Supporting Information

Abrams et al. 10.1073/pnas.0914540107

SI Text

Methods. ***Derivation and justification of temperature-accelerated molecular dynamics.*** Consider the coupled system of equations (Eqs. **2** and **5** in main text)

$$m_i\ddot{x}_i = -\frac{\partial V(\boldsymbol{x})}{\partial x_i} - \kappa\sum_{j=1}^{m}[\theta_j^*(\boldsymbol{x}) - \theta_j]\frac{\partial\theta_j^*(\boldsymbol{x})}{\partial x_i} - \gamma m_i\dot{x}_i + \eta_i(t;\beta)$$

$$\tilde{\gamma}\tilde{m}_j\dot{\theta}_j = \kappa[\theta_j^*(\boldsymbol{x}) - \theta_j] + \xi_j(t;\tilde{\beta}). \qquad \textbf{[S1]}$$

This system of equations governs the coupled evolution of $\boldsymbol{x}(t)$ and $\boldsymbol{\theta}(t)$ over the extended potential

$$U_\kappa(\boldsymbol{x},\boldsymbol{\theta}) = V(\boldsymbol{x}) + \frac{\kappa}{2}\sum_{j=1}^{m}[\theta_j^*(\boldsymbol{x}) - \theta_j]^2 \qquad \textbf{[S2]}$$

with the collective variables (CVs)$\boldsymbol{\theta}(t)$ being overdamped.

If we set the friction coefficient $\tilde{\gamma}$ large enough so as to make the CVs $\boldsymbol{\theta}(t)$ slow compared to the original variables $\boldsymbol{x}(t)$, we can make $\boldsymbol{x}(t)$ adiabatic to $\boldsymbol{\theta}(t)$. This means that $\boldsymbol{x}(t)$ are at equilibrium conditional on $\boldsymbol{\theta}(t)$ being fixed at their current value, i.e., they are distributed according to the density

$$\rho_\kappa(\boldsymbol{x};\boldsymbol{\theta}(t)) = Z_\kappa^{-1}(\boldsymbol{\theta}(t))\exp[-\beta U_\kappa(\boldsymbol{x},\boldsymbol{\theta}(t))], \qquad \textbf{[S3]}$$

where $Z_\kappa(\boldsymbol{\theta}(t)) = \int\exp[-\beta U_\kappa(\boldsymbol{x},\boldsymbol{\theta}(t))]d\boldsymbol{x}$ is a normalization factor. Note that it is the physical temperature β^{-1} that enters Eq. **S3** because we use $\boldsymbol{\eta}(t;\beta)$ in the equation for $\boldsymbol{x}(t)$ in Eq. **S1**. Taking $\tilde{\gamma}$ large enough also guarantees that, at any given moment, the CVs $\boldsymbol{\theta}(t)$ only feel the average effect of $\boldsymbol{x}(t)$. Concretely, this means that the evolution of $\boldsymbol{\theta}(t)$ can be captured by an effective equation in which the right-hand side of the equation for $\boldsymbol{\theta}(t)$ in Eq. **S1** is averaged out with respect to Eq. **S3**. With a little algebra this equation can be written as

$$\tilde{\gamma}\tilde{m}_j\dot{\theta}_j = -\frac{\partial F_\kappa(\boldsymbol{\theta}(t))}{\partial\theta_j} + \xi_j(t;\tilde{\beta}), \qquad \textbf{[S4]}$$

where we defined

$$F_\kappa(\boldsymbol{\theta}) = -\beta^{-1}\ln Z_\kappa(\boldsymbol{\theta}). \qquad \textbf{[S5]}$$

Using Eq. **S3** as well as the definition of $F(\boldsymbol{\theta})$, it is easy to see that $F_\kappa(\boldsymbol{\theta})$ is a smoothed version of the actual free energy $F(\boldsymbol{\theta})$:

$$\exp[-\beta F_\kappa(\boldsymbol{\theta})] = \int\exp[-\beta F(\boldsymbol{\theta}')]\exp\left[-\frac{\beta\kappa}{2}\sum_{j=1}^{m}[\theta_j' - \theta_j]^2\right]d\boldsymbol{\theta}'. \qquad \textbf{[S6]}$$

In other words, for κ large enough, $F_\kappa(\boldsymbol{\theta})$ gives an accurate approximation of $F(\boldsymbol{\theta})$. In this case, the existence of the effective equation (Eq. **S4**) guarantees that the CVs $\boldsymbol{\theta}(t)$ in Eq. **S1** explore the free energy landscape in a way that is consistent with the density $\exp[-\tilde{\beta}F_\kappa(\boldsymbol{\theta})] \approx \exp[-\tilde{\beta}F(\boldsymbol{\theta})]$ (up to an irrelevant constant), where $F(\boldsymbol{\theta})$ is the free energy at the physical temperature β^{-1}, $F_\kappa(\boldsymbol{\theta})$ is the filtered version of this free energy, and $\tilde{\beta}$ appears because we use $\boldsymbol{\xi}(t;\tilde{\beta})$ in the equation for $\boldsymbol{\theta}(t)$ in Eq. **S1**.

To sum up, when $\tilde{\beta}^{-1} > \beta^{-1}$, temperature-accelerated molecular dynamics (TAMD) permits one to hyperthermally explore the free energy calculated at the physical temperature without distorting it. This requires adjustment of both the friction coefficient $\tilde{\gamma}$ and the coupling constant κ in Eq. **S1**. How to do so systematically will be discussed below.

Free-energy barrier estimation. Because the slow variables evolve at the fictitious temperature $\tilde{\beta}^{-1}$, points in CV space are visited in accordance with the Boltzmann factor $\exp[-\tilde{\beta}F]$. Barriers to transitions out of the local vicinity of the initial state are therefore easily overcome in finite time so long as they are less than $\tilde{\beta}^{-1}$. This suggests that we can in fact use TAMD runs at various values of $\tilde{\beta}^{-1}$ to roughly quantify free-energy barriers without actually going on to compute the free-energy surface.

Consider first that the rate α at which a barrier of height ΔF is crossed at temperature $\tilde{\beta}$ is given by

$$\alpha = \alpha_0\exp(-\tilde{\beta}\Delta F), \qquad \textbf{[S7]}$$

where α_0 is the frequency factor, or the number of attempted barrier crossings per unit time. This means that the probability that a dynamical trajectory begun at the state point from which the barrier height is measured has a probability of not yet having crossed this barrier in a finite time t given by

$$P(t,\tilde{\beta}) = \exp[-\alpha_0\exp(-\tilde{\beta}\Delta F)t]. \qquad \textbf{[S8]}$$

Assuming that the prefactor α_0 is independent of $\tilde{\beta}$, and noting that ΔF is also strictly independent of β (because TAMD determines F at the physical temperature β), when $\alpha_0 t$ is large, $P(t,\tilde{\beta})$ is a slowly varying function of $\alpha_0 t$ but a rapidly varying function of $\tilde{\beta}^{-1}$, which transitions sharply from 0 to 1 at a value of $\tilde{\beta}^{-1}$ given by

$$\tilde{\beta}^{-1} = \Delta F\frac{1}{\ln(\frac{\alpha_0 t}{\ln 2})}. \qquad \textbf{[S9]}$$

(This is found by determining $\tilde{\beta}^{-1}$ at which $P = 0.5$.) We show a plot of $P(\alpha_0 t,\tilde{\beta}^{-1})$ in Fig. S1. This relation suggests that we can bracket ΔF by means of estimating P by either 0 or 1 across TAMD simulations at various values of $\tilde{\beta}^{-1}$ but all of the same duration t. For instance, if we observe no transition in time t for a particular $\tilde{\beta}_1$, but we do observe a transition in the same time t for a different $\tilde{\beta}_2$, this would mean that with high probability

$$\tilde{\beta}_1^{-1}\ln\left(\frac{\alpha_0 t}{\ln 2}\right) < \Delta F < \tilde{\beta}_2^{-1}\ln\left(\frac{\alpha_0 t}{\ln 2}\right). \qquad \textbf{[S10]}$$

To make this relation useful, it remains to estimate the prefactor α_0, e.g., by monitoring the rate of barrier crossing attempts in the TAMD simulations. As explained above, this estimation can be rough because $\ln(\alpha_0 t)$ varies slowly in t when $\alpha_0 t$ is large.

As an example, consider the particular case of the rigid body twisting of the apical domain of GroEL, which for the 6 kcal/mol TAMD simulation is shown in the green trace of Fig. 2 in the main article. Measured using our internal coordinate (IC) definitions (see below), this angle transitions from about −30° to almost −120°. Following the trace closely, we see the transition seems hindered around −50° for the first 20 ns of the simulation and then suddenly transitions. If we infer the existence of a free-energy barrier against this motion and consider the trace a dynamical trajectory, it is clear that the system attempts to surmount this barrier O(10) times in the first 20 ns (as evidenced by the number of times the trace dips to about −60° and then rebounds), which implies that $\ln(\alpha_0 t) \approx 2$. Because this transition does not occur for $\tilde{\beta}^{-1} = 4$ kcal/mol but does occur for $\tilde{\beta}^{-1} = 6$ kcal/mol, we estimate the barrier to this transition is greater than about 8 kcal/mol but less than 12 kcal/mol.

This is, of course, a very rough estimate predicated on essentially a guess of the prefactor, α_0. More accurate barriers could be computed using a string method variant to determine the minimal free-energy path (1).

Preparation of simulation systems. Initial coordinates for the **t**-state GroEL subunit were extracted from chain A of the crystallographic data of the $(\text{GroEL-KMgATP})_{14}$ complex (2) [Protein Data Bank (PDB) code 1kp8]. The ATP molecule with its associated Mg^{2+} and K^+ ions was retained. Hydrogen atoms were inserted assuming pH 7.0. All molecular dynamics (MD) and TAMD simulations employed a time step of 2 fs and Langevin-based temperature control at 310 K with a coupling constant of 5 ps^{-1}. Long-range electrostatics were handled using particle-mesh Ewald summation with a grid spacing of 1 Å. Nonbonded interactions were cut off at 9 Å. Solvation using TIP3P water (3) was performed in boxes of size $107 \times 103 \times 116$ $Å^3$ (approximately 38,000 water molecules with 17 Na^+) for production runs and in boxes of size $81 \times 77 \times 90$ $Å^3$ (approximately 15,000 water molecules) for preequilibration and testing runs. We also created a system with the GroEL subunit initially in the $\mathbf{r}''$ crystallographic conformation, extracted from the $\mathbf{R}''$ ring (chains A–G) in the 1AON structure (4). When solvated and ionized, this box measured approximately $98 \times 104 \times 130$ Å. As a base case, we conducted a single traditional all-atom explicit-solvent MD simulation of the **t**-state GroEL subunit of duration 20 ns.

Initial coordinates of HIV-1 gp120 were extracted from chain G of the crystallographic data of the gp120-sCD4-Fab(17b) complex (5) (PDB code 1g9n). The solvent box was $82 \times 77 \times 80$ $Å^3$ (approximately 14,000 water molecules with 6 Na^+ and 8 Cl^-). Using the same simulation parameters as were used with GroEL, we conducted a single MD simulation of solvated gp120 of duration 20 ns.

Subdomains and composite ICs. To define subdomain memberships for a particular protein, we employ a gyration-radius–based clustering algorithm, similar in spirit to K-means clustering (6). Let the number of residues to be assigned to subdomains be N, and the number of subdomains to assign be m. The first N/m residues sequentially along the given segment(s) are assigned to subdomain 1, the second to subdomain 2, etc. Then, the radius of gyration of each subdomain is computed using Cartesian coordinates of Cα in a reference structure. We then enter a Monte Carlo loop in which the trial moves are swaps of residues between randomly selected pairs of subdomains. Any swap that lowers the average radius of gyration of the subdomains is accepted; otherwise, it is rejected. For our treatments of GroEL and gp120, 10^6 attempted swaps is adequate to determine optimally sized subdomains.

Residue memberships in each of the nine subdomains for GroEL and gp120 are reported in Tables S1 and S2, respectively. We adhered to a few guiding principles to define the subdomains. First, because the motion we seek to activate generally involves rotation and translation of domains, each domain requires minimally three subdomain centers. A minimal (i.e., rigid-body) but meaningful description of domain conformation in the context of a multidomain protein requires specification of nine variables; conventionally these are comprised of three for Cartesian coordinates of a center of mass, three for the lengths of the three principal axes, and three associated Euler angles, describe the orientation of the three axes in the lab frame. It is no less general to choose Cartesian coordinates of three distinct mapping points on a domain, which is what we have done here. Second, we rely on the knowledge (or consensus) of domain structure; for example, the GroEL subunit is comprised of three easily recognizable domains (equatorial, intermediate, and apical), and gp120 is comprised of two (inner and outer). Third, we require that subdomains are mutually exclusive and collectively exhaustive over all residues in a given protein.

It is true that we need to know something about the system in order to assign CVs, but no more than one needs to know to conduct standard MD, i.e., an initial configuration. One then postulates that domains move more or less as rigid bodies and asks what kinds of motion can be activated using modest fictitious temperatures. This to us seems fairly general and broadly applicable to many other systems, so long as one has an independent means to identify domains given a crystal structure.

In the specific case of gp120, we were interested to know if the interface between helices $\alpha 1$ and $\alpha 5$ in the inner domain was easily disrupted, so we purposefully defined the first subdomain as only $\alpha 5$ (residues 475–486) and then used the sorting algorithm to apportion remaining inner domain residues into a further eight subdomains. The outer domain was sorted directly into 14 subdomains.

The ICs monitored in the GroEL TAMD simulations are defined as follows (numbers refer to subdomains as listed in Table S1):

1. Equatorial-interfacial hinge angle: angle along 1–4–6
2. Interfacial-apical hinge angle: angle along 4–6–8
3. Apical twist angle: Dihedral along 4–6–8–7
4. Binding pocket angle: angle along 3–4–5.

It is worth pointing out that these composite ICs could have been accelerated using TAMD, rather than accelerating the Cartesian coordinates of the subdomain centers. We experimented with several scenarios in which these angles were indeed accelerated in early implementations of the method. We were generally unable to observe large-scale motion completely free of domain unfolding in these cases, which we attributed to the fact that very small changes in angle due to thermal fluctuations resulted in periodically large forces that led to larger-than-tolerable displacements of the centers. Interdomain dihedral angles proved especially troublesome because their values can fluctuate significantly and nearly uncontrollably when the two central points move only a little. We reasoned that, although they are convenient metrics to monitor conformation, such angles may not in fact be the best choice for CVs to accelerate using TAMD.

Determining the value of fictitious friction $\bar{\gamma}$ and spring constant κ. We demonstrate here that a fictitious friction $\bar{\gamma}$ of 50 ps^{-1} is adequate by computing the running average $G_j(N)$ of the restraining force for each CV during an MD simulation in which the CVs are restrained at their initial values:

$$G_j(N) = \frac{\kappa}{N} \sum_{i=1}^{N} [\theta_j^*(\boldsymbol{x}(t)) - \theta_j]. \qquad \textbf{[S11]}$$

We show $G_j(N)$ for all 27 CVs in the GroEL subunit in Fig. S2 from a restrained MD simulation in the **t**-state with $\kappa = 100$ kcal/mol · $Å^2$. $G_j(N)$ is seen to saturate for all CVs well before 25,000 steps; hence, a friction of 50 ps^{-1} is long enough to allow accurate self-averaging of local free-energy gradients.

The value of the spring constant κ should be sufficiently large such that $\theta_j^*(\boldsymbol{x}(t)) \approx \theta_j$ but not so large as to introduce numerical instabilities. To arrive at our value of κ, we considered earlier work using the related method of "steered MD" where evolution of an all-atom system is restrained such that one CV follows a predefined trajectory (7). In order to extract single-CV restricted free energies from such calculations, it is also a requirement that desired and measured CV values closely approximate one another. For a single CV, Park et al. (7) demonstrated this criterion was met when κ was between ≈7 and 500 kcal/mol · $Å^2$, guaranteeing mean deviations between desired and measured CVs of less than about 1 Å. We selected an intermediate value of 100 kcal/mol · $Å^2$. For $\bar{\beta}^{-1} = 6$ kcal/mol and $\kappa = 100$ kcal/mol · $Å^2$, we observe fluctuations in $\theta_j^*(\boldsymbol{x}(t)) - \theta_j$ of about 0.22 Å for GroEL and 0.20 Å for gp120. This check verifies that the restraints function properly, and demonstrates that the degree of precision in the relationship $\theta_j^*(\boldsymbol{x}(t)) \approx \theta_j$, being much less than 1 Å, is tolerable. We point out, however, that under perfect

adiabatic separation of the slow variables and collective variables, the fluctuation values reported above would be somewhat smaller.

Results and Discussion. ***TAMD simulations of the GroEL subunit.*** The majority of the MD and TAMD simulations of the GroEL subunit considered in the main article were initiated from configurations representing the equilibrium **t**-state crystallographic conformer [i.e., they were subunits excised from the 1KP8 crystal structure of Wang and Boisvert (2)]. For the TAMD simulations with fictitious temperatures $\tilde{\beta}^{-1}$ of 4 and 6 kcal/mol, several tens of nanoseconds of simulation resulted in conformations that closely resembled the crystallographic **r**″-state, coming to within 5 Å rms. Because the 6 kcal/mol TAMD simulation produced a fully rotated apical domain, we focus here on it. The conformations closest to the **r**″ conformer differed primarily in an approximately 30° "outward" tilt of the intermediate and apical domains relative to the equatorial—outward in the sense that the apical and intermediate domains would move away from the sevenfold axis were the subunit oriented in a heptamer ring. We show a detailed alignment of a TAMD-generated conformation that closely resembles the **r**″ conformer in Fig. S3. It is possible that this inward tilting requires the subunit to pay an energy penalty beyond what TAMD at $\tilde{\beta}^{-1} = 6$ kcal/mol can easily pay in a few tens of nanoseconds. In such a case, we would expect that the additional subunit–subunit interactions in a heptamer of GroEL might act to stabilize the high-energy inward-tilt of the **r**″ conformation. As a partial test of this hypothesis, we launched a 6 kcal/mol TAMD simulation beginning from an equilibrated **r**″ conformation [i.e., it was excised from the 1AON crystal structure (4) with ADP bound].

We observe in this simulation that the first major conformational change under TAMD is indeed an outward tilt of both the intermediate and apical domains to open the nucleotide pocket. At about only 3 ns of integration time, we observe that the conformations sampled converge to the most **r**″-like conformations of the previous TAMD simulations to within 2.3 Å of rms. We show snapshots overlaid in Fig. S4. This indicates that both TAMD simulations, launched from very different starting conformations, converge to essentially the same conformation: one that is **r**″-like in the upward tilt and rotation of the apical domain (60 and 90°, respectively) but displaying an open binding pocket due to an outward tilting of the intermediate and apical domains about the equatorial-intermediate hinge.

It is interesting to note the driving force for the **T**-to-**R**″ transition for a heptamer ring in the GroEL tetradecamer remains somewhat equivocal. It is not necessarily true that ATP hydrolysis is required for the transition, but it seems the binding free energy of ATP and cochaperonin GroES combine to stabilize **R**″. This is evident from the fact that GroES increases the Hill coefficient (degree of cooperativity) of the sevenfold hydrolysis reaction in a ring (8), and it is difficult to see how this can be true if hydrolysis is required to drive the ring to the state in which GroES can bind in the first place. There is therefore some support for the speculation that the GroEL subunit visits **r**″-like conformations with ATP bound prior to ATP hydrolysis. The rate at which it visits such conformations might be extremely low, and might also be enhanced by "activation" via release of the binding free energy of ATP.

Now, considering the transitions we observe, it must be the case that activating the translation of subdomain centers (which activates interdomain reorientation) drives the transitions because without activation the transitions do not occur. It is not a problem that the **t**-state monomer with ATP bound is driven to visit the **r**″ state because, as we reason above, this very transition must occur under normal conditions for the tetradecamer. But we must stress that there is no transduction of binding free energy from ATP in the simulation; ATP is already present in the binding pocket and the system is thermalized before beginning the calculations. The ATP is merely "along for the ride" in the **t**-state–initiated TAMD simulations, and it is the explicit activation of interdomain motion provided by TAMD that drives the transition. Another equivalent way of thinking about how TAMD drives the transition is that TAMD allows for rapid exploration of conformations (i.e., points in CV space) that are thermally accessible given the user's choice of the fictitious temperature, $\tilde{\beta}^{-1}$. At this point, it is also important to note that the activation of interdomain reorientation by TAMD is strictly unbiased, and the random kicks provided by the TAMD forces to the subdomains are apparently rectified along directions that yield productive transitions in finite time simulations. Considering the newer simulation beginning from the **r**″-state with ADP bound, the situation is analogous; the ADP is not an active part of the transition.

Counterrotation of inner and outer domains of gp120. Out of the four 4-ns TAMD trajectories at $\tilde{\beta}^{-1} = 6$ kcal/mol, three resulted in domain–domain motion significant enough to break the hydrogen bonds connecting the $\beta 2/3$ and $\beta 20/21$ halves of the bridging sheet. These trajectories displayed qualitatively similar domain–domain motion, characterized by clockwise rotation (when viewed along an inner-to-outer domain axis) of the outer domain by a maximum of about 30° exposing almost $1,000$ Å^2 of interdomain buried surface area. We illustrate this motion in Fig. S5*A*, which shows three gp120 conformations aligned along inner domain: the 1g9n crystal structure and two samples from the TAMD trajectories (postequilibrated for a further 4 ns with standard MD). The close overlap of inner domains confirms the fact that TAMD is not disrupting domain structure, but instead inducing domain–domain motion, here indicated by the clockwise rotation of the outer domain (ochre-to-red "marshmallow" renderings). In Fig. S5*B* this motion is presented schematically.

A second common feature in these conformational changes is the direct contact between the hairpin turn on $\beta 20/21$ and the C-terminal end of $\alpha 1$ with the $\beta 2/3$ stem some 15 Å away from a position that would allow crystallographic registry of the $\beta 2/3$–$\beta 20/21$ hydrogen bonds. In other words, direct interaction between inner and outer domain appears to stabilize a conformation incommensurate with bridging sheet formation due to a rotational displacement of inner domain with respect to outer domain in the unliganded structure.

Comparisons among unliganded gp120 structures. Here, we present comparisons between one of the TAMD-generated unliganded configurations of HIV-1 gp120 and two experimentally derived unliganded structures: The first is the low-resolution EM map of the native unliganded envelope spike (9) and the second is the unliganded crystal structure of the closely homologous simian immunodeficiency virus (SIV)-1 gp120 (10). In Fig. S6*A*, we show alignments of our predicted unliganded gp120 structure into the EM map of unliganed trimer. These alignments were performed by simple rmsd minimization against the gp120 structures already fit into the maps. The figure shows a surface for one particular density threshold used in Fig. 2*E* of the EM work (9). The two major conclusions from this fitting are that (*i*) the bulk of the core (inner and outer domain) fits well into the map, and (*ii*) the $\beta 2/3$ hairpin bridges monomers. It has been shown that gp120s engineered to lack V1/2 result in nonfunctional spikes (11), leading us to speculate that V1/2 does in fact mediate essential inter-gp120 contacts in trimer.

In contrast, our unliganded HIV-1 gp120 model differs from the unliganded SIV structure (10) in several ways. We show an alignment of these two structures in Fig. S6*B*. The outer domain structures are nearly superimposable, except for suface loops, as would be expected, because it is already well-appreciated that the SIV outer domain superimposes on the CD4-bound HIV-1 gp120

outer domain, and our TAMD method does not significantly perturb intradomain structure. However, the inner domains and bridging sheet elements ($\beta 2/3$ and $\beta 20/21$) are drastically different between the two. As a particular example, the SIV structure shows the tip of the $\beta 20/21$ interacting with the N-terminus of $\alpha 1$, whereas our model places it at the bottom. Also, our $\beta 2/3$ protrudes away from the body of the protein, whereas in Harrison's structure the $\beta 2/3$ lies almost along the $\alpha 1$ axis against $\alpha 1$, placing its turn (like that of $\beta 20/21$) near the N-terminus of $\alpha 1$. Interestingly, however, it is apparent that much of the discrepancy in the orientation of inner domain helices $\alpha 1$ and $\alpha 5$ relative to outer domain in these two structures would vanish if one were to rotate the SIV inner domain counterclockwise by approximately 45° (with respect to the view in Fig. 5*A* in the main paper and Fig. S6*B* here). Note, however, this does not speak to the very different conformations of the bridging sheet elements, and the relative orientations of the two helices $\alpha 1$ and $\alpha 5$ in the SIV structure is significantly different than in our model or the existing HIV-1 gp120 structures. Nevertheless, taken together these findings hint at the strong likelihood that large-scale conformational flexibility of the two domains of gp120 involves counterrotation of the inner and outer domains.

Based on these assessments, we believe our unliganded model agrees reasonably well with unliganded spike EM map of Liu et al (9). The outer domain match between our model and SIV structure (10) is unsurprising, but with the realization that much of the mismatch seems to be attributable to counterrotating domains, it is possible we have uncovered a general gp120 conformational change mechanism. We point out that, while this contribution was under review, a crystal structure of the gp120 core including the N- and C-terminal gp41-interactive regions was published (12). In this article, the authors propose a gp120 conformational change mechanism that may be characterized by motion of three individual "layers" of inner domian, each of which is a loop that emanates and returns to the β-sandwich distal to the bridging sheet. Each layer-loop is of a strand-helix-strand topology, with helix $\alpha 0$ in layer 1, helix $\alpha 1$ in layer 2, and helix $\alpha 5$ in layer 3. $\alpha 1$ and $\alpha 5$ are resolved in all gp120 crystal structures in the CD4-bound conformation to date, and the structure involving layer 1 ($\alpha 0$) is new. We are currently testing the "layer-motion" hypothesis of Pancera et al. (12) using new TAMD simulations.

Implemenation issues. TAMD requires updating the slow variables θ each time step of a running MD simulation and communicating the resulting restraining forces back to the atoms. The latter of these two requirements was met using the *tclforces* interface to freely available protein simulation package NAMD to implement TAMD. This interface allows the NAMD user to predefine an arbitrary number of selections of atoms whose position data is accessible and to whom additional forces can be communicated at each MD time step. The former requirement was met using a custom-written combination of TcL scripts and C code.

This implementation carries an overhead of a few percent, at most. We did not make a systematic study, but on runs on up to 64 cores simultaneously, the master NAMD process does idle a bit more than the slave processes as coordinate data is copied into and out of the shared-object data space. There is essentially no cost associated with the update of slow-variable values and forces because in number these are many times smaller than the number of atoms the MD kernel handles.

Because our slow variables are Cartesian coordinates, they have units of Å, as do all positional variables in NAMD. We retain the NAMD conventions of computing forces in units of $\text{kcal/mol} \cdot \text{Å}^2$, but use a time unit of ps for the slow-variable evolution. This is not a problem because we use the value of the MD timestep in units of ps (0.002) when updating the slow variables and we measure the fictitious friction $\bar{\gamma}$ in units of ps^{-1}. We are not interested here in the true dynamics of the slow variables but only on sampling CV space according to the equilibrium free-energy surface; hence, mass (within the limits of computational stability) is irrelevant. Using the unit system described here, unit mass is equivalent to approximately 400 atomic mass units (amu) (i.e., $1\ \text{kcal} \cdot (\text{mol})^{-1} \cdot \text{ps}^2 \cdot \text{Å}^{-2} \approx 418$ amu).

1. Vanden-Eijnden E (2009) Some recent techniques for free energy calculations. *J Comput Chem*, 30:1737–1747.
2. Wang J, Boisvert DC (2003) Structural basis for GroEL-assisted protein folding from the crystal structure of $(\text{GroEL-KMgATP})_{14}$ at 2.0 Å resolution. *J Mol Biol*, 327:843–855.
3. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 79:926–935.
4. Xu Z., Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-$(\text{ADP})_7$ chaperonin complex. *Nature*, 388:741–750.
5. Kwong P, Wyatt R, Majeed S, Robinson J, Sweet R, Sodroski J, Henderickson W (2000) Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Struct Fold Des*, 8:1329–1339.
6. Forgy E (1965) Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications. *Biometrics*, 21:768–780.
7. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys*, 119:3559–3566.
8. Gray TE, Fersht AR (1991) Cooperativity in ATP hydrolysis by GroEL is increased by GroES *FEBS Lett*, 292:254–258.
9. Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S (2008) Molecular architecture of native HIV-1 gp120 trimers *Nature*, 455:109–U76
10. Chen B, Vogan EM, Gong HY, Skehel JJ, Wiley DC, Harrison SC (2005) Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*, 433:834–841.
11. Bontjer I, Land A, Eggink D, Verkade E, Tuin K, Baldwin C, Pollakis G, Paxton W, Braakman I, Berkhout B, Sanders RW (2009) Optimization of human immunodeficiency virus type 1 envelope glycoproteins with V1/V2 deleted, using virus evolution. *J Virol*, 83:368–383.
12. Pancera M, et al. (2010) Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc Natl Acad Sci USA*, 107:1166–1171.

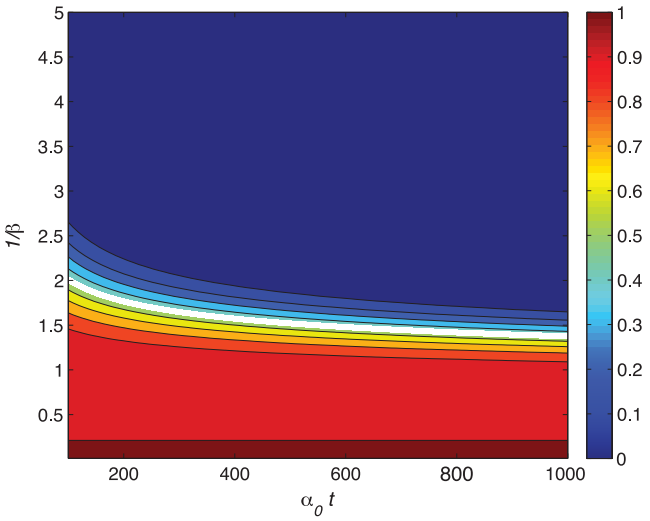

Fig. S1. A map of $P(t, \bar{\beta})$ (Eq. **S8**). The white curve is Eq. **S9**.

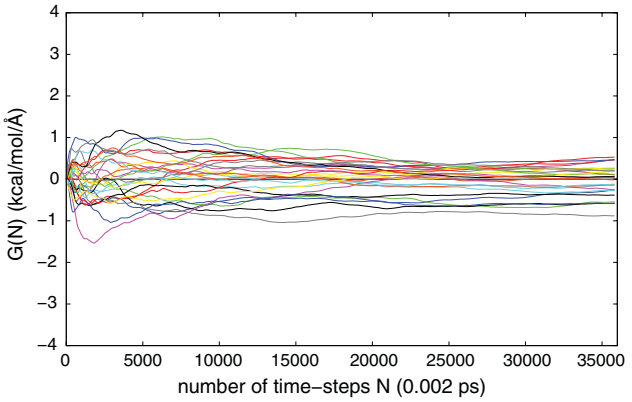

Fig. S2. Running time-averaged restraining force for fixed CVs for the all-atom, explicitly solvated GroEL subunit.

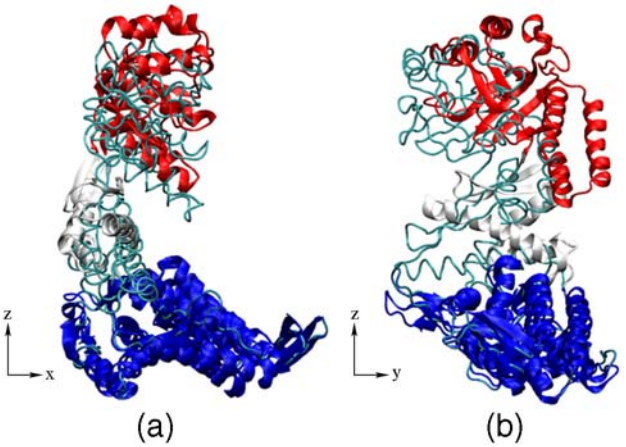

Fig. S3. Equatorial-domain–aligned overlays of a TAMD-generated conformer (blue, equatorial domain; white, intermediate domain; red, apical domain) and a single GroEL subunit from the R″ ring of the 1AON crystal structure (cyan "tube") (4). The overall rmsd of the two conformers shown here is 5 Å.

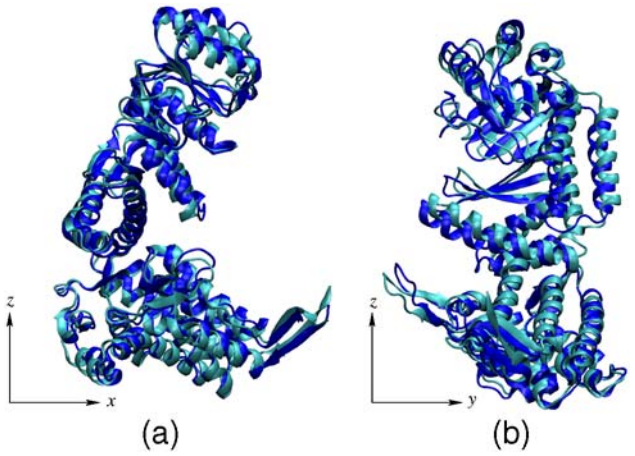

Fig. S4. Aligned overlays of TAMD-generated conformers of the GroEL subunit viewed from two orthogonal directions. The cyan conformer was generated from approximately 30 ns of TAMD with $\tilde{\beta}^{-1} = 6$ kcal/mol beginning from an equilibrated **t**-state conformation, and the blue conformer was generated from approximately 3 ns of TAMD with $\tilde{\beta}^{-1} = 6$ kcal/mol beginning from an equilibrated **r**″ conformation. The two conformers agree to within 2.3 Å rms, and both are about 5 Å rms from the **r**″ conformation.

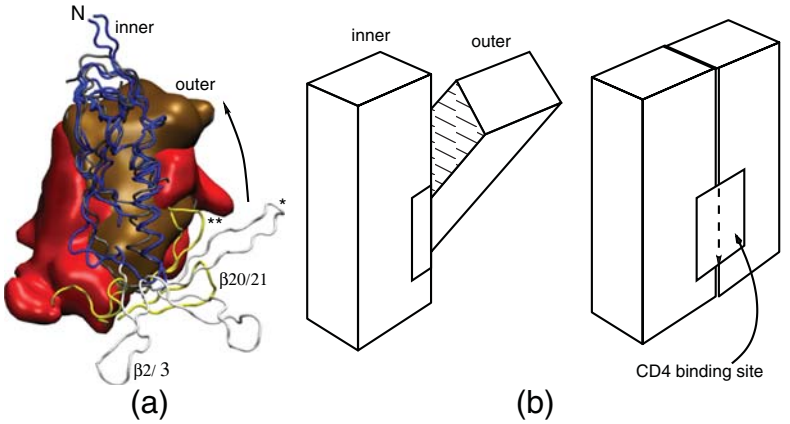

Fig. S5. (*A*) Gp120 conformations aligned along inner domain residues [blue tubes; gray is the 1g9n crystal structure (5)]. Outer domain rendered in marshmallow showing relative orientations in liganded (ochre) and unliganded (red) conformations. Bridging sheet elements are shown in white (β2/3) and yellow (β20/21), with asterisks (*) denoting conformations in the 1g9n crystal structure. (*B*) Schematic rendering of the unliganded-to-liganded conformation predicted for gp120 from TAMD simulations.

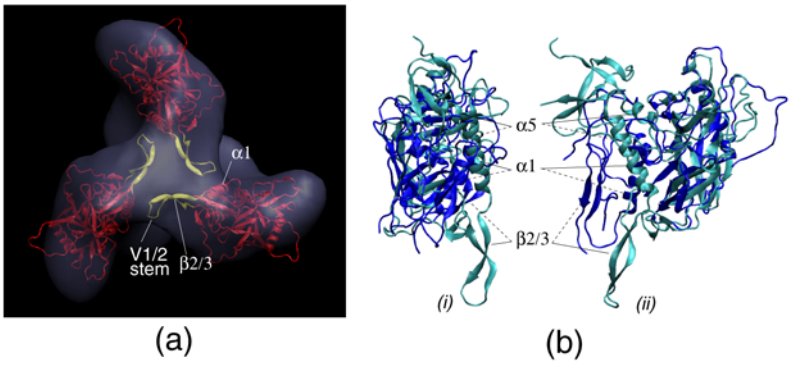

Fig. S6. (*A*) Alignment of our unliganded HIV-1 gp120 into the EM map of unliganded gp120 trimer, formatted for direct comparison to Fig. 2*E* of Liu et al. (9). (*B*) "Front" view of our unliganded HIV-1 gp120 model (cyan) with outer domain aligned against the unliganded SIV gp120 structure of Chen et al. (10) (blue); inner domain is in front and outer domain in back. (*B*) Side view of same alignment.

Table S1. Subdomain memberships in the GroEL subunit

Subdomain	Mass, kDa	Residues
1	7.79	2–8, 11, 12, 14–27, 38–50, 54–80, 82, 83, 520–526
2	7.37	9, 10, 13, 96–129, 426–445, 505–519
3	7.42	28–37, 51–53, 81, 84, 86–95, 415, 416, 419, 420, 422–424, 446–473, 476–486, 491, 503, 504
4*	5.54	85, 130–140, 142–144, 146, 147, 150, 151, 171, 401–403, 405–414, 417, 418, 421, 425, 474, 475, 487–490, 492–502
5	5.19	141, 145, 148, 149, 152–164, 167, 177–186, 379–400
6†	6.45	165, 166, 168–170, 172–176, 187–195, 220, 291–297, 319–323, 331–346, 350, 367–378, 404
7	4.94	198–218, 244–246, 248, 256, 259–275, 324–327
8	5.03	219, 221–223, 226–243, 247, 249–251, 254, 255, 257, 258, 298, 299, 301, 303, 306–318
9	5.40	196, 197, 224, 225, 252, 253, 276–290, 300, 302, 304, 305, 328–330, 347–349, 351–366

*Equatorial-intermediate hinge subdomain.
†Intermediate-apical hinge subdomain.

Table S2. Subdomain memberships in HIV-1 gp120

Subdomain	Mass (kDa)	Residues
1	1.64	475–486
2	1.65	92–97, 231–238
3	1.64	98–110, 253
4	1.51	113, 117–121, 200–207
5	1.30	122–129, 194–199
6	1.65	111, 112, 114–116, 208–213, 252, 254, 255
7	1.48	220–226, 244, 487–492
8	1.52	86–91, 228–230, 239–243
9	1.55	214–219, 227, 245–251
10	1.27	259–263, 291–293, 447–451
11	1.49	272–279, 281–285
12	1.45	264–271, 286–290
13	1.34	294–297, 299, 329–331, 417, 418, 444–446
14	1.55	337–347, 360, 395
15	1.46	348–359, 464
16	1.21	363–369, 371–373, 386–388
17	1.70	370, 374–377, 382–385, 419–422
18	1.67	361, 362, 389–394, 396–398, 407, 408
19	1.42	332–336, 409–416
20	1.56	423–435
21	1.24	298, 378–381, 436–443
22	1.30	256–258, 452–455, 469–474
23	1.45	280, 456–463, 465–468