	-2.75	4.578	-3.42	3.762	-3.83	✗	-2.37	N
0.347	0.394	-0.63	0.886	-0.87	2.802	0	✓	2.93	Y
-3.37	-3.47	-1.83	0	-2.26	2.216	-2.99	✗	-11.70	N
0.588	0.394	-2.9	1.506	-3.56	0	-2.12	✗	-6.09	N
-1.24	-0.94	-0	1.626	0	2.962	-1.75	!	0.65	Y
-1.53	-1.69	0	0.619	-0.31	3.389	-2.81	✗	-2.33	Y
0.781	0.841	-3.43	2.372	-3.68	3.801	-3.34	✗	-2.66	N
0.685	0.692	-0.99	2.333	-1.27	3.955	-3.82	✓	1.59	N
-3.37	-3.47	-3.43	0	-3.68	0	-3.83		-11.70	
0.781	0.841	0	4.578	0	4.202	0		4.96	
0	0	-0.9	2.052	-1.5	3.587	-2.6		0.65	
1	1	1	1	1	1	1		3.94	

gp120 models which were co-crystallized with CD4 and/or 17b get better scores (2.28 on average) than those which are not co-crystallized (-4.2 on average). Correlation of the overall score and a (1/-1) labeling of the structures is 0.7363 (tail probablity 0.000214)

Figure S1: (Correspond to Fig 1 and Materials and Methods of main paper.) **(Top row) Illustration of scoring terms and controls/calibration** In **(A)**, we show two different models of gp120, and their interface with CD4 (yellow). The blue model was generated by Swiss-model (Schwede et al., 2003) using 1G9N:G as template, and our optimized model is shown in red. **(B)** comparison of the binding footprints of CD4 on two different gp120 models. The one on top is a model generated by Swiss-model (Schwede et al., 2003) using the PDB 3IDY as template, the one at the bottom is our final model. **(C)** provides a graphical understanding of external total ratio (ETR) which measures the fraction of atoms of a model lying outside a isosurface of the density map. We used the recommended isovalue (1.0) (Liu et al., 2008) and rendered with transparency so that atoms lying outside will be distinguishable from those inside. The blue model was generated by Swiss model using 1GC1 chain G as template and the V3 loop is almost completely outside the isosurface. Our optimized model, shown in red, has only a few atoms outside the isosurface. In **(D)**, we compare two models in terms of how much of the density map they are covering. The model on top is actually a crystal structure (1GC1:Chain G) and has poor MIS 53.78 compared to the model below (optimized model) whose MIS is 64.39. **(Bottom Row) Controls and calibration of scoring methods.** Our control or benchmark consist of 20 gp120 structures from the PDB. They are listed by their PDBIDs in the tables. The table in **(E)** details the z-scores for each term in $s_{internal}$. The rows beside 'Control', describe the min, max, mean and standard deviation of the z-scores of the 20 models in the control set. At the bottom of the table, the mean (avg) and standard deviation of scores for models with different resolutions, as reported in (Bhattacharya et al., 2007), are given. On average, the scores for the models in our benchmark correspond to the low-resolution (< 3.5Å) class and agrees with the actual resolutions (shown in the last column of the table). Also, the sum of z-scores (i.e. $s_{internal}$) for individual models in the control set, have a high correlation with corresponding actual resolutions. Hence, $s_{internal}$ is a statistically sound metric for tertiary structural and stereochemical quality. The table in **(F)** reports the interface and fitting based scores ($s_{external}$). Again, notice that the overall score is highly correlated with whether or not the gp120 model is in a co-crystallized state with CD4 and/or 17b (mentioned in the last column of the table). Hence, the validity of the interface and fitting based scores to distinguish correct interface/neighborhood is also established. Finally, the star, check, exclamation and cross icons visually highlight scores which are $\geq \mu + 2\sigma$, $\geq \mu + \sigma$, $\geq \mu$ and $< \mu$ respectively.

Details on calibration of scoring and search protocols

Benchmark preparation and z-score computation The success of any prediction algorithm, which optimizes a specific scoring scheme and/or uses some type of statistical evaluation of the prediction, hinges on the quality of the scoring scheme. One must ensure that the scoring metric has significant statistical correlation to ground truth. Bhattacharya et al. (Bhattacharya et al., 2007) prepared a benchmark of crystal and NMR structures with less than 50% similarity with each other and computed PROCHECK g-factors, Verify3D, ProSA and MolProbity composite scores. The composite scores were converted to z-scores defined as the number of standard deviations the score of a specific structure is from the mean score across the benchmark. Finally, they plotted the distribution, mean and standard deviation of the z-scores for structures with different resolutions. This provides a nice way to evaluate if the score of a given model lies within the range of say structures having < 3.5Å resolution (see bottom-left part of Fig S1(E)).

For the other terms, we use our own benchmark to compute the z-scores. We collected 20 structures of gp120 deposited in the PDB and computed raw scores for the other terms. Then the mean (μ), min (m), max (M) and standard deviation σ of these raw scores were used to define z-scores. Usually, a model with an average score gets a z-score of 0 by definition. However, complete gp120 models will have more than 100 extra residues as compared to the partial models in the control set. The interface area, MIS (which is maximized when larger portion of the density map is covered by a model upon fitting) and residue-contacts (negative and positive) for the complete models are thus expected to be higher than all the models in the control set. Hence, we use the extreme values of the control set as the expected value while computing z-scores for these terms. For example, the z-score for MIS, z_{MIS}, is defined as $(s_{MIS} - M_{MIS})/\sigma_{MIS}$ (see Fig S1(F)). Note that as far as ranking and comparison of models are concerned this is no different from using the mean.

Calibration of scoring metrics and thresholds Bhattacharya et al. (Bhattacharya et al., 2007) prepared a benchmark of crystal and NMR structures with less than 50% similarity with each other and reported the mean and variance of PROCHECK g-factors, Verify3D, ProSA and MolProbity composite scores. We verified if the distribution of the scores reported in (Bhattacharya et al., 2007) accurately represent the quality of the structures in our control set (see Fig S1A). We found that most of the structures in our benchmark had scores which lie within 2σ from the mean score (μ) for structures in the low-resolution class (Bhattacharya et al., 2007) of PSVS (resolution between $2.5 - 3.5$Å). The actual resolution of the structures in our set (shown in the resolution column in Fig S1E) ranges from 1.89 to 3.51, and hence the z-scores and ranges prescribed by PSVS are validated. More importantly, we found that $s_{internal}$ can correctly distinguish between low and high resolution crystal structures within the control set. The Pearson correlation coefficient of $s_{internal}$ and corresponding resolutions across the 20 models is -0.5927 which corresponds to a tail probability of 0.006175. The correlation is statistically significant and hence if the $s_{internal}$ of a model is higher than the average value (-8.14) of the control set, we can accept it, with high confidence, as a high resolution and stereochemically accurate model. Models which at are within 1 standard deviation and 2 standard deviations away from this average are considered medium and low quality models. Models with even lower scores are completely unacceptable. In some tables these classifications are shown using star, check, exclamation and cross icons.

The objective of $s_{external}$ is to distinguish between correct and incorrect interfaces/sites offered to the binding partners (e.g. CD4 and 17b), and correct and incorrect conformations for fitting/alignment to the EM map. The last column in Fig S1F shows which gp120 models were co-crystallized with CD4 and 17b and hence have a correct site topology. The correlation between this labeling and $s_{external}$ is 0.7363 and is statistically even more significant.

In conclusion, if a model is rated as high quality under both $s_{internal}$ and $s_{external}$, then the model is indeed high quality with high probability.

Search protocol calibration

Our multi-resolution docking/fitting algorithm and scoring/ranking scheme was validated by applying it to predict the correct binding interactions with gp120 and other molecules from a large set of co-crystallized structures. We first compute the best fitting of 1GC1, a complex that includes gp120 core,

CD4 and 17b, to the density map EMD5020. Then, for each crystal structure, we computed the best rigid body transformation which would align the gp120 chain to the fitted gp120. The transformation was applied to the entire structure (including other chains like CD4, 17b etc.). After the alignment, we kept the position of gp120 fixed, and created a copy of the other molecules and applied a random transformation to each. The randomly moved molecule is used as ligand, and the fixed gp120 is used as receptor for our multiresolution docking prediction tool. The position of the other molecule before applying the random transformation is considered the native state. If a predicted pose is within 4Å RMSD from this reference state, then the prediction is considered acceptable.

gp120-17b		First Near-Native Pose		Best Pose	
Receptor(PDBID-Chain)	Ligand(PDBID-Chain)	Rank	RMSD	Rank	RMSD
1G9M-G	1G9M-HL	71	4.1	441	3.3
1G9N-G	1G9N-HL	1	1.3	1	1.3
1GC1-G	1GC1-HL	1	2.4	13	2.1
1RZJ-G	1RZJ-HL	1	1.7	1	1.7
1YYL-G	1YYL-HL	1	2.3	168	1.4
1YYM-G	1YYM-HL	1	0.9	1	0.9
2I5Y-G	2I5Y-HL	1	1.4	1	1.4
2I60-G	2I60-HL	1	2.3	152	1.5
2NXZ-A	2NXZ-CD	7	2.1	19	1.8
2NY0-A	2NY0-CD	1	2	1	2
2NY1-A	2NY1-CD	1	1.3	1	1.3
2NY2-A	2NY2-CD	1	1.5	1	1.5
2NY3-A	2NY3-CD	1	2.2	1190	2.1
2NY4-A	2NY4-CD	1	1.8	31	1.6
2NY5-A	2NY5-CD	1	2.1	1	2.1
2NY6-A	2NY6-CD	56	4.7	276	1.7

Table ST1: (Correspond to Protocol Calibration section of main paper.) **Performance of multi-resolution docking (F2Dock + PF3Fit). on predicting 17b binding.** Result of applying F2Dock+PF3Fit to predict the orientation of 17b w.r.t. gp120. Both 17b and gp120 were extracted from co-crystallized structures available in the PDB. The co-crystallized structure was fitted to the density map. Then, given a randomly transformed 17b and keeping gp120 fixed at the fitted position, the docking protocol predicted the orientation of 17b which maximizes a scoring term similar to $s_{external}$ mentioned in the text. A predicted pose was considered acceptable (near-native) if it was within 4Å RMSD of the fitted position of 17b (before it was randomly moved). The rank of the first such pose, in the ordered list of predictions, and its RMSD is reported in the table. The best pose, on the other hand, is defined as the pose with the lowest RMSD across the first 2000 predictions. Notice that our docking protocol successfully predicted a near-native pose as the top-ranked solution in 13/16 cases. Hence, it is strongly validated that the scoring method clearly distinguishes between native and non-native poses, and the algorithm successfully samples the conformational space to find the native conformation. Also, in most cases, the top solution is the best solution as well.

For each of the complexes, we applied the above protocol and observed the rank and RMSD of the first acceptable prediction. An acceptable solution at high rank indicates that the scoring functions are capable of discriminating between native and non-native poses. A low RMSD indicates that the conformational space is sampled sufficiently.

Our docking protocol successfully predicted a near-native pose as the top-ranked solution in 13/16 cases for gp120-17b interactions and 11/16 cases for gp120-CD4 interactions (with 3 more cases having a correct solution within top 10). In 18 out of the overall 32 cases, our method picked the lowest RMSD solution as the top solution. Hence, it is strongly validated that the scoring method clearly distinguishes between native and non-native poses, and the algorithm successfully samples the conformational space to find the native conformation. Tables ST1 and ST2 provide further details.

gp120-cd4		First Near-Native Pose		Best Pose	
Receptor(PDBID-Chain)	Ligand(PDBID-Chain)	Rank	RMSD	Rank	RMSD
1G9M-G	1G9M-C	1	1.3	1	1.3
1G9N-G	1G9N-C	116	3.7	116	3.7
1GC1-G	1GC1-C	1	1.6	1	1.6
1RZJ-G	1RZJ-C	1	1.2	1	1.2
2B4C-G	2B4C-C	4	2.6	4	2.6
2NXZ-A	2NXZ-B	1	1.2	1	1.2
2NY0-A	2NY0-B	1	0.8	1	0.8
2NY1-A	2NY1-B	1	2.4	2	0.9
2NY2-A	2NY2-B	1	1.2	1	1.2
2NY3-A	2NY3-B	1	1.7	1	1.7
2NY4-A	2NY4-B	1	1.5	1	1.5
2NY5-A	2NY5-B	1	1.5	1	1.5
2NY6-A	2NY6-B	4	1.9	4	1.9
2QAD-A	2QAD-B	1	1.9	1	1.9
3JWD-A	3JWD-C	9	3	9	3
3JWO-A	3JWO-C	0	10	0	10

Table ST2: (Correspond to Protocol Calibration section of main paper.) **Performance of multi-resolution docking (F2Dock + PF3Fit) on predicting CD4 binding** An experiment similar to the one reported in Table ST1 with CD4, instead of 17b. Again, we get a near-native solution at rank 1 in 11/16 cases and within rank 10 in 14/16 cases. 3JWO is the only case where our method could not find any near native solutions.

Detailed description of different stages of the modeling protocol applied to gp120

Quality of initial stage template-based models Swiss-model Schwede et al. (2003) and I-TASSER Roy et al. (2010) are two state of the art protein modeling software that have additionally performed well at CASP challenges. When we tried to generate models using UniProtKB sequence P04578 (the same sequence as 1GC1:Chain G) to produce models which contained the variable regions, the sequence alignment and template identification tools picked 3JWD and 4NCO as templates. However, neither the model produced by Swiss-model nor the 5 models produced by I-TASSER scored within 2 standard deviations of the expected $s_{external}$ and $s_{internal}$ values. Then, we generated more models (61 in total) by manually selecting different templates (gp120 cores from different crystal structures). Figure S2A-B shows a superposition of all the models generated in this stage.

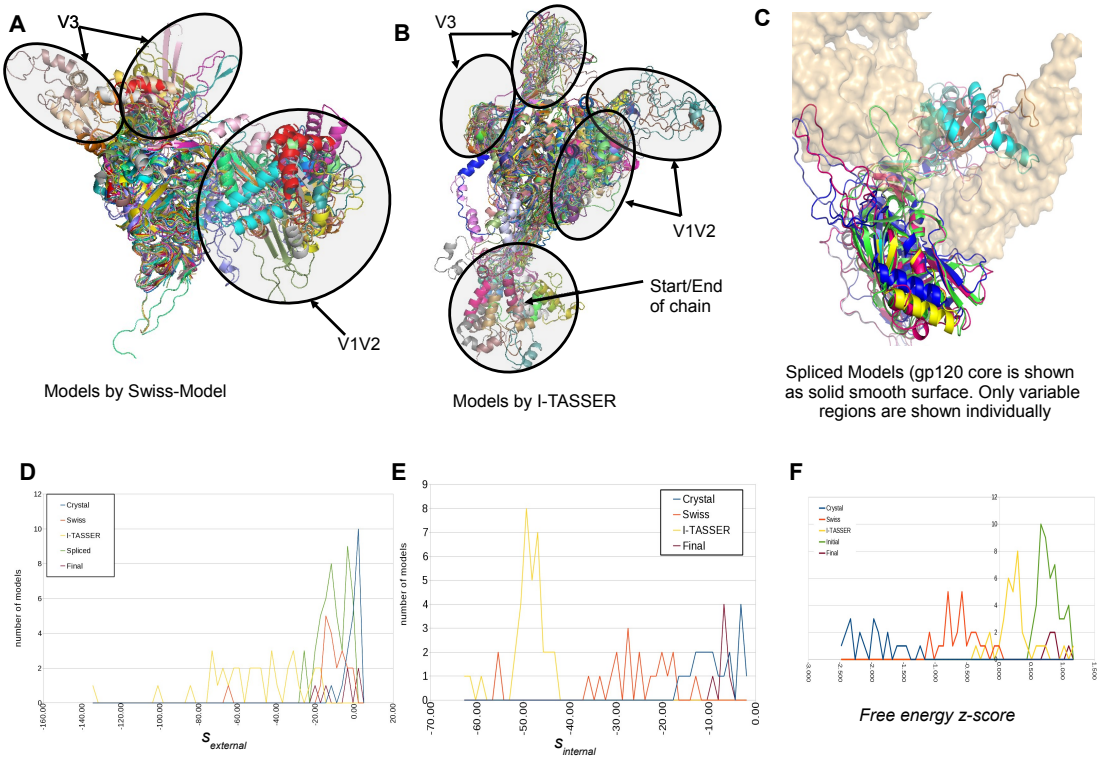


Figure S2: (Correspond to Figure 1, and the section titled 'Modeling gp120 in complex with CD4 and 17b' in the main paper) **(Top Row)** Superposition of initial stage models produced by Swiss-Model (Schwede et al., 2003) **(A)** and I-TASSER (Roy et al., 2010) **(B)**, and spliced models **(C)** created by assembling top ranked cluster representative for each fragment. Notice that the V1V2 loops in most I-TASSER model comes close to the gp120 core and shields the CD4 binding site. Such models score poorly under our $s_{external}$ score which penalizes models that does not offer suitable binding interfaces for CD4 and 17b, and also because they are generally protrude outside the chosen isosurface of the density map. Models for the start/end sections of the gp120 chain produced by I-TASSER were also poor in terms of fitting with the density map. **(Bottom Row)** Distribution of the quality of the models generated during different pahses (more details in Figure S3). We plotted the distribution of models generated by direct application of Swiss-model (Schwede et al., 2003) and I-TASSER (Roy et al., 2010) with different templates (stage 1), the models made by selected assembly (stage 2), the optimized models (stage 3) and crystal structures (control) in terms of $s_{external}$ **(D)**, $s_{internal}$ **(E)**, and free energy z-score **(F)**. The distributions show that the optimized models get, in general, similar scores as crystal structures. Swiss models and spliced model distributions are slightly lower quality.

In Figure S2D-F we have plotted the distribution of these models and compare the distribution with the crystal structures in our control set with respect to the range of $s_{external}$ and $s_{internal}$ values. The plots show that the I-TASSER models are in general poorer quality under both the scoring terms. Swiss models on the other hand have better scores, but $s_{internal}$ is still not close to the crystal structures. Actually there was only one swiss-generated model which was assessed as low quality in terms of $s_{internal}$, and every other model was assessed as unacceptable. Also there was only 1 medium and 1 low quality model in terms of $s_{external}$. See Tables ST3 and ST4 for details. Energetically, however, the I-TASSER models have lower solvation energy (Figure S2F). We noticed that in most of the I-TASSER models, the

		# of models			
	Average	$\geq \mu + 2\sigma$	$\geq \mu + \sigma$	$\geq \mu$	$< \mu$
Swiss (Stage 1)	-27.86	0	0	1	19
I-TASSER (Stage 1)	-48.57	0	0	0	35
Spliced (Stage 2)	-8.05	0	2	5	41
Optimized (Stage 3)	-6.39	4	1	0	0
x-ray Structures (Control)	-8.14	10	7	3	0

Table ST3: The second column of this table reports the average $s_{internal}$ scores of all models generated in different stages of the pipeline. The last four columns reports the number of models whose scores are $\geq \mu + 2\sigma$, $\geq \mu + \sigma$, $\geq \mu$ and $< \mu$ respectively. Please refer to the section on calibration of the pipeline for details on the computation of μ and σ.

		# of models			
	Average	$\geq \mu + 2\sigma$	$\geq \mu + \sigma$	$\geq \mu$	$< \mu$
Swiss (Stage 1)	-10.93	0	1	1	10
I-TASSER (Stage 1)	-50.25	0	0	0	35
Spliced (Stage 2)	-9.09	0	14	13	21
Optimized (Stage 3)	-5.24	2	0	0	4
x-ray Structures (Control)	0.65	3	11	1	5

Table ST4: The second column of this table reports the average $s_{external}$ scores of all models generated in different stages of the pipeline. The last four columns reports the number of models whose scores are $\geq \mu + 2\sigma$, $\geq \mu + \sigma$, $\geq \mu$ and $< \mu$ respectively. Please refer to the section on calibration of the pipeline for details on the computation of μ and σ.

V1V2 loop conformation brought it close to the core, introducing clashes with CD4 and disagreement with the EM map as evidenced by poor ETR, clash and MIS scores (see Figure S3), but increasing the contacts and interface area and also lowering the energy. This highlights the fact that unless the EM density map and/or the neighboring proteins are considered at while modeling a protein, the resulting low energy state may not be the native one.

Note that we have also attempted an alternate apporach where the bound configuration of gp120, CD4 and 17b (from 1GC1) together were used as templates to achieve better $s_{external}$ by ensuring fewer clashes with CD4 and 17b. However, these attempts produced even worse results (I-TASSER), or no results at all (Swiss-model).

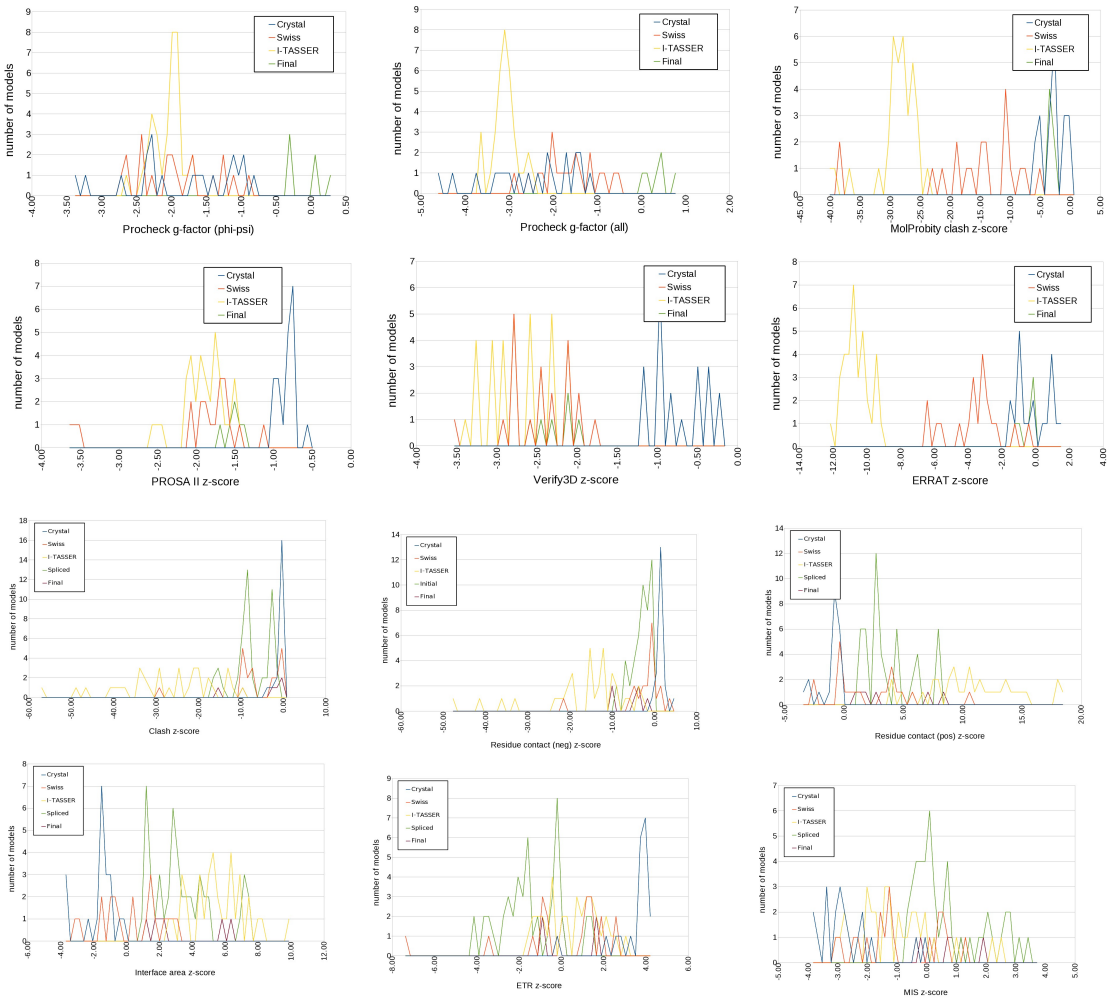
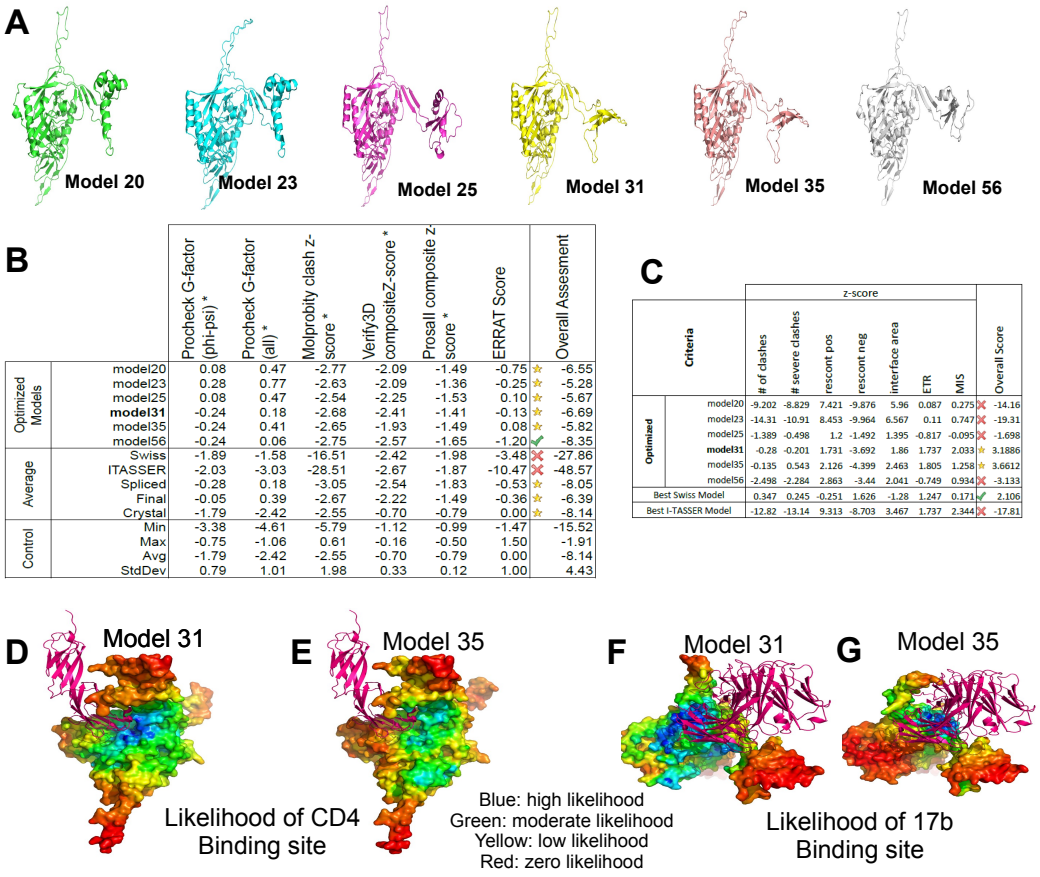


Figure S3: (Correspond to Figure 1, and the section titled 'Modeling gp120 in complex with CD4 and 17b' in the main paper) **Comparison of the distribution of models for individual terms in $s_{internal}$ and $s_{external}$.** (In top-bottom, left-right order) PROCHECK g-factor ($\phi - \psi$), PROCHECK g-factor (all), MolProbity clash z-score, ProSA II z-score, Verify3D z-score, ERRAT z-score, Clash, interface area, positive residue contacts, negative residue contacts, external total ratio and mutual information score z-scores

Quality of fragments and models generated by fragment assembly Each of the initial models were decomposed into fragments (core, V1V2, V3, V4, C-termini and N-termini). Then the fragments were clustered based on similarity under TM-score Xu and Zhang (2010). Some clusters for V1V2 and the four selected structures are shown in Figure 1(A-B) of the main text. Interestingly, we found that even though an initial model scored poorly, it might contain a fragment which is locally quite feasible and gets a high score when considered on its own.

Since all the component fragments are already fitted into the density map in their correct relative orientations, assembly does not require any major reconfiguration. However the structures are not stereochemically sound as the bond lengths/angles at the joint are too far from ideal (see Figure 1C in the main text for example). However after local structural refinement and energy minimization, the stereochemical and energetic quality of the models are significantly improved. Also, $s_{external}$ scores are better than initial model, which is unsurprising given the procedure of fragment selection employed in our protocol. Please see Tables ST3 and ST4 for a summary of the scores, as well as Figures S2D,E. Figure S2C shows a superposition of all the spliced models.

We clustered the spliced models based on TM-scores and then selected six models which were in different clusters. Among the cluster representatives, two (Model31 and Model56) were rated medium quality, 1 model (Model25) was rated low quality, and the other three (Model20, Model23 and Model35)



B

		Procheck G-factor (phi-psi) *	Procheck G-factor (all) *	Molprobity clash z-score *	Verify3D compositeZ-score *	Prosall composite z-score *	ERRAT Score		Overall Assesment
Optimized Models	model20	0.08	0.47	-2.77	-2.09	-1.49	-0.75	☆	-6.55
	model23	0.28	0.77	-2.63	-2.09	-1.36	-0.25	☆	-5.28
	model25	0.08	0.47	-2.54	-2.25	-1.53	0.10	☆	-5.67
	model31	-0.24	0.18	-2.68	-2.41	-1.41	-0.13	☆	-6.69
	model35	-0.24	0.41	-2.65	-1.93	-1.49	0.08	☆	-5.82
	model56	-0.24	0.06	-2.75	-2.57	-1.65	-1.20	✓	-8.35
Average	Swiss	-1.89	-1.58	-16.51	-2.42	-1.98	-3.48	✗	-27.86
	ITASSER	-2.03	-3.03	-28.51	-2.67	-1.87	-10.47	✗	-48.57
	Spliced	-0.28	0.18	-3.05	-2.54	-1.83	-0.53	☆	-8.05
	Final	-0.05	0.39	-2.67	-2.22	-1.49	-0.36	☆	-6.39
	Crystal	-1.79	-2.42	-2.55	-0.70	-0.79	0.00	☆	-8.14
Control	Min	-3.38	-4.61	-5.79	-1.12	-0.99	-1.47		-15.52
	Max	-0.75	-1.06	0.61	-0.16	-0.50	1.50		-1.91
	Avg	-1.79	-2.42	-2.55	-0.70	-0.79	0.00		-8.14
	StdDev	0.79	1.01	1.98	0.33	0.12	1.00		4.43

C

	Criteria	z-score							Overall Score
		# of clashes	# severe clashes	rescont pos	rescont neg	interface area	ETR	MIS	
Optimized	model20	-9.202	-8.829	7.421	-9.876	5.96	0.087	0.275	✗ -14.16
	model23	-14.31	-10.91	8.453	-9.964	6.567	0.11	0.747	✗ -19.31
	model25	-1.389	-0.498	1.2	-1.492	1.395	-0.817	-0.095	✗ -1.698
	model31	-0.28	-0.201	1.731	-3.692	1.86	1.737	2.033	☆ 3.1886
	model35	-0.135	0.543	2.126	-4.399	2.463	1.805	1.258	☆ 3.6612
	model56	-2.498	-2.284	2.863	-3.44	2.041	-0.749	0.934	✗ -3.133
Best Swiss Model		0.347	0.245	-0.251	1.626	-1.28	1.247	0.171	✓ 2.106
Best I-TASSER Model		-12.82	-13.14	9.313	-8.703	3.467	1.737	2.344	✗ -17.81

Figure S4: (Correspond to Figure 1-3, and the section titled 'Modeling gp120 in complex with CD4 and 17b' in the main paper) **Closer look at the optimized models**. **(A)** Structure of the six models which got selected, based on clustering and a combined $s_{external} + s_{internal}$ score, for energy minimization and optimization. **(B)** presents a summary of the $s_{external}$ scores for the optimized models. Notice that both Model31 and Model35 scored very high. **(C)** presents the $s_{internal}$ of the optimized models. All but one model scores very high and hence must have excellent stereochemistry and comparable to x-ray models with resolution between 2.5 to 3.5Å (please see the section on protocol calibration in this supplement). **(D-G)** Binding site analysis visualized. Left two figures (D,E) show likelihood of each point on the surface of Model31 and Model35 of being on the CD4 binding site (as predicted by docking). Each point on the surface of the models are colored by the probability of it being at the binding interface, where blue-red gradient (like a rainbow) is used to show high-low probabilities. Model31 shows higher affinity at the correct site. The right two figures (F,G) show the 17b binding sites. Both models have high probabilities near the correct site. Model31 also have slightly lower probability of binding at the b12 epitope as well, which is not desired. Model35 on the other hand have a very focused predicted site, but slightly offset from the correct site. Note that in all the pictures the correct location of the binding sites is implied by rendering the CD4 and 17b chains (as ribbons) in their native poses.

were rated unacceptable. These six models are shown in Figure S4A. Note that poorer models were also selected to ensure diversity, and in fact Model35 became high quality after co-optimization further energy minimization (see next section).

Quality of models after co-optimization We first applied our docking Chowdhury et al. (2013) and fitting Bettadapura et al. (2012); Bajaj et al. (2013) protocols to improve the relative configuration of gp120, CD4 and 17b with each other as well as with respect to the EM map EMD5020. As a result, the score for almost all the terms improved significantly. Breakdown of the scores for each term for each of these models are contrasted with the average scores of models from previous stages in Figures S4B-C. These data also reveal that all the optimized models have very good energetics and $s_{internal}$ scores. However, two models (Model20 and Model23) have quite poor $s_{external}$ scores.

Model31 and Model35 both are assessed as high quality in terms of both $s_{internal}$ and $s_{external}$ scores, and considered medium quality in terms of energy. Note that high quality means that they scored better than average crystal structures in our control, and according to the calibration and correlation mentioned before, their qualities are equivalent to resolutions better than 3.5Å. However, our binding site analysis (see next section) revealed that Model31 was better than Model35 in terms of the binding interface offered

to CD4 and 17b. So, Model31 was was further refined using side-chain repositioning at the interfaces to remove any clashes and improve residue contacts. The quality of this final model is described separately in a later section.

Binding site analysis The binding site analysis was performed by docking (using F2Dock Chowdhury et al. (2013)) CD4 and 17b with the optimized models, and F2Dock reports the top 1000 possible binding poses. We define the parts of the surface of gp120 which is in contact with the CDR loops of CD4 and 17b in a docking pose as the footprint/site of that particular pose. For each point on the surface of gp120 model, we compute the ratio of the number of poses whose footprint includes the point, and the total number of poses as the probability of that point being on the binding site. The binding site score is then defined as the sum of the probailities of all the points on the surface which would be in contact with CD4 (or 17b) in their native poses. In other words, a model which has high likelihood of having the binding site at the correct region, scores high. Model31 had more specificity near the correct CD4 and 17b binding sites compared to model35 (see Figures S4D-G for a visualization) and was reported as the final model.

Detailed quality assessment of the final model

Here, we first detail the quality assessment scores of the final model under the metrics used in our protocol. Then, for independent verification, we report further validation scores based on other metrics which have not been used in the protocol. Note that, unless specified otherwise, all scores reported here pertains to the entire trimeric complex.

Ramachandran plot analysis by Procheck MacArthur et al. (1993) showed 89.6% residues in most favored, 7.6% in additionally allowed, 1.4% in generously allowed and only 1.4% in disallowed regions. A summary plot is given in Fig S5A. Overall Procheck g-factor is -0.22 for $\phi - \psi$ angles only and -0.04 for all, both of which is extremely favorable and correspond to high resolution ($< 2\mathring{A}$) structures Bhattacharya et al. (2007). In total 10 band-contacts were reported and 3.6% residues were found to have bad planarity.

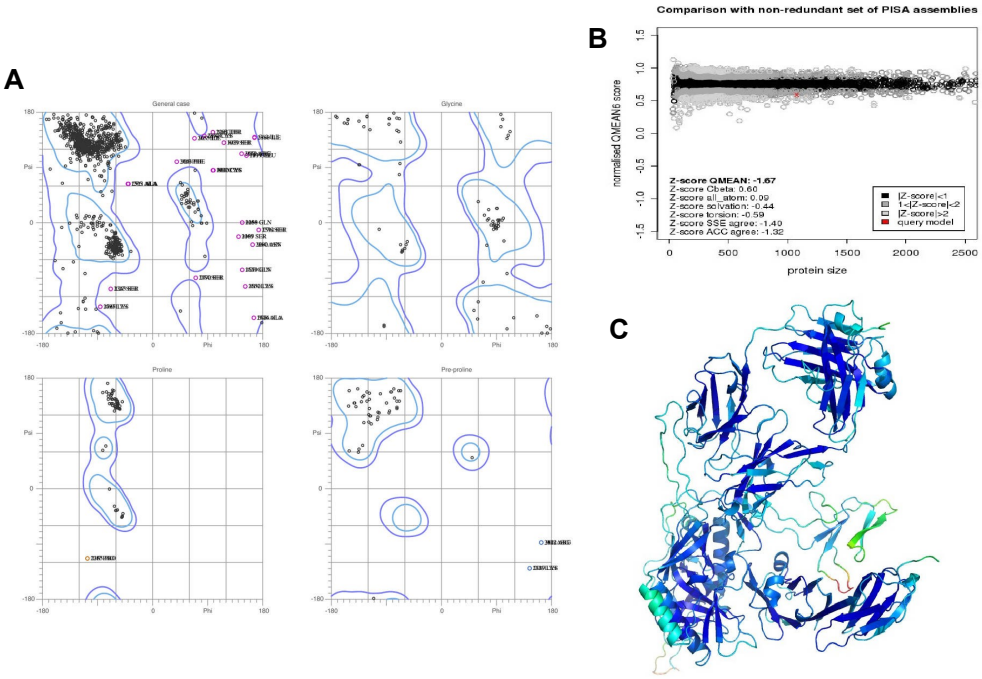


Figure S5: (Correspond to Figure 1 and the section on model quality assessment in the main paper) **(A)** Summary Ramachandran plot of the final model of the trimer. 89.6% residues are in most favored, 7.6% in additionally allowed, 1.4% in generously allowed and only 1.4% in disallowed regions. **(B)** The Qmean z-score (Benkert et al., 2011) of the final model of the trimer of gp120+CD4+17b is -1.666. In the figure, each circle represents an x-ray model and the red cross represents our model. In the plot, our model lies in the range where models have average level of confidence. **(C)** the surface of the model (1 part of the trimer) is colored based on ANOLEA score (Melo et al., 1997). Only small segments have poor quality which is quite expected, primarily since those regions are far from CD4 and 17b and hence fewer constraints are available to improve their qualities.

ProsaII Sippl (1993) composite score for the model is 0.76 which is also representative of high resolution structures Bhattacharya et al. (2007). MolProbity Davis et al. (2007) composite score for the model is 30.44 (z-score -3.70), which in general indicates that a model is in the low resolution range. However, we note that among existing x-ray models of gp120, 1G9M, 1G9N, 1GC1, 1RZJ and 3RJQ all have worse MolProbity scores. Verify3D Luthy et al. (1992) reports that more than 71% of the residues have a 1D-3D score above 0.2, which is in acceptable range according to Verify3D's guidelines.

We used the PDB validation software (ADIT), Modeval Shen (2006) and Qmean z-score Benkert et al. (2011) to provide independent validation of the quality. PDB validation software (ADIT) reports RMS deviation for bond angles at 0.7 degrees and bond length deviation of 0.003Å, both of which is quite acceptable. ModEval Shen (2006) predicted an RMSD of 3.378 (for the gp120 chain only). The Qmean z-score was -1.666 which is within the acceptable range for a protein of this size. A plot showing the quality of our model with respect to existing x-ray models in terms of Q-mean z-scores in given in Figure S5B. The model, colored by per residue error under the ANOLEA Melo et al. (1997) score, is also shown in Figure S5C.

Supplementary discussion regarding implication of the final model

Configuration of the V1V2 loops with respect to antibodies binding at CD4bs

We computed the footprint of different antibodies whose x-ray structures in bound state with gp120 (core or complete) are available. We transformed the bound gp120-antibody complex such that the gp120 chain aligns with our gp120 model. A detailed list of the number of contacts and clashes for different antibodies is reported in Figure S6A.

A

Epitope	Antibody	Source Model	# of atoms in contact	# of atoms In clash
Around CD4 binding Site	CD4	new model	169	0
	NIH45-46	4JDT	161	8
	NIH45-46	3U7Y	241	8
	PGV04	3J5M	189	7
	PGV04	3SE9	241	7
	VRC03	3SE8	355	39
Around 17b binding site	17b	new model	135	0
	48d	4DVR	131	1
	48d	3JWD	172	1
	PGT122	4NCO	70	3
	X5	2B4C	87	2
Others	b12	2NY7	573	117
	b13	3IDX	524	128
	b13	3IDY	507	125
	f105	3HI1	539	156
	PG9	3U2S	90	36

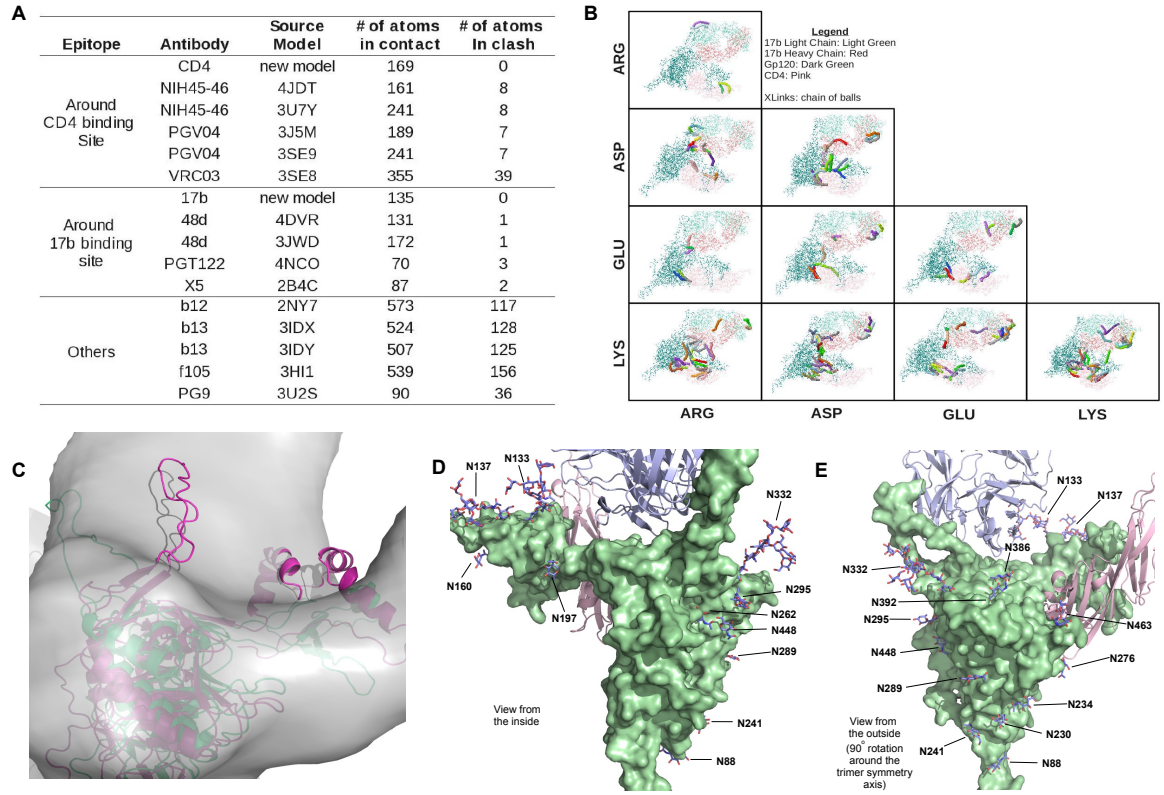


Figure S6: (Correspond to Fig 5 and 6 of main paper) **Implications derived from the model. (A)** Comparison of the footprint of different antibodies on our gp120 model. The antibodies are grouped by their epitope. The last two columns report the number of atoms of gp120 which come in contact with the antibody, and the number of atoms of gp120 which clash with the antibody. Two atoms a and b are considered to be in contact if the distance between their centers is less than $r_a + r_b + 1$ where r_a and r_b are the radii of the atoms. Two atoms are considered to be clashing if the distance between their centers is less than $min(r_a, r_b)$. The contacts and clashes with the antibodies were computed after aligning our gp120 model to the x-ray model containing the antibody bound with gp120 (core or complete). As expected, antibodies that bind to the b13 and f105 epitopes have heavy clash with the V1V2 loop, since in our model the loop partially occludes these epitopes. But the interesting aspect is that all antibodies which bind at the CD4 binding site, also have clashes with the V1V2 loop. So, the V1V2 loop must be in a different configuration when gp120 binds with these antibodies. For example, in EMD5021 Liu et al. (2008), we see that when b12 binds with gp120, the V1V2 loop stays in a similar ternary configuration as seen in the unliganded state. **(B)**The computationally predicted inter-domain (gp120-CD4, gp120-17b and CD4-17b) crosslinks are shown. Each crosslink is shown as a series of contiguous balls, and the actual molecules are shown only using sticks (each stick representing a bond). Colors of the crosslink are only meant to keep a distinction between them, and do not carry any other meaning/interpretation. The figures are arranged like a matrix where each cell represents crosslinks between residues of the types indicated beside the row and column. For example, the second from left figure in the bottom row shows crosslinks between LYS and GLU residue types. We observe a large number of crosslinks between the V1V2 region and CD4, as well as the V3 region and the light chain of 17b; but very few crosslinks between 17b and the V1V2 region. **(C)** Comparison of the best models generated with and without using the EM map (EMD 5020). The prescribed isocontour of the EM map is shown as transparent smooth surface. The best model with ETR and MIS terms in the scoring is shown as ribbon model in green color. The best model without ETR and MIS terms in the scoring is shown as ribbon model in purple color. **(D-E)** show glycans derived from 1RZJ Huang et al. (2004), and 4NCO Lyumkis et al. (2013) mapped onto our model. Note that the configurations of the glycans are distal from the CD4 and 17b chains.

Possible cross-links indicate clues to conformational change

Chemical cross-linking is often used to generate low resolution distance constraints between parts of a protein (or multiple proteins). We used Xwalk Kahraman et al. (2011), a computational tool that

mimics cross-linking experiments by calculating the distance between two residues along the surface of the proteins, to identify inter-domain(gp120-CD4 and gp120-17b) cross-links in our model. We considered only cross-links between ARG, ASP, GLU and LYS residues whose C-beta atoms were within 25Å of each other (Figure S6B).

As expected, we observed a high number of cross-links between residues at the CD4bs with CD4, and 17bbs with the heavy chain of 17b. However, we also identified a large number of cross-links between CD4 and the V1V2 region, and a few cross-links between the light chain of 17b with the V3 region. The cross-links between 17b-CD4, and 17b-V1V2 were very few. The lack of predicted cross-link constraints between 17b and the V1V2 region, and the presence of high number of predicted cross-link constraints between CD4 and V1V2, may be considered as another structural explanation for the conformational motion of the loop, especially the preference of the V1V2 to move away from the 17b binding site (or by extension, the CCR5 binding site).

Modeling glycans

HIV-1 Env is heavily glycosylated. Although possible glycosylation sites are not difficult to identify, the exact type and configuration of the glycan, which can vary wildly between strains and depending on the bound partner Bonomelli et al. (2011), is difficult to model with high-confidence unless finer resolution EM-maps are available. Glycans have recently become targets of many recent antibodies that bind to glycans near V3 loop Kong et al. (2013); Julien et al. (2013a), V1V2 loop McLellan et al. (2011); Pancera et al. (2013); Doores and Burton (2010), gp41-gp120 interface Blattner et al. (2014); Louise Scharf and Bjorkman (2014) etc. However, glycan dependence of the CD4 and 17b binding interactions is still not clear.

We have extracted glycans from 1RZJ for the glycosylation sites on the core and 4NCO for those on the variable loops. The glycans were grafted to their respective locations on the new model (see Figure S6D-E). While all of the glycan sites including the ones on the variable loops seem distal from the CD4 and 17b binding sites in this static model, a more rigorous study of the glycan configurations is warranted for a better understanding of the possible effect of the glycans on CD4 and 17b binding. For example, recently it was reported that some antibodies like PGT121, which binds far from the CD4bs, still manages to prevent CD4 binding by induced conformational changes to the glycans and variable loops Julien et al. (2013b). We are currently exploring new scoring models that would use recently developed databases Sunhwan Jo (2013) of glycan structures to effectively quantify the quality of specific glycan configurations and their interactions with the protein chains.

Best scoring model without using EM map

If the EM map is not used, the protocol may not be able to prune out possibly incorrect ternary configurations of the variable domains early in the pipeline. The contact-based terms and interface area does help guiding the protocol to generate feasible and energetically stable configurations, they are often not sufficient to reach the native-like (in-vitro) state available from EM. In this particular case, we found that if one does not use the EM-based scoring terms (ETR+MIS) in the protocol, then a pretty close model to the one we reported is found. However, when the model is fitted back into the EM (for validation), we found that it has large portions of both V1V2 and V3 regions outside the prescribed isocontour of the EM map (see Fig S6C).

References

Bajaj, C., Bauer, B., Bettadapura, R. and Vollrath, A. (2013). Nonuniform Fourier Transforms for Rigid-Body and Multi-Dimensional Rotational Correlations. SIAM J. of Sc. Comput. *35*, B821–B845.

Benkert, P., Biasini, M. and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics *27*, 343–350.

Bettadapura, R., Bajaj, C. and Vollrath, A. (2012). PF3Fit: Hierarchical Flexible Fitting in 3D EM. Technical Report 12-18 ICES, UT at Austin.

Bhattacharya, A., Tejero, R. and Montelione, G. T. (2007). Evaluating protein structures determined by structural genomics consortia. Proteins *66*, 778–795.

Blattner, C., Lee, J. H., Sliepen, K., Derking, R., Falkowska, E., Peña, A. T. d. l., Cupo, A., Julien, J.-P., Gils, M. v., Lee, P. S. and et al. (2014). Structural Delineation of a Quaternary, Cleavage-Dependent Epitope at the gp41-gp120 Interface on Intact HIV-1 Env Trimers. Immunity *40*, 669–80.

Bonomelli, C., Doores, K. J., Dunlop, D. C., Thaney, V., Dwek, R. A., Burton, D. R., Crispin, M. and Scanlan, C. N. (2011). The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. PLoS ONE *6*, 1–7.

Chowdhury, R., Rasheed, M., Keidel, D., Moussalem, M., Olson, A., Sanner, M. and Bajaj, C. (2013). Protein-Protein Docking with F2Dock 2.0 and GB-Rerank. PLoS ONE *8*, e51307.

Colovos, C. and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Prot. Sc. *2*, 1511–1519.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. and et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nuc Acids Res. *35*, W375–W383.

Doores, K. J. and Burton, D. R. (2010). Variable loop glycan dependency of the broad and potent HIV-1-neutralizing antibodies PG9 and PG16. J. Vir. *84*, 10510–10521.

Glaser, F., Steinberg, D. M., Vakser, I. A. and Ben-Tal, N. (2001). Residue Frequencies and Pairing Preferences at Protein-Protein Interfaces. Proteins *43*, 89–102.

Huang, C.-c., Venturi, M., Majeed, S., Moore, M. J., Phogat, S., Zhang, M.-Y., Dimitrov, D. S., Hendrickson, W. A., Robinson, J., Sodroski, J. and et al. (2004). Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. PNAS *101*, 2706–2711.

Julien, J.-P., Cupo, A., Sok, D., Stanfield, R. L., Lyumkis, D., Deller, M. C., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P. and et al. (2013a). Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. Science *342*, 1–12.

Julien, J.-P., Sok, D., Khayat, R., Lee, J. H., Doores, K. J., Walker, L. M., Ramos, A., Diwanji, D. C., Pejchal, R., Cupo, A. and et al. (2013b). Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. PLoS Path. *9*, e1003342.

Kahraman, A., Malmström, L. and Aebersold, R. (2011). Xwalk: computing and visualizing distances in cross-linking experiments. Bioinformatics *27*, 2163–2164.

Kong, L., Lee, J. H., Doores, K. J., Murin, C. D., Julien, J.-P., McBride, R., Liu, Y., Marozsan, A., Cupo, A., Klasse, P.-J. and et al. (2013). Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. Nature Struc. Mol. Biol. *20*, 796–803.

Liu, J., Bartesaghi, A., Borgnia, M. J., Sapiro, G. and Subramaniam, S. (2008). Molecular architecture of native HIV-1 gp120 trimers. Nature *455*, 109–113.

Louise Scharf, Johannes F. Scheid, J. H. L. A. P. W. J. C. C. H. G. P. N. G. R. M. M. S. S. A. B. W. M. C. N. and Bjorkman, P. J. (2014). Antibody 8ANC195 Reveals a Site of Broad Vulnerability on the HIV-1 Envelope Spike. Cell Rep. *7*, 785–795.

Luthy, R., Bowie, J. U. and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. Nature *356*, 83–85.

Lyumkis, D., Julien, J.-P., Val, N. d., Cupo, A., Potter, C. S., Klasse, P.-J., Burton, D. R., Sanders, R. W., Moore, J. P., Carragher, B. and et al. (2013). Cryo-EM Structure of a Fully Glycosylated Soluble Cleaved HIV-1 Envelope Trimer. Science *1484*.

MacArthur, M. W., Moss, D. S., Laskowski, R. A. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. App. Cryst. *26*, 283–291.

McLellan, J. S., Pancera, M., Carrico, C., Gorman, J., Julien, J.-P., Khayat, R., Louder, R., Pejchal, R., Sastry, M., Dai, K. and et al. (2011). Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature *480*, 336–343.

Melo, F., Devos, D., Depiereux, E. and Feytmans, E. (1997). ANOLEA: a www server to assess protein structures. Int. Conf. Intelli. Sys. Mol. Biol. *5*, 187–190.

Pancera, M., Shahzad-Ul-Hussan, S., Doria-Rose, N. a., McLellan, J. S., Bailer, R. T., Dai, K., Loesgen, S., Louder, M. K., Staupe, R. P., Yang, Y. and et al. (2013). Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. Nat. Struc. Mol. Bio. *20*, 804–13.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004). UCSF Chimera–a Visualization System for Exploratory Research and Analysis. Journal of Computational Chemistry *25*, 1605–12.

Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Prot. *5*, 725–738.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nuc. Acids Res. *31*, 3381–5.

Shatsky, M., Hall, R., Brenner, S. and Glaeser, R. (2008). A method for the alignment of heterogeneous macromolecules from electron microscopy. Journal of Structural Biology *166*, 67–78.

Shen, M. (2006). Statistical potential for assessment and prediction of protein structures. Prot. Sc. *15*.

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins *17*, 355–362.

Sunhwan Jo, Hui Sun Lee, J. S. W. I. (2013). Restricted N-glycan Conformational Space in the PDB and Its Implication in Glycan Structure Modeling. PLoS Comp. Bio. *9*, 1–10.

Vasishtan, D. and Topf, M. (2011). Scoring functions for cryoEM density fitting. Journal of Structural Biology *174*, 333–343.

Xu, J. and Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics *26*, 889–895.