On the irrationality of rational design of an HIV vaccine in light of protein intrinsic disorder

Vladimir N. Uversky

Abstract

The lack of progress in finding an efficient vaccine for a human immunodeficiency virus (HIV) is daunting. In fact, this search has spanned nearly four decades without much success. There are several objective reasons for such a failure, which include the highly glycosylated nature of HIV-1, the presence of neotopes, and high mutation rates. This article argues that the presence of highly flexible and intrinsically disordered regions in both human anti-HIV-1 antibodies and the major HIV-1 immunogen, its surface glycoprotein gp120, represent one of the major causes for the lack of success in utilization of structure-based reverse vaccinology.

Structure-based reverse vaccinology [52, 63] represents an important constituent of rational vaccine design, where an attempt is made to produce a vaccine using information from the observed crystallographic structure of neutralizing monoclonal antibodies (mAbs) bound to their complementary epitopes. Such structural vaccinology is believed to represent a way to facilitate rational design of better antigens able to act as vaccine immunogens. Computer-based reverse engineering methodology utilizes the available structures of pathogenic proteins and antigen-antibody complexes and uses docking and modeling studies to predict epitopes to reconstruct an epitope capable of mAb binding, where the structure of mAb is used as a template in a process similar to rational drug design, where the 3D structure of a biological target is used for designing molecules capable of the selective binding to and specific inhibition of the biological activity of a target molecule [12]. It should be noted that mAbs used in this context must have been shown to be broadly neutralizing (in vitro). Here, an assumption is made that such an antigen “rationally” designed to fit the mAb would have the desired immunogenic potential and would be able to induce polyclonal antibodies (Abs) that would possess neutralizing properties similar to those of the mAb used as the template for the computational reconstruction of the antigen. Alternatively, one can utilize available information on the whole genomes/proteomes of various pathogens and use computational approaches to identify the bacterial/viral surface antigens that are most likely to be vaccine candidates. The rationale of this approach is based on the assumption that virtually all of the proteins of the pathogenic microorganisms are likely to contain antigenic sites that can be predicted by computational means. Here, the entire bacterial or viral proteome is subjected to in silico analysis in order to find all of the antigens that a bacterial or viral pathogen is able to express [52, 63]. Again, structural vaccinology is utilized here for prediction and characterization of structural epitopes of immunogenic antigens. At the next stage of both approaches, the ability of individual antigens to elicit immunity in animal models is tested [13, 65].

There is no doubt that reverse vaccinology has open new horizons, and this is supported by multiple research studies and dedicated reviews that emphasize the strengths and successes of this approach for rapid targeted identification of novel vaccine antigens as well as for improving the immunogenicity and safety of vaccine antigens [12, 13, 52, 63, 65]. Examples include development of strongly immunogenic vaccine candidates based on the respiratory syncytial virus (RSV) glycoprotein and the spike protein of Middle
were grafted onto computationally designed scaffolds [5, 6]. The culmination of the success of this strategy is the successful development of designer antigens, where structural information pertaining to existing bNAbs-antigen complexes was utilized to design linear and discontinuous HIV epitopes that were grafted onto computationally designed scaffolds [5, 6]. Despite all these and many similar success stories of the application of rational vaccine design, and despite the fact that several broadly neutralizing anti-HIV-1 Abs have been found (e.g., IgG1 b12 targeting the HIV-1 envelope glycoprotein gp120 of the viral spike [95] and IgG 2G12, which recognizes a carbohydrate epitope on gp120 [92]), not a single promising HIV vaccine has been found after nearly 40 years of research utilizing various approaches, including structure-based reverse vaccinology [22, 27, 28, 91]. Among the accepted explanations for these failures are the high glycosylation level of the HIV surface (e.g., HIV is the most glycosylated virus known [87], with glycans composing up to half of the only surface HIV-1 protein, the Env protein [49]) and the presence of rather recently described neotopes (or neo-antigenic sites) in HIV-1 gp120 (neotopes have been described in many viruses since 1966 [85], but only recently in HIV [34]). Here, neotopes are novel transient epitopes that are not present in the viral protein monomers but arise from the quaternary structure of a multimeric assembly of identical subunits of viral proteins, either from the juxtaposition of residues in neighboring subunits that are recognized by Abs as a single epitope or from conformational changes induced in the protein by intersubunit interactions [56]. Curiously, although the term neotope was coined by Prof. Marc H. V. van Regenmortel in 1966 [85], for a long time, this phenomenon was mostly studied by plant virologists studying the serological properties of plant viruses. In the field of animal/human viruses, the importance of neotopes became commonly recognized much later [86]. As a result, until quite recently, vaccine designers looked only at the monomeric forms of pathogen-derived proteins, despite the recognition that the neotope can be formed via several mechanisms. For example, complex-specific, neo-antigenic sites that are not detectable on monomeric proteins but are found in complexes of target proteins with receptors are rather common [83, 84]. The neo-epitopes can also appear due to the dissociation of the native oligomeric forms of proteins into free subunits (e.g., C-reactive protein, CRP [69]). Another common mechanism of neotope formation includes specific cleavage of target proteins by proteases [11]. Also, neotopes can originate due to specific posttranslational modifications of target proteins [1, 4]. Finally, neotopes can be generated as a result of structural perturbations of target proteins induced by free radicals and oxidative damage [30, 73], as is the case in autoimmune rheumatoid disease [30, 73]. It is also implied that HIV infection typically promotes an immune response against highly variable immunodominant epitopes (e.g., the V1, V2, and V3 loops of the envelope glycoprotein gp160) that does not provide protection against diverse HIV strains [65]. Like other retroviruses, HIV utilizes an error-prone polymerase for replication, and the resulting high rate of mutation accounts for many of the difficulties in the search for the HIV vaccine [27, 28, 91]. It seems that this situation is similar to that of influenza virus, which serves as a well-known illustration of the correlation between high mutation rates and difficulties in finding effective vaccines [27]. However, the influenza vaccine analogy is not strictly applicable to HIV, since the efficacy of the correctly anticipated influenza vaccine mix can be as high as 69-83% [58], whereas the maximal efficacy of an HIV vaccine has so far reached only about 30% [22]. This is an interesting observation taking into account that, since the late 1980s, more than 100 candidate prophylactic vaccines have been tested in phase 1 clinical trials but only a few of those have advanced to the stage of phase 2b or phase 3 efficacy studies [22], of which only one candidate vaccine, ALVAC, has shown modest efficacy of 31.2% [4, 22, 75]. Recently, a modified version of this partially effective HIV vaccine failed to show efficacy in preventing HIV, and the corresponding $100 million clinical trial HVTN 702 in South Africa evaluating it has been stopped early (https://www.niaid.nih.gov/news-events/experimental-hiv-vaccine-regimen-in-ineffective-preventing-hiv). Importantly, not all retroviruses are without vaccines, as effective vaccines for equine infectious anemia virus (EIAV) have been developed [19, 50, 51, 55, 82, 89, 90], whereas there are still no vaccines for the non-retroviral RNA virus hepatitis C virus (HCV) or the DNA virus herpes simplex virus (HSV) [7, 10].

The major immunogens of HIV-1 are trimeric spikes formed by Env heterodimers of gp41 and gp120 originating during Env maturation as a result of cleavage of the full-length gp160 protein by a protease. The surface of the virus is decorated by 14 such spikes [9, 72]. Although many bNAbs to the Env protein of HIV-1 interact directly with glycans (e.g., the aforementioned IgG 2G12, which recognizes a carbohydrate epitope on gp120 [92]), some of the bNAbs do interact with the proteinaceous part of gp120 (e.g., IgG1 B12 [45, 95]). Let us consider some structural aspects of
such antigen-Ab interactions, which constitute the foundation of structure-based reverse vaccinology and rational structural vaccinology in general.

We will start with a brief description of the Ab structural organization. Although there are five different classes of Abs, which are commonly known as immunoglobulins, IgM, IgD, IgG, IgA, and IgE, they all have a comparable structural organization of functional units, and we will focus here on IgG. A well-recognized image of a typical Ab is a Y-shaped molecule consisting of three equal-sized parts that are connected by flexible linkers (Fig. 1A) and have different functions, with the arms (or variable [V] regions, since they vary among different antibody molecules, or antigen-binding fragment, F_{ab}) being involved in specific antigen binding, and with the stem part (or constant fragment, F_{c}) being less variable and responsible for interactions with various effectors. Importantly, since in the Abs, the “arms” are connected to the “stem” via flexible linkers, the resulting Y-shaped molecules are highly flexible (with the hinge regions connecting the F_{ab} fragments to the F_{c} domains being especially mobile), and therefore, known Ab crystal structures can only be considered “snapshots” of the broad range of conformations available to these proteins in solution [71]. This high conformational flexibility of the hinges connecting F_{ab} fragments with the F_{c} fragment represents a major reason for the sparsity of resolved crystal structures of full-length immunoglobulins.

Zooming in to the “active sites” of Abs reveals that they have two identical binding clefts located within the N-terminal regions of their H and L chains. Each of these binding clefts is made up of 50-70 hypervariable residues and includes several overlapping paratopes; i.e., binding subsites of 10-20 residues, which are structurally and chemically complementary to the epitopes; i.e., certain patches of residues present at the surface of the target protein [29]. These paratopes are built from short stretches of residues located on six complementarity determining regions (CDRs, L1, L2, L3 on the light chain and H1, H2, and H3 on the heavy chain) that form discontinuous binding sites. Therefore, these six CDRs of each Ab arm are able to come into contact with the antigen. One should keep in mind, though, that although as many as all six CDRs can contribute amino acid residues to the contact surface with the antigen, often, fewer than six CDRs actually do so, and as few as four CDRs might contribute to contact with the antigen. Furthermore, the effectiveness of camelid V_{H}-only antibodies suggests that perhaps fewer than four CDRs could be involved in contacting antigen in some complexes. Typically, the antigen-binding sites of anti-protein Abs are relatively flat. However, H3 loops (which are located in the center of the binding site) in human Abs are often extended [8, 41, 70, 71], allowing them better access to the canyons and clefts on the antigen surface [71, 77]. Many of these H3 loops in human Abs are characterized by unique structural features [64]. This is illustrated by an important study of known 3D structures of human Abs (1,779 structures with 4,989 chains), which revealed that the H3 loops contain on average 10 times more unique conformations than the other loops, with more than a thousand four-residue-long fragments of H3 adopting conformations not seen in any other structure [64]. This enormous polymorphism of the available structures of the H3 loop is likely a reflection of its highly flexible or even disordered nature in solution, where this loop exists as a highly dynamic conformational ensemble, and each crystal structure represents a snapshot showing one member of this ensemble. In line with these considerations are the results of an analysis of the predisposition for intrinsic disorder in both the H and L chains of a typical IgG (human antibody IgG1 B12) (Fig. 1B and C, respectively). This analysis revealed that both chains contain several intrinsically flexible or disordered regions (i.e., regions with a predicted disorder score (PDS) ≥ 0.5 or 0.15 ≤ PDS < 0.5, respectively), with the CDRs being predicted to be either flexible or disordered. This strongly suggests that the information about conformational plasticity is encoded in the amino acid sequences of these important regions. Figure 1B and C show another interesting feature of the H and L chains, namely the presence of redox-sensitive regions (i.e., cysteine-containing regions capable of undergoing disorder-to-order or order-to-disorder transitions in response to changes in the redox state of the environment [54]).

Figure 2 shows that, in different anti-HIV Abs of human or bovine origin, the H3 loop length can vary over a very wide range, from four residues in the non-neutralizing HIV antibody 13H11 (PDB ID: 3MO1) to 16 residues in the broadly reactive and potent HIV-1-neutralizing human antibody PG9 (PDB ID: 3U1S; [53]), and to 60 residues in the potent HIV-1 bNAb NC-Cow1 (PDB ID: 6OO0; [81]). Importantly, the overall close resemblance of these structures is determined by their high sequence similarity (sequence identity of these chains ranges from 48.33% to 88.61%). Furthermore, a multiple structural alignment revealed that these 10 structures, ranging in length from 215 to 273 residues, can be aligned over the region of 98 residues (corresponding to the N-terminal domain of the V_{H}; i.e., the V domain of the H chain) with a root-mean-square deviation (RMSD) of 0.83 Å, and the region of the successful structural alignment (with an RMSD below 1 Å) can increase
Protein intrinsic disorder and HIV vaccine design

Figure 1 Structure and disorder in human IgG. A. Crystal structure of the intact human IgG B12 with broad and potent activity against primary HIV-1 isolates (PDB ID: 1HZH) [71]. This Y-shaped structure originates from the specific packing of four protein chains, two identical heavy (H) chains (blue and cyan structures) and two identical light (L) chains (red and orange structures), each composed of variable and constant regions located at the N- and C-terminal parts of the chains, respectively. In IgG, each L chain is made up of two independent domains with a typical immunoglobulin fold, whereas each H chain has four such domains. The antigen-binding V region of the Ab (the Fab fragment) is made of the variable V domains of the H and L chains (VH and VL, respectively), whereas the Ab constant region (the Fc fragment) is made of the constant C domains of the H and light L (CCH and CL, respectively). The overall structure is maintained not only by specific chain-chain interactions but also by a network of disulfide bonds that link the two H chains to each other in the Fc region and also link each H chain to an L chain in each Fab region. Utilization of the two identical H chains and the two identical L chains to form a Y-shaped structure results in the formation of two identical antigen-binding sites in any given immunoglobulin molecule, which are able to simultaneously interact with two identical structures. B. Intrinsic disorder predisposition and the presence of redox-sensitive regions (i.e., cysteine-containing regions capable of undergoing disorder-to-order or order-to-disorder transitions in response to changes in the redox state of the environment) in a heavy chain of the human antibody IgG1 B12. A redox-sensitive region is shown with dark red shading. C. Intrinsic disorder predisposition and the presence of the redox-sensitive regions in a light chain of the human antibody IgG1 B12. Profiles shown in plots B and C were generated using the IUPred2A platform [54].

Figures 2 and 3 show that despite the high content of cysteine residues, this CDR H3 of NC-Cow1 is predicted to be rather flexible, whereas Figure 3B suggests that this region is characterized by a strong redox-sensing potential and is expected to undergo an order-to-disorder transition when disulfides are reduced. Figure 3C and D illustrate that the CDR H3 loop of different Abs shown in Figure 2 are all predicted to be flexible or disordered, with the degree of disorderliness increasing with the length of the H3 loop. One should also keep in mind that with a very high probability, the observed structures of these long Fab appendices are stabilized (or even induced) by interaction of the Ab with antigens or by the crystal lattice. Here, “a protein crystal lattice consists of surface contact regions, where the interactions of specific groups play a key role in stabilizing the regular arrangement of the protein molecules” within a crystal [94]. Obviously, some of these interactions between specific groups in the surface contact regions of proteins within the crystal lattice can reduce conformational flexibility and induce structuration. These observations strongly suggest that conformational flexibility and intrinsic disorder are crucial for the functionality of anti-HIV-1 Abs (or at least for the binding efficiency of their H3 loops).

It is time now to consider intrinsic disorder-related features of the major immunogen of HIV-1, its gp120 protein, which is a constituent of the trimeric spikes of this virus and which attaches the virus to the host lymphoid cell by binding to the primary receptor CD4. Figure 4A shows the disorder profile of the full-length gp120 of HIV-1 group M subtype B (isolate YU-2), which is a 467-residue-long surface protein containing variable regions V1 through V5 (residues 98-122, 123-160, 260-293, 348-373, and 416-426), which are the most genetically diverse regions of the entire HIV-1 genome, as well as a CD4-binding loop (residues 327-337). All of these functional regions of gp120 are located within or in close proximity to disordered regions. It is of interest to...
note that, in the majority of the crystallization experiments, various shorter forms of the protein were used instead of the full-length gp120. The disorder profiles for two illustrative examples of these so-called gp120 core forms are shown in Figure 4B (the extended gp120 core, 376 residues) and Figure 4C (the gp120 core from the HXBc2 laboratory-adapted isolate, 306 residues). Comparison of the disorder profiles shown in Figure 4 clearly indicates that some disordered/flexible regions had been removed from the gp120 core constructs used in the crystallization experiments. For example, in the extended gp120 core, these removed regions are residues 90-165, and 265-288 of the full-length gp120, which clearly correspond to its variable regions V1, V2, and a significant part of V3, whereas in the gp120 core from the HXBc2 laboratory-adapted isolate, the removed regions are residues 1-50, 96-159, 264-293, and 448-467, which, in addition to the N- and C-tails, include the variable regions V1, V2, and V3. To illustrate the structural variability of the gp120 cores utilized in crystallographic experiments, Figure 5 shows the results of multiple structural alignment for 10 such chains ranging in length from 288 to 343 residues and shows that these structures can be accurately aligned over the region of 234 residues with an RMSD of 0.66 Å. This analysis shows that although 70 to 80% of these structures are very similar, the remaining parts of each chain show unique structural features.
Similar to other rational structure-based drug discovery approaches, rational structural vaccinology uses the known 3D structure of an Ab to find epitopes that would form specific epitope-paratope complexes. However, the data considered in this article indicate that both partners that are used in such structure-based rational design of anti-HIV vaccines (gp120 and Abs) are characterized by high conformational plasticity and have regions with considerable intrinsic disorder. This raises serious doubts about the overall applicability of such computational methods in this case. In fact, structure-based rational design is based on considering complexes (protein-ligand, protein-protein, protein-nucleic acid or antigen-antibody, as in this particular case) in terms of rigid, motionless structures with steric complementarity to each other. This approximation goes back to the famous lock-and-key model of enzyme catalysis proposed by Hermann Emil Fischer [23, 48] and the “side-chain theory” suggested by Paul Ehrlich for antigen-antibody complexes more than a century ago [18, 21, 33, 46]. However, protein-protein binding involves a process of induced complementarity and fit resulting from the mutual adjustments of the two partners that involve important side-chain movements and changes in the backbone conformation, as in the flexible keys and adjustable locks model that was proposed by Edmundson et al. in 1987 [20] based on the induced-fit model suggested earlier by Daniel Edward Koshland for enzyme catalysis [37]. In this induced-fit model, the fit between an active site and a substrate is brought about by substrate binding, since the spatial positioning and the 3D relationship of the amino acids at the active site of an enzyme can be changed by binding of the substrate, and such substrate-induced structural changes bring the catalytic groups into the proper orientation for the reaction, which does not occur with a non-substrate [37]. This also indicates that a binding site is a relational entity, undergoing “fine tuning” in response to interaction with a partner, and it is not solely defined by intrinsic structural features that are identifiable independently of the relationship with a particular partner. Obviously, these considerations apply to all partners involved in complex formation, since their binding sites are engaged in mutual tuning. One should also keep in mind that the scale of such mutual tuning can range from rather minimal structural adjustments to global binding-induced folding.

Finally, there is another intrinsic-disorder-related angle in the mystery of the HIV-1 vaccine failure. It has been pointed out that the HIV-1 matrix protein (p17) is expected to be highly disordered [25]. In fact, based on computational analysis, it was concluded that, depending on a strain, the percentage of intrinsic disorder (PID, which is the percentage of residues with PDS ≥ 0.5) in p17 can be as high as 70%, a level encountered very rarely in the outer shell of other viruses [25, 26]. This is further supported by an analysis of the shell disorder status of over 300 viruses in the publicly available database [24, 28], which indicates that, in addition to HIV-1, the outer-shell proteins of herpes simplex virus (HSV) and hepatitis C virus (HCV) are highly disordered as well [27]. Curiously, successful vaccines have not been established for these three viruses (HIV-1, HSV, and HCV) with highly disordered outer shells, suggesting that the motions arising from the disordered outer shell might lead to the inability of antibodies to bind tightly to the polysaccharides on the viral surface proteins or to the viral surface proteins themselves, rendering the immune response inadequate [27]. Therefore, the failure of vaccines based on the HIV glycoprotein gp120 and other vaccines can be traced back to a lack of understanding of the important role of shell disorder in immune evasion by such viral shape-shifters [27].

Recently, the results of a systematic analysis of the intrinsic disorder in shells of various viruses in relation to vaccine development were reported [27]. It was shown that successful vaccines have been developed for influenza, smallpox, rabies, polio, and yellow fever, all of which are caused by viruses characterized by a relatively low level of intrinsic disorder in the outer shell. On the other hand, there are no efficient vaccines against EIAV, HIV-1, HIV-2, HSV-1, HSV-2, or HCV, all of which show higher intrinsic disorder in their shells. In line with these considerations is the recent success with the development of the anti-SARS-CoV-2 vaccines. In fact, although the maximal proportion of intrinsically disordered residues in the outer shells of HIV-1, HCV, and HSV reaches 70, 53, and 63%, respectively, in SARS-CoV and SARS-CoV-2, the corresponding values are much lower: 8 and 6%, respectively [27].

One should keep in mind that in accounting for the lack of success to date in the quest for an HIV-1 vaccine, it is impossible to determine, even roughly, the relative contributions of the high mutation rate of HIV or the specific structural features of gp120, such as its heavily glycosylated state, and the appearance of quaternary-structure-dependent neotopes versus the contribution of conformational flexibility or the presence of intrinsically disordered regions. This is because all of these seemingly structure-related features are intertwined with intrinsic disorder and structural flexibility. In fact, in viral proteins (and proteins in general), mutations most commonly occur in flexible or disordered regions. Quaternary
structure formation is frequently associated with folding upon binding events; i.e., flexible or disordered regions of protomers undergo a transition to a more ordered state as a result of oligomerization or complex formation. Therefore, at least some neotopes originate due to this binding-induced folding. Finally, the sites of many posttranslational modifications (glycosylation included) are located within disordered, or at least flexible, regions.

All of these considerations suggest that attempts to use rational structure-based design for the development of HIV vaccines are rather spurious. In fact, since neither the Ab nor gp120 (nor, as a matter of fact, any other antigen) has a static rigid structure, the use of such rational structure-based design in these cases resembles searching in vain under the streetlight for keys lost in a dark alley. Curiously, intrinsic disorder is considered the major reason for the existence of the dark proteome comprising proteins that are not amenable to experimental structure determination by existing means and that are inaccessible to homology modeling [40]. These considerations are very important, cannot be ignored, and clearly should be taken into account while thinking about novel approaches to HIV vaccine design. It is time to move away from playing with the motionless toys. Reality is more complex than the static “lock-and-key”. In fact, it is even more complex than the “flexible keys and adjustable locks” model.

Fig. 3 Intrinsic disorder predispositions of the H chains of human and bovine Fabs. A. Multifactorial analysis of the intrinsic disorder predisposition of the H chain of the potent HIV-1 bNAb NC-Cow1. The disorder profile was generated using a DiSpI web crawler designed to aggregate the results from a number of well-known disorder predictors: PONDR® VLXT [67], PONDR® VL3 [61], PONDR® VLS2 [60], PONDR® FIT [93], IUPred2 (Short), and IUPred2 (Long) [16, 17]. This tool enables the rapid generation of disorder profile plots for individual polypeptides as well as arrays of polypeptides. Positions of various secondary structure elements in the mini-domain (knob) located at the tip of the long CDR H3 are shown. Here, β-strands are found at residues 111-113, 128-130, and 142-143, whereas α-helices are located at residues 116-121, 131-133, and 139-141. In reference 82, it was pointed out that “the knob domain begins with a conserved CPED motif (CPDG in the germline) containing a type I β-turn around PEDY, followed by three very short, antiparallel β-strands (D6-D8, D23-D25, and D37-D38) with two intervening loops of 14 and 11 residues. Loop 1 forms a single turn of helix, while loop 2 has two small helical turns. The knob has three disulfide bonds with 1-4, 2-5, and 3-6 connectivity (D2-D23, D12-D32, and D21-D37).”

B. Evaluation of the redox sensitivity of the H chain of the potent HIV-1 bNAb NC-Cow1 (i.e., the presence of the cysteine-containing regions capable of undergoing disorder-to-order or order-to-disorder transitions in response to changes in the redox state of the environment) of the H chain of the potent HIV-1 bNAb NC-Cow1. A redox-sensitive region is shown with dark red shading.

C. Aligned intrinsic disorder profiles of the H chains of various human and bovine Fabs generated using PONDR® VLXT [67].

D. Aligned intrinsic disorder profiles of the H chains of various human and bovine Fabs generated using PONDR® VLS2 [60]. Gaps in these plots correspond to gaps in the multiple sequence alignment of the corresponding chains.
Fig. 4 Characterization of intrinsic disorder predisposition of various gp120 forms used in the crystallization experiments. A. Intrinsic disorder profile of the full-length (467 residues) gp120 of HIV-1 group M subtype B (isolate YU-2). B. Intrinsic disorder profile of the extended gp120 core (376 residues). C. Intrinsic disorder profile of the gp120 core of a laboratory-adapted HXBc2 isolate (306 residues).

Disorder profiles were generated using a DiSpi web crawler designed to aggregate the results from a number of well-known disorder predictors: PONDR® VLXT [67], PONDR® VL3 [61], PONDR® VLS2 [60], PONDR® FIT [93], IUPred2 (Short), and IUPred2 (Long) [16, 17]
Fig. 5 Structure of gp120 in different bound forms. Comparison of the gp120 structures complexed with the CD4-binding-site antibody F105 (PDB ID: 3HI1:G) [14]; complexed with CD4 and the induced neutralizing antibody 17B (PDB ID: 1G9N:G) [44]; complexed with CD4 and the induced neutralizing antibody (PDB ID: 1RZK:G) [31]; complexed with CD4M33, a scorpion-toxin mimic of CD4 and anti-HIV-1 antibody 17B (PDB ID: 1YYL:G) [32]; unliganded HIV-1 gp120 core (PDB ID: 3TGQ:A) [42]; in complex with Fab 48d and NBD-557 (PDB ID: 4DVR:G) [43]; complexed with CD4-mimetic miniprotein M48U1 (PDB ID: 4JZW:G) [3]; in complex with the CD4-mimetic miniprotein M48U1 and the llama single-domain, broadly neutralizing, co-receptor binding site antibody JM4 (PDB ID: 4LAJ:A) [2]; bound to CD4 and 17B Fab (PDB ID: 4RQS:G) [36]; and in a complex with VH1-46 germline-derived CD4-binding site-directed antibody 8ANC131 (PDB ID: 4RWY:A) [96]. Structures are shown as ribbon diagrams. In the middle of the figure is a multiple structure alignment of the indicated structures (4LAJ, 343 residues; 4JZW, 339 residues; 4RWY, 337 residues; 3TGQ, 336 residues; 1RZK, 306 residues; 1G9N, 306 residues; 1YYL, 301 residues; 4RQS, 295 residues; 3HI1, 289 residues; 4DVR, 288 residues) that was made using the MultiProt algorithm (http://bioinfo3d.cs.tau.ac.il/MultiProt/) [74]. The alignment of 234 residues achieved an RMSD of 0.66 Å.

Funding The author did not receive support from any organization for the submitted work.

Availability of data and material All data generated or analyzed during this study are included in this published article.

Compliance with ethical standards

Conflict of interest The author certifies that he has no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Abidi M, Mir AR, Khan F, Ali A, Uddin M (2020) Glycoxidation of LDL generates cytotoxic adducts and elicits humoral response in type 2 diabetes mellitus. Glycobiology. https://doi.org/10.1093/glycob/cwaa077
2. Acharya P, Luongo TS, Georgiev IS, Matz J, Schmidt SD, Louder MK, Kessler P, Yang Y, McKee K, O’Dell S, Chen L, Baty D, Chames P, Martin L, Mascola JR, Kwong PD (2013) Heavy chain-only IgG2b llama antibody effects near-pan HIV-1 neutralization by recognizing a CD4-induced epitope that includes elements of coreceptor- and CD4-binding sites. J Virol 87:10173–10181
3. Acharya P, Luongo TS, Louder MK, McKee K, Yang Y, Kwon YD, Mascola JR, Kessler P, Martin L, Kwong PD (2013)
40. Kulkarni P, Uversky VN (2018) Intrinsically disordered proteins: the dark horse of the dark proteome. Proteomics 18:e1800061

41. Kunert R, Rucker F, Katinger H (1998) Molecular characterization of five neutralizing anti-HIV type 1 antibodies: identification of unconventional D segments in the human monoclonal antibodies 2G12 and 2F5. AIDS Res Hum Retroviruses 14:1115–1128

42. Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee BK, Moore LR, Schmidt SD, Stuckey J, Yang Y, Zhou T, Zhu J, Vicic DA, Debnath AK, Shapiro L, Bewley CA, Mascola JR, Sodroski JG, Kwong PD (2012) Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. Proc Natl Acad Sci USA 109:5663–5668

43. Kwon YD, LaLonde JM, Yang Y, Elban MA, Sugawara A, Courtier JR, Jones DM, Smith AB 3rd, Debnath AK, Kwong PD (2014) Crystal structures of HIV-1 gp120 envelope glycoprotein in complex with NBD analogues that target the CD4-binding site. PLoS ONE 9:e85940

44. Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (2000) Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. Structure 8:1329–1339

45. Kwong PD, Mascola JR, Nabel GJ (2013) Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. Nat Rev Immunol 13:693–705

46. Lapage G (1952) Paul Ehrlich. Nature 169:362

47. Lee JH, Andradi R, Su CY, Yasmeen A, Julien JP, Kong L, Wu NC, McBride R, Sok D, Pautner M, Cottrell CA, Nieuwsa T, Blattner C, Paulson JC, Klasse PJ, Wilson IA, Burton DR, Ward AB (2017) A broadly neutralizing antibody targets the dynamic HIV envelope trimer apex via a long, rigidified, and anionic beta-hairpin structure. Immunity 46:690–702

48. Lemieux RU, Spohr U (1994) How Emil Fischer was led to the lock and key concept for enzyme specificity. Adv Carbohydr Chem Biochem 50:1–20

49. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J Biol Chem 265:10373–10382

50. Leroux C, Cadore JL, Montelaro RC (2004) Equine infectious anemia virus (ELAV): what has HIV’s country cousin got to tell us? Vet Res 35:485–512

51. Li F, Craigo JK, Howe L, Steckbeck JD, Cook S, Issel C, Montelaro RC (2003) A live attenuated equine infectious anemia virus proviral vaccine with a modified S2 gene provides protection from detectable infection by intravenous virulent virus challenge of experimentally inoculated horses. J Virol 77:7244–7253

52. Massignani V, Rappuoli R, Pizza M (2002) Reverse vaccinology: a genome-based approach for vaccine development. Expert Opin Biol Ther 2:895–905

53. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchali R, Sastry M, Dai K, O’Dell S, Patel N, Shahzad-ul-Hussain S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Dwanji D, Georgiev I, Kwon YD, Lee D, Louder MK, Moqin S, Schmidt SD, Yang ZY, Bonsignore M, Crump JA, Kaipa SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480:336–343

54. Meszaros B, Erdos G, Dosztanyi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 46:W329–W337

55. Montelaro RC, Grund C, Raabe M, Woodson B, Cook RF, Cook S, Issel CJ (1996) Characterization of protective and enhancing immune responses to equine infectious anemia virus resulting from experimental vaccines. AIDS Res Hum Retroviruses 12:413–415

56. Neurath AR, Rubin BA (1971) Viral structural components as immunogens of prophylactic value. Karger, Basel

57. Ofek G, Tang M, Sambor A, Katinger H, Mascola JR, Wyatt R, Kwong PD (2004) Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. J Virol 78:10724–10737

58. Osterholm MT, Kelley NS, Sommer A, Belongia EA (2012) Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. Lancet Infect Dis 12:36–44

59. Pallesen J, Wang N, Corbett KS, Wrapp D, Kirchdoerfer RN, Turner HL, Cottrell CA, Becker MM, Wang L, Shi W, Kong WP, Andres EL, Klettenbach AN, Denison MR, Chappell JD, Graham BS, Ward AB, McLellan JS (2017) Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. Proc Natl Acad Sci USA 114:E7348–E7357

60. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obra-dovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. J Bioinform Comput Biol 3:35–60

61. Peng K, Radivojac P, Vucetic S, Dunker AK, Obra-dovic Z (2006) Length-dependent prediction of protein intrinsic disorder. BMC Bioinform 7:208

62. Purtscher M, Trkola A, Gruber G, Buchacher A, Predl R, Steinidl F, Tauer C, Berger R, Barrett N, Jungbauer A et al (1994) A broadly neutralizing human monoclonal antibody against gp41 of human immunodeficiency virus type 1. AIDS Res Hum Retroviruses 10:1651–1658

63. Rappuoli R (2001) Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 19:2688–2691

64. Regep C, Georges G, Shi J, Popovic B, Deane CM (2017) The H3 loop of antibodies shows unique structural characteristics. Proteins 85:1311–1318

65. Rinaudo CD, Telford JL, Rappuoli R, Seib KL (2009) Vaccinology in the genome era. J Clin Invest 119:2515–2525

66. Robinson LN, Tharakaraman K, Rowley KJ, Costa VV, Chan KR, Katinger H, Mascola JR, Wyatt R, Ofek G, Tang M, Sambor A, Katinger H, Mascola JR, Wyatt R, Ofek G, Tang M, Sambor A, Katinger H, Mascola JR, Wyatt R, Ofek G, Tang M, Sambor A, Katinger H, Mascola JR, Wyatt R, Ofek G, Tang M, Sambor A, Katinger H, Mascola JR, Wyatt R (1995) Directed selection of recombinant human monoclonal antibodies to herpes simplex virus glycoproteins from phage display libraries. Proc Natl Acad Sci USA 92:6439–6443

67. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Nievergelt B, Wang YH, Ong LC, Tan HC, Koch T, Cui D, Kirloskar R, Viswanathan K, Liew CW, Tissire H, Ramakrishnan B, Myette JR, Babcock GJ, Sasekumar V, Alonso S, Chen J, Lescar J, Shriver Z, Ooi EE, Sasekumar R (2015) Structure-guided design of an anti-dengue antibody directed to a non-immunodominant epitope. Cell 162:493–504

68. Romero P, Obadovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. Proteins 42:38–48

69. Saini SS, Allore B, Jacobs RM, Kaushik A (1999) Exceptionally long CDR3H region with multiple cysteine residues in functional human immunoglobulin J domains. Nature 401:1115–1128

70. Sanna PP, Williamson RA, De Logu A, Bloom FE, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480:336–343

71. Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, Rudd PM, Dwek RA, Stanfield RL, Burton DR, Wilson IA (2001) Protein intrinsic disorder and HIV vaccine design
