

**Solvation dynamics-powered structure and function of multi-molecular cellular systems exemplified by non-equilibrium cereblon-degrader-CK1 $\alpha$  ternary complex formation**

Hongbin Wan, Vibhas Aravamuthan, Sarah Williams<sup>‡</sup>, Charles Wartchow<sup>‡</sup>, José S. Duca,  
Robert A. Pearlstein

Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, MA 02139

<sup>‡</sup>Novartis Institutes for BioMedical Research, 5959 Horton Street, Suite 900, Emeryville, CA 94608

Corresponding author: Robert A. Pearlstein, Ph.D.

Phone: +1 617-871-7293

Email: [robert.pearlstein@novartis.com](mailto:robert.pearlstein@novartis.com)

Keywords: Non-equilibrium, kinetics, solvation free energy, solvation field, non-linear dynamics

## Abstract

Cellular functions are executed via a form of analog computing that is based on the switchable covalent and non-covalent states of multi-molecular fluxes (i.e., time-dependent species/state concentrations) operating in the non-linear dynamics regime. We and others have proposed that the non-covalent states and state transitions of aqueous fluxes are powered principally by the storage and release of potential energy to/from the anisotropic H-bond network of solvating water (which we refer to as the “solvation field”), which is a key tenet of a first principles theory on cellular structure and function (called Biodynamics) that we outlined previously. This energy is reflected in water occupancy as a function of solute surface position, which can be probed computationally using WATMD software. In our previous work, we used this approach to deduce the structural dynamics of the COVID main protease, including substrate binding-induced enzyme activation and dimerization, and product release-induced dimer dissociation. Here, we examine:

- 1) The general relationships between surface composition/topology and solvation field properties for both high and low molecular weight (HMW and LMW) solutes.
- 2) The general means by which structural dynamics are powered by solvation free energy, which we exemplify via binding between the E3 ligase CUL4A/RBX1/DDB1/CRBN, LMW degraders, and substrates. We propose that degraders organize the substrate binding surface of cereblon toward complementarity with native and neo substrates, thereby speeding the association rate constant and incrementally slowing the dissociation rate constant.
- 3) Structure-activity relationships (SAR) based on complementarity between the solvation fields of cognate protein-ligand partners exemplified via LMW degraders.

## Introduction

The equilibrium state distributions of biomolecular species present in the in vitro regime are widely assumed to extend to the native cellular regime. However, many cellular functions depend on time- and space-dependent variation in the concentrations/number densities of their constituent molecular and ionic species (which we refer to as “fluxes”), where equilibrium principles and metrics (e.g.,  $\Delta G$ ,  $IC_{50}$ ,  $K_d$ ) break down. Molecular fluxes can be described symbolically as:

$$dC_{ik}(t)/dt = \text{buildup rate} - \text{decay rate} \quad [1]$$

where  $C$  is the time-dependent concentration of a given molecular species  $i$  within a given compartment  $k$  that builds and decays at specific rates (which we refer to as “extrinsic rates”) that are distributed over one or more intramolecular, intermolecular, and covalently modified states according to their individual rates of buildup and decay (which we refer to as “intrinsic rates”).<sup>1</sup> The most populated states are those exhibiting similar intrinsic and extrinsic buildup and decay rates.<sup>1</sup> The equilibrium regime is the special case in which the species concentrations are fixed,  $dC(t)/dt = 0$  and the buildup and decay rates are equal. We have developed a multi-scale first principles, physics-based theory called Biodynamics describing the general means by which cellular behaviors are conveyed by dynamic multi-molecular and multi-ionic fluxes operating in the non-linear dynamics (NLD) regime (where the behavior of the whole system  $\gg \Sigma(\text{behaviors of the parts})$ ).<sup>1,3,4</sup> Our theory carries major implications for:

- 1) Molecular biology, including protein-protein interactions and the relationship between mutation-induced protein dysfunction and cellular dysfunction/disease.
- 2) Cell biology, including the means by which cellular functions are conveyed by multi-flux systems.

- 3) Drug R&D, including structure-activity relationship (SAR) analysis of low molecular weight (LMW) analogs, pharmacokinetic-pharmacodynamic (PK-PD) analysis, and structure-based drug design.

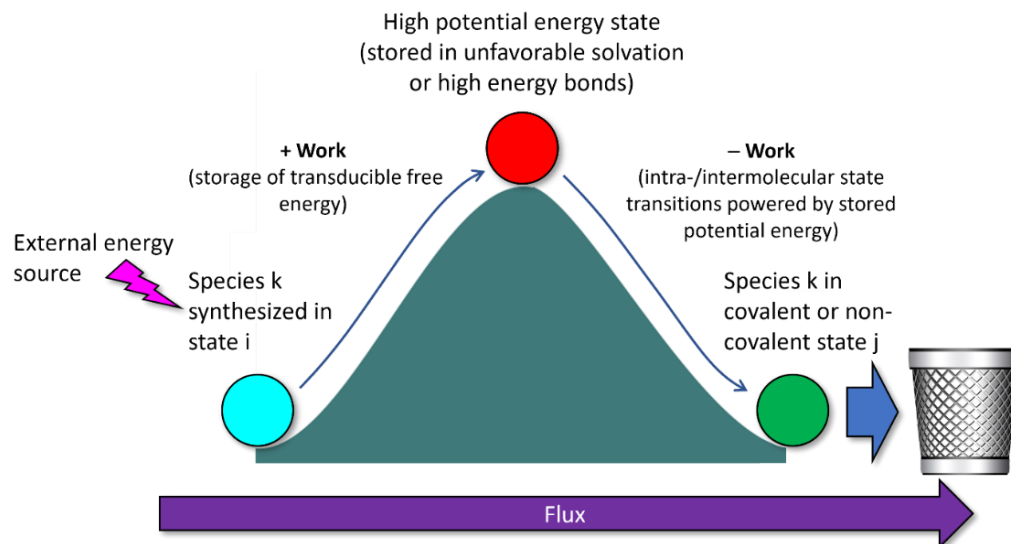
In our previous work, we applied Biodynamics principles and WATMD software (a first principles method for probing water occupancy across solute surfaces) to qualitatively deduce the solvation free energy-powered non-covalent non-equilibrium intra- and intermolecular state transitions of the COVID main protease ( $M^{pro}$ ) and dynamic blockade of the human ether-a-go-go related gene potassium channel (hERG).<sup>4,5</sup> In this work, we:

- 1) Briefly review the basic properties of fluxes and the common architecture of multi-flux systems exemplified by the cereblon-mediated proteasomal protein degradation pathway (the sink for many protein fluxes).
- 2) Overview our ongoing progress in deciphering the general mechanisms by which solvation free energy is transduced into the intrinsic rates of non-covalent intra- and intermolecular state transitions (vis-à-vis the status quo paradigm) under aqueous non-equilibrium conditions, which we extend to protein folding and the low molecular weight (LMW) regime.
- 3) Expand the application of our theory and WATMD to deduce the solvation free energy-powered non-equilibrium binary and ternary states of cereblon (the substrate-recruiting subunit of the CUL4A/RBX1/DDB1/CRBN E3 ligase) bound to native and synthetic LMW degraders and neo substrates.

## The moving parts of cellular systems consist of molecular and ionic fluxes multiplexed into non-linear dynamic systems

Whereas molecules and ions constitute the tangible moving parts (i.e., fluxes) of cellular systems, the intangible behaviors of such systems emerge from NLD operation. Cellular systems respond to perturbations via acceleration or deceleration of specific fluxes and/or the state transitions thereof.<sup>2</sup> Molecular fluxes are generated in a source  $\rightarrow$  sink fashion powered by transducible covalent free energy sources (e.g., ATP) and undergo dynamic solvation-powered non-covalent intra- and intermolecular state transitions (including enzyme-substrate binding) throughout their lifetimes (Figure 1A). Protein fluxes, the workhorses of cellular function, are produced via transcription and translation and cleared via proteasomal, lysosomal and other degradation mechanisms (Figure 1B) (noting that fast fluxes depend on activation of pre-synthesized protein or zymogen pools). The lifetimes of proteins that decay via proteasomal degradation are governed by the relative rates of poly-ubiquitylation versus deubiquitylation catalyzed by E3 ligases and deubiquitylases (DUBs), respectively (and possibly other mechanisms) that we address later in this work.

A



B

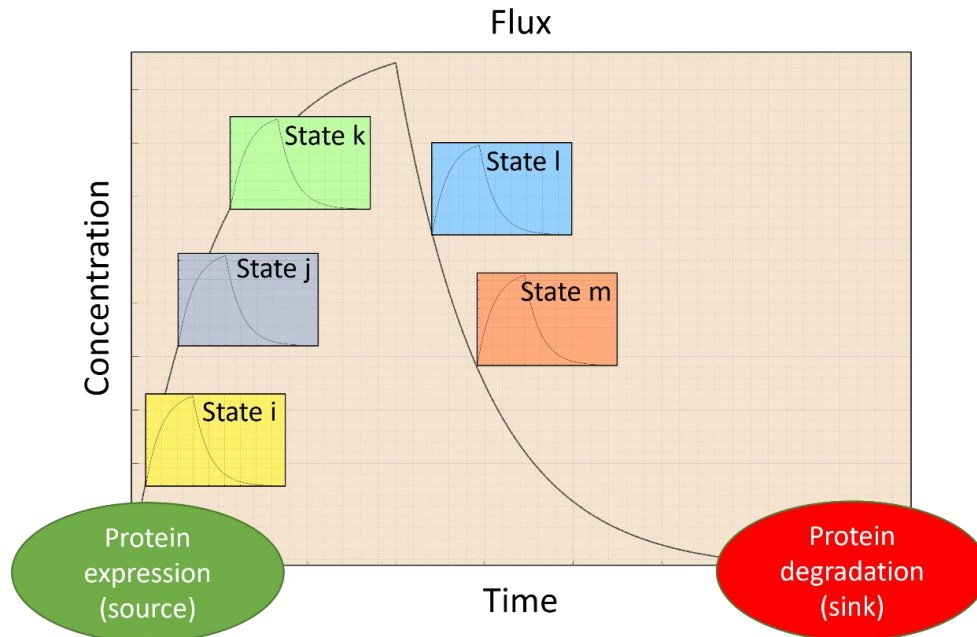


Figure 1. (A) Fluxes are powered by various forms of transducible potential and kinetic energy, beginning with synthesis or inward intercompartmental translocation, storage of transducible free energy forms (including unfavorable solvation and high energy bonds), state transitions, and degradation or outward intercompartmental translocation. (B) Protein fluxes build via expression and decay via degradation in a source-sink fashion and undergo a range of covalent and non-covalent solvation free energy-powered intra- and intermolecular state transitions that build and decay throughout their lifetimes.

The states of each flux are populated dynamically according to their rates of buildup and decay, the physical form of which we refer to as “molecular differential equations” (MDEs), and the canonical symbolic form of which is given in equation 1. Fluxes exist in one or more intra- and intermolecular states that build and decay over time, analogous to the flow of fluid to/from the fixtures of a plumbing system, which depend on the rates of entry and exit to/from the system

(analogous to the extrinsic flux rates) and the rates of entry and exit to/from each fixture (analogous to the intrinsic flux rates).<sup>2</sup> The non-covalent and enzyme-mediated covalent states of a given flux necessarily build and decay on the overall timescale of the cellular function in which they participate (e.g., the open, inactivated, and closed states of cardiac ion channels build and decay on the ~350 ms timescale of the action potential<sup>6,7</sup>). Cellular behaviors consist of the “physical solutions” of networked time- and space-dependent multi-molecular or multi-ionic fluxes (Figure 2), which we refer to as “physical math”.<sup>2</sup> Cellular systems comprised of three or more multiplexed MDEs operate in the NLD regime, giving rise to emergent behaviors that are  $\gg \Sigma$ (behaviors of the moving parts). NLD systems lack analytical solutions and must be solved computationally via symbolic mathematical simulations or physically via analog computing in the case of living cells (where the solutions/outputs consist of various dynamic behaviors). The “cellular operating system” (CellOS) is thus analog, wherein the hardware and software are one and the same. From this perspective, fluxes act as analog signal carriers, analogous to amplitude and frequency modulated (AM and FM) radio waves. Cellular behaviors are initiated by intra- and/or extracellular input signals that propagate through and switch the states of the corresponding network via alteration of one or more extrinsic and/or intrinsic flux rates.<sup>2</sup> Abnormal genetic mutation-induced alterations of these rates constitutes the fundamental basis of cellular disease (e.g., slowed protein synthesis has been shown to play a role in Huntington’s disease<sup>8</sup>). Molecular and ionic fluxes build and decay exponentially in time (i.e., the solutions of both the physical and symbolic forms of the MDEs), which has crucial implications for cellular pathway architecture<sup>1-3</sup> (Figure 3). Counter-balanced flux-anti-flux pairs that we refer to as “Yins” and “Yangs” are necessitated by such behavior.

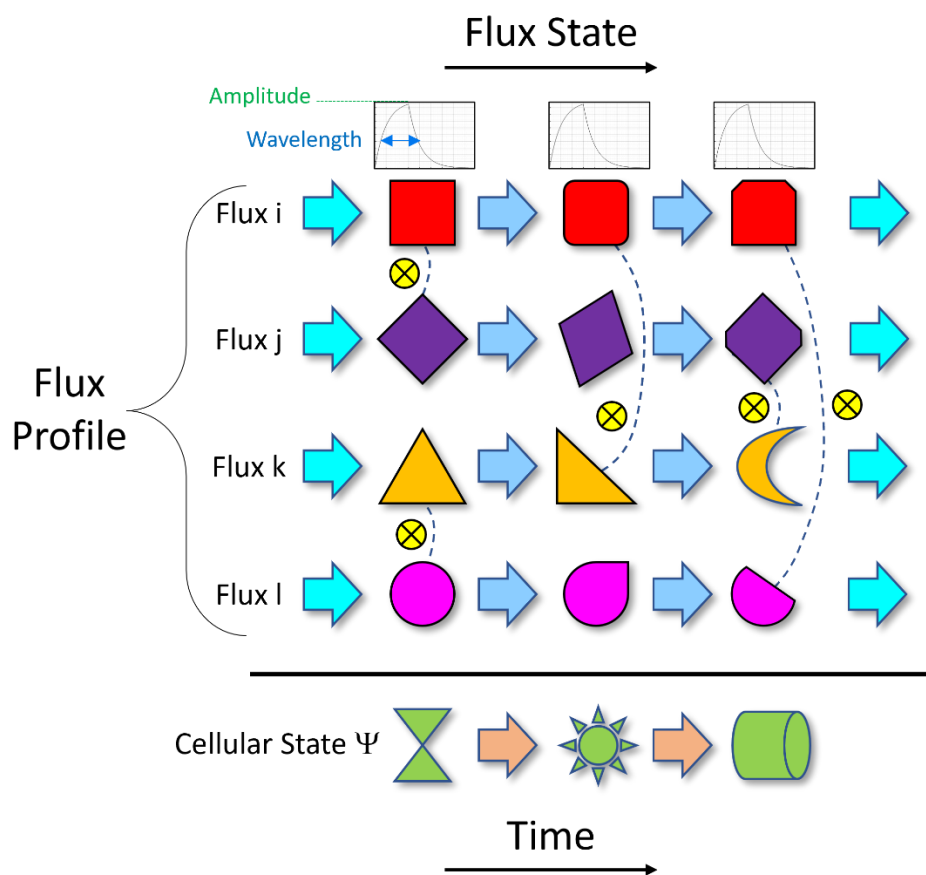


Figure 2. Time-dependent cellular states are defined by the dynamic state transitions of their constituent multi-flux systems, which are subject to a variety of transient perturbations/signals (typically initiated by stimulation of intracellular and extracellular receptors). Each participating flux (denoted by the geometric shapes along the time axis) encodes a signal in the amplitudes and wavelengths of its covalent and non-covalent intra- and intermolecular states, intrinsic and extrinsic rates, and state-specific inter-connections (analogous to a switchboard, denoted by dotted lines). Non-linear behaviors result from interdependence between the dynamic states and multiplexed configurations of each multi-flux system.



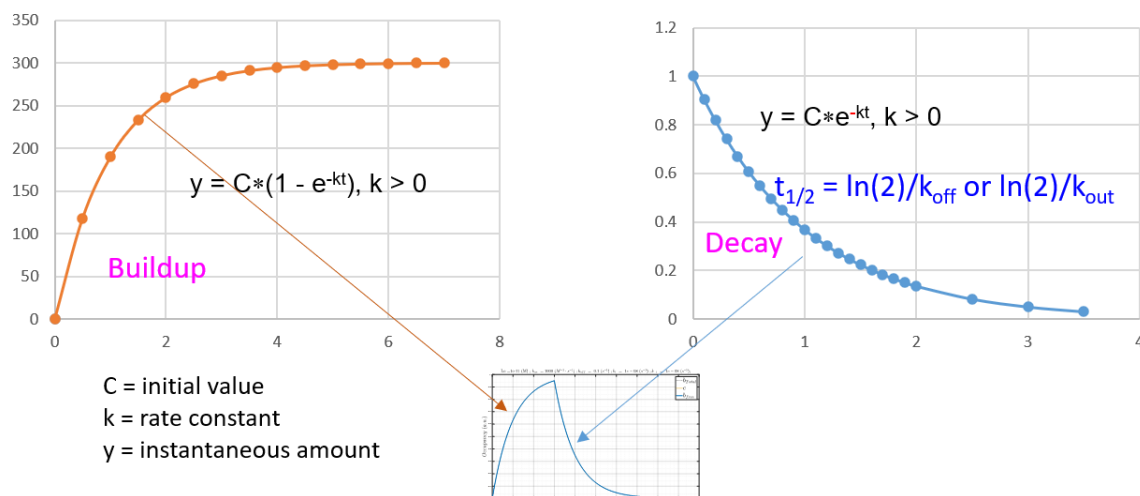


Figure 3. Fluxes and their populated states build and decay exponentially in time.

The instantaneous net rate of product formation in the case of enzyme Yin-Yang pairs (e.g., kinase-phosphatase, ligase-ubiquitinase) depends on the relative Yin and Yang activities, which operate in a dynamic push-pull fashion (a molecular tug-of-war) governed by the relative Yin and Yang flux rates/amplitudes, together with the phasing of the Yang buildups relative to the corresponding Yin-catalyzed products (Figure 4). Yang buildups, in turn, are driven by de novo expression or activation by downstream species (e.g., phosphatase activation by a downstream kinase), which we refer to as “Clocks”. The Yin-Yang balance is tipped in the forward and reverse directions by slow and fast Clocks, respectively. The states of each Yin-Yang-Clock module are governed by the rates of entry and exit to/from the available intra- and intermolecular states of the module.

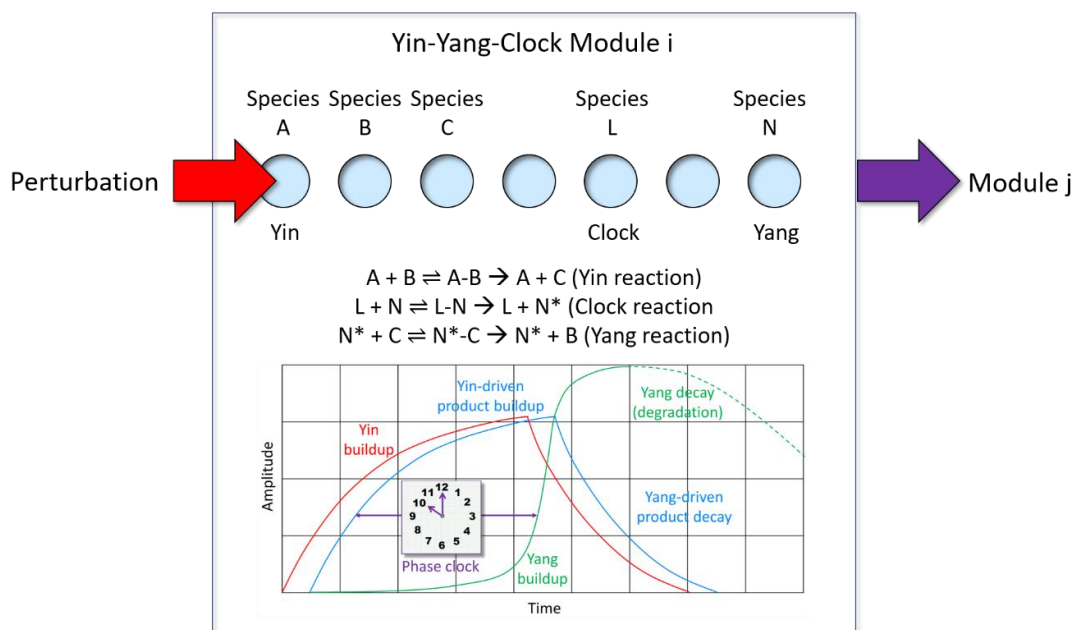


Figure 4. Cellular pathways are comprised of sets of sequential enzyme-catalyzed reactions (fluxes) among the participating species (blue filled circles, denoted as A, B, C, etc.). Pathways are putatively organized into sets of Yin-Yang-Clock modules (exemplified here for a hypothetical module i), in which the Yin, Yin product, Clock, and Yang, consist of species A, B, L, and N, respectively. The forward and reverse fluxes are maintained in an out-of-phase relationship via the Clock-dependent Yang buildup (green curve) that lags behind that of the corresponding Yin-catalyzed product (blue curve). The Yin-Yang balance is respectively tipped toward the Yin and Yang during the product buildup and decay phases.

## **The status quo concept of structure-free energy relationships is challenged by Biodynamics theory**

Molecular dynamics (MD) simulations, free energy perturbation (FEP), docking scores, and other force-field-based methods have been used extensively for studying the intra- and intermolecular structures and state transitions of proteins and other biomolecules.<sup>9,10,19–22,11–18</sup> However, the widely held view that non-covalent molecular structure and function are powered by van der Waals, electrostatic, H-bond, hydrophobic, hydrophilic,  $\pi$ -stacking,  $\pi$ -cation, and other types of interatomic energy contributions is being increasingly questioned by our findings<sup>2–4,23,24</sup> and those of others.<sup>25</sup> Mounting evidence suggests that, under both static equilibrium and dynamic non-equilibrium conditions, the energetically preferred sterically-accessible structural states of aqueous HMW and LMW biomolecular solutes and non-covalent HMW-LMW and HMW-HMW complexes are powered principally by solvation free energy, and in particular, solvation H-bond free energy relative to that of bulk solvent. Molecular structures reside within a Goldilocks zone of favorable and unfavorable free energy contributions comprised of:

- 1) Stabilizing/H-bond enriched solvation (“cold spots”), the free energy of which is significantly exceeded by that of bulk solvent.
- 2) Destabilizing/H-bond depleted solvation (“hot spots”), the free energy of which significantly exceeds that of bulk solvent.

Overly H-bond enriched states are non-rearrangeable (e.g., rigid protein structures), whereas overly H-bond depleted states are poorly or randomly populated (manifesting as disorder, and in extreme cases, precipitation or aggregation). Furthermore, the barriers to/from a given non-covalent state (denoted as  $\Delta G_{entry}^{\ddagger}$  and  $\Delta G_{exit}^{\ddagger}$ , respectively) consist predominantly of the local

desolvation and resolution costs incurred respectively in entering and exiting that state (i.e.,  $\Delta G_{desolv}^{\ddagger}$  and  $\Delta G_{resolv}^{\ddagger}$ ), rather than  $\Delta G$  per se.<sup>3,4,23,24,26</sup> Interatomic contact energies are overestimated by empirical force-field (FF) and quantum mechanics (QM)-based methods relative to the H-bond free energy of liquid water, consistent with the vastly higher melting and boiling points of water per unit molecular volume compared with those of liquid organic substances.<sup>3</sup> The predominant contribution of solvation free energy to  $\Delta G$  is therefore underestimated in classical interatomic contact models, resulting in:

- 1) Poor structure predictions based on calculated free energies.
- 2) Inapplicability to the non-equilibrium cellular setting.
- 3) Incorrect or suboptimal interpretation of structure-free energy relationships derived from measured data (e.g., kinetic barriers and rate constants measured via surface plasmon resonance,  $\Delta G$  measured via isothermal titration calorimetry, and biochemical assay data).

Pure liquid bulk water behaves as a high density, isotropic H-bond network to which each water molecule contributes an average of ~3.5 moderately strong H-bonds.<sup>3</sup> The H-bond network of liquid water resides within a Goldilocks zone of fluidity and strength, behaving as a quasi-solid when subjected to high frequency perturbations. Water minimizes its enthalpic and entropic free energy contributions when perturbed by HMW and LMW solutes, air, surfaces, cells, etc. relative to those of bulk solvent by:

- 1) Altering its rate of entry/exit to/from water-non-water interfaces and bulk solvent.
- 2) Increased proton dissociation.<sup>3</sup>

- 3) Ordering into a wide range of non-isotropic two- and three-dimensional H-bond network geometries in conformance with the local water-solute interfacial geometry.
- 4) Driving non-covalent solute rearrangements (including protein folding and binding) toward the minimum solvation H-bond free energy, which is incorrectly attributed under the conventional wisdom to hydrophilic, hydrophobic, and other types of interatomic interactions.

Mounting evidence suggests that the primary free energy source powering solute structure and function consists of solvation free energy (water H-bond free energy, in particular), rather than interatomic contacts.<sup>1,8,24–26</sup> A spectrum of free energies, ranging from favorable to unfavorable relative to that of bulk solvent, is stored in the H-bond network of solvating water as a function of local solute surface structure and composition (Figure 5), as follows:

- 1) **Strained solvation:** diffuse H-bond/enthalpically depleted, entropically neutral solvation at non-polar or single/weak H-bond donor- or acceptor-containing solute regions, corresponding to a wide range of local environment-dependent positive potential energies.
- 2) **Partially relaxed solvation:** bulk-like, entropically neutral solvation at moderately polar solute regions, corresponding to zero strain.
- 3) **Fully relaxed solvation:** H-bond/enthalpically enriched, entropically depleted solvation at multi-donor- and/or acceptor solute environments exhibiting a wide range of local environment-dependent negative/favorable potential energies.

The absolute free energy ( $G_j^{abs}$ ) of a given non-covalent sterically-accessible state  $j$  consists predominantly of the H-bond enriched + depleted solvation free energy of that state relative to that of bulk solvent (i.e.,  $G_j^{abs} \approx G_j^{bulk\ solvent} - (G_j^{enriched} + G_j^{depleted}) = \delta G_j$ ). States in which  $(G_j^{enriched} + G_j^{depleted}) \gg G_j^{bulk\ solvent}$  are unfavorable, and therefore poorly populated. The chemical structures of solutes (Figure 6, left) are effectively “imprinted” in their solvating water (Figure 6, right), loosely analogous to the manner in which heavenly bodies are imprinted in space-time, resulting in gravity. From this perspective, the water H-bond network equates to a field (henceforth referred to as the “solvation field”) that responds to the regional loss and gain of H-bonds across solute surfaces (noting that solvating water is capable of energetically resolving solute differences on the order of a single carbon atom)<sup>3</sup>.

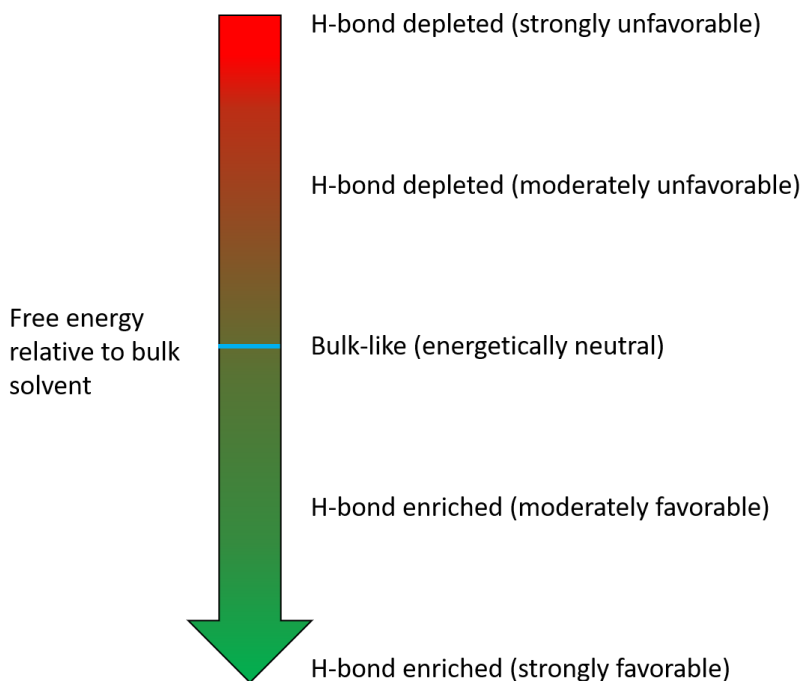


Figure 5. The free energy of bulk solvent is split into a spectrum of favorable and unfavorable contributions across the internal (buried) and external (exposed) surfaces of solutes relative to that

of bulk solvent, corresponding to H-bond depleted/strained (potential energy > bulk solvent), bulk-like/relaxed (energetically neutral), and fully relaxed H-bond enriched (potential energy < bulk solvent) water-water-solute configurations. Potential energy stored in H-bond depleted solvation is released via the replacement of such water by non-polar solute groups that themselves incur little desolvation cost during intra- or inter-solute rearrangements.

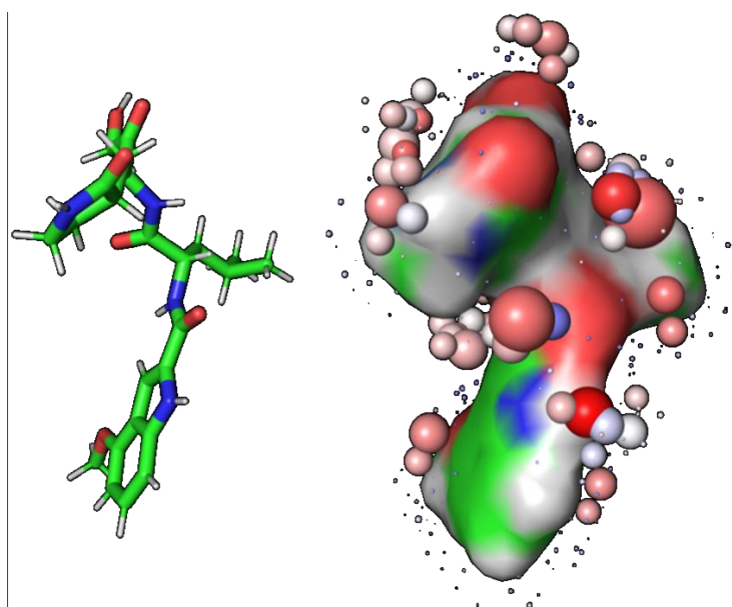


Figure 6. (Left) Non-covalent intra- and intermolecular energies are commonly interpreted on the basis of chemical structure and interatomic contacts. (Right) In reality, such energies are contributed largely by solute/state-specific solvation fields represented here by colored spheres (color-coding explained in Materials and methods) denoting H-bond depleted/low water occupancy (tiny spheres) and enriched/high water occupancy (large spheres) regions of the field. Solutes are driven toward their lowest free energy states corresponding to minimized H-bond depleted solvation and maximized H-bond enriched solvation content. Solubility and rearrangeability depend on a counter-balanced mixture of H-bond enriched and depleted solvation within a Goldilocks zone of total solvation free energy.

**Relaxation of the solvation field results from the replacement of interfacial H-bond depleted solvation by non-polar solute groups during intra- or inter-solute rearrangements.** All non-covalent rearrangements are powered predominantly by the drive of each solvating water to achieve H-bond free energies  $\leq$  that of bulk solvent (Figure 7). We proposed previously that “hydrophobicity” and “lipophilicity are misnomers for non-polar-non-polar interactions, which in practice, are less favorable than those between water and non-polar moieties.<sup>3</sup> Sterically accessible solute states residing significantly below the Goldilocks zone of H-bond enriched solvation are non-rearrangeable, whereas those significantly above that of H-bond depleted solvation form slowly or not at all. H-bond enriched and depleted solvation “flavors” are comingled (non-randomly) and dynamically counter-balanced, as follows:

- 1) Overall solubility (neglecting dissolution) is promoted by H-bond enriched solvation residing largely on concave regions of solute surfaces.
- 2) Intra-/intermolecular rearrangeability (i.e., sensitivity to perturbations). Populated states reside within a Goldilocks zone of H-bond enriched versus depleted solvation. Such states form the fastest (i.e., have the lowest barriers or appropriate H-bond replacement groups), while achieving sufficient kinetic stability to perform their functions (e.g., enzyme-substrate complexes are sufficiently stable to turn over prior to substrate dissociation).



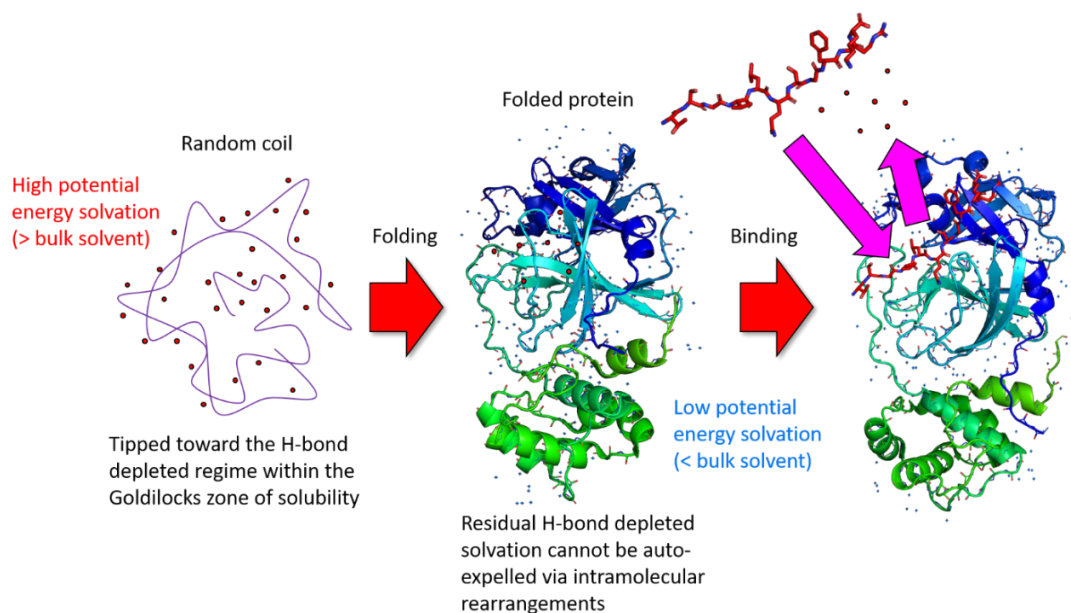


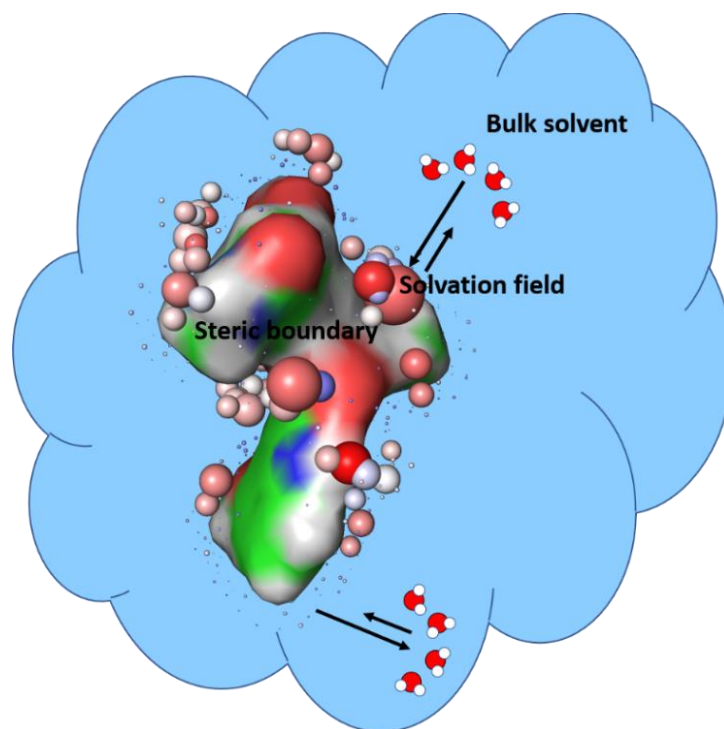
Figure 7. Cartoon depicting the auto-desolvation of high potential energy H-bond depleted solvation (denoted by red filled circles) (left) via folding of HMW and conformational rearrangements of LMW solutes (exemplified for the apo form of COVID M<sup>pro</sup>) (middle), and via non-covalent binding (right) (exemplified for the substrate-bound form of COVID M<sup>pro</sup>)<sup>4</sup>. Rearrangement-induced desolvation of H-bond enriched water (denoted by blue filled circles) typically incurs a loss of H-bond free energy, but is at best, a zero sum game of solute H-bond gains and water H-bond losses. Crystal waters correspond to bulk-like, H-bond enriched, or trapped solvation.

Continuous water exchanges between the solvation field and bulk solvent (Figure 8A), and possibly between H-bond depleted and enriched regions of the same field (Figure 8B), are driven by potential energy difference between these regions. Water transfer rates depend on the degree of enrichment or depletion of water-water H-bonds at each solute surface position, as follows:

- 1) Fully exposed water distributed around the external surfaces of LMW and HMW solutes, which is readily exchangeable with bulk solvent. Water occupancy is bulk-like at iso-bulk solvent H-bond positions (3.5 moderately strong H-bonds per water), high at H-bond enriched/energetically relaxed positions (3.5-4 strong H-bonds per water) corresponding to fast entry/slow exit, and low at H-bond depleted/energetically strained positions (< 3.5 weak H-bond per water) corresponding to slow entry/fast exit.
- 2) Partially exposed water occupying buried surface-connected channels that form transiently in HMW solutes, which exchanges slowly with bulk solvent due to impeded diffusion within the channels. Such channels appear in time-averaged molecular dynamics trajectories and some experimentally determined structures.
- 3) Water trapped within transient highly strained surface-disconnected buried cavities (“bubbles”) within HMW solutes, which is non-exchangeable with bulk solvent. Such volumes likewise appear in time-averaged MD trajectories and some experimentally determined structures.

We previously reported a qualitative first principles computational method called WATMD for predicting the distribution of H-bond depleted, enriched, and bulk-like solvation across internal and external solute surfaces relative to that of bulk solvent based on all-atom, explicit water MD simulations referred to in this work as “solvation dynamics” (SD) simulations to emphasize the focus on water exchanges versus molecular state transitions.<sup>3,4</sup> We used this approach to deduce the inhibition and functional state transitions of COVID M<sup>PRO</sup> based on a set of crystal structures in which the putative functional states were captured.<sup>4</sup>

A



B

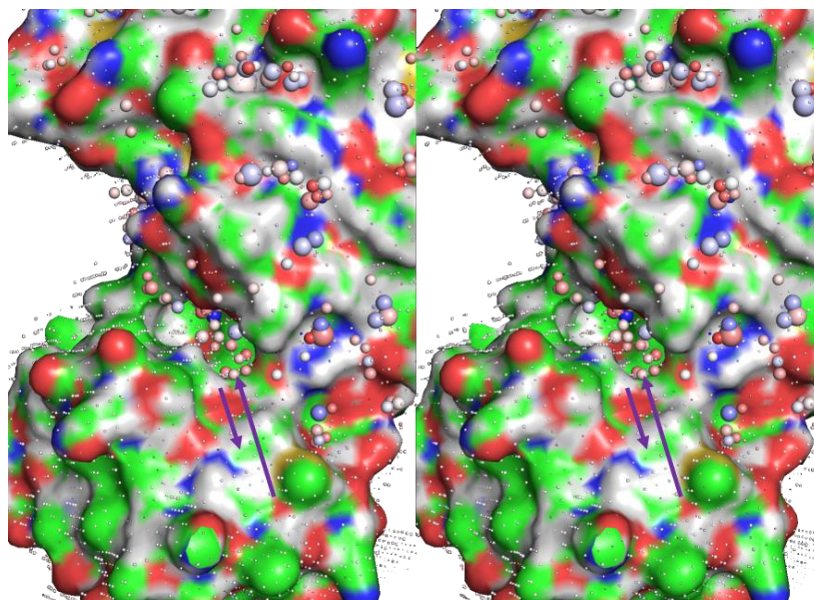


Figure 8. Cartoons showing the hypothetical exchanges between bulk solvent and H-bond enriched/high water occupancy (denoted by large spheres) and depleted/low water occupancy

(denoted by small spheres) regions of the solvation field for LMW and HMW solutes (color-coded by atom type). The rates of water exchange to/from solvation and bulk solvent are governed by free energy barriers that are proportional to the loss of interfacial H-bond free energy relative to that of bulk solvent. (A) The solvation field of the COVID M<sup>pro</sup> inhibitor PF-00835231 overlaid on a volume of bulk solvent (blue cloud). Fast water transfer from bulk solvent/slow transfer to bulk solvent occurs at H-bond enriched positions of the solvation field, whereas slow water transfer from bulk solvent/fast transfer to bulk solvent occurs at H-bond depleted positions. (B) Fast water transfer laterally across the surfaces of HMW solutes (exemplified here by COVID M<sup>pro</sup>)<sup>4</sup> may likewise occur between H-bond depleted and H-bond enriched positions (stereo view of the M<sup>pro</sup> surface color-coded by atom type).

### **From equilibrium $\Delta G$ in vitro to non-equilibrium $\Delta G^\ddagger$ in living cells**

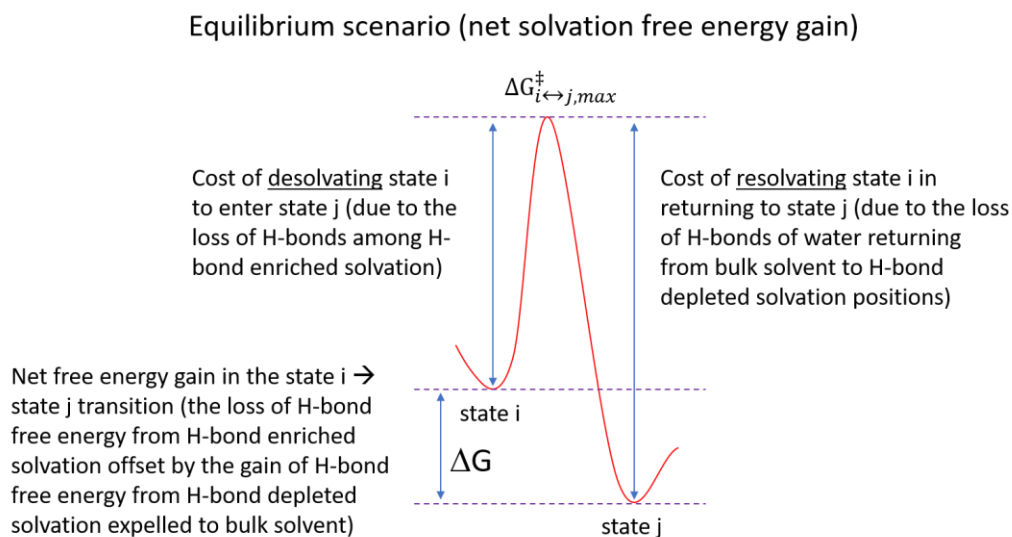
Under equilibrium conditions, the sterically-accessible configurations of HMW and LMW solutes are distributed in a stationary fashion in accordance with  $\Delta G = \Delta G_{i \rightarrow j}^\ddagger - \Delta G_{j \rightarrow i}^\ddagger$  (Figure 9A). However, under non-equilibrium conditions in cells (the basis of signal-carrying molecular and ionic fluxes), the configurations are distributed transiently in proportion to the separate  $\Delta G_{i \rightarrow j}^\ddagger$  and  $\Delta G_{j \rightarrow i}^\ddagger$  entry/exit barriers to/from each state (i.e., the intrinsic rates) (Figure 9B). The most populated states of a given species consist of those that build the fastest and/or decay the slowest relative to the rates of buildup and decay of the species itself (i.e., the extrinsic rates).<sup>1</sup> We proposed more than a decade ago that:<sup>23</sup>

- 1)  $\Delta G_{i \rightarrow j}^\ddagger$  equates principally to the desolvation cost of H-bond enriched solvation, which decreases as the H-bonds of such water are replaced increasingly optimally by intra-solute H-bonds during the  $i \rightarrow j$  state transition (Figure 9B). We define  $\Delta G_{i \rightarrow j, \max}^\ddagger$  as the

maximum cost of expelling H-bond enriched solvation in the complete absence of intra- or inter-solute water H-bond replacements, and  $\Delta G_{j \rightarrow i, max}^\ddagger$  as the maximum cost of re-solvating H-bond depleted positions during the  $i \rightarrow j$  state transition. An additional cost is incurred during the  $i \rightarrow j$  state transition when H-bond depleted solvation is generated de novo in state j due to the addition of solvent exposed non-polar groups and/or the loss of solvent exposed polar groups.

- 2)  $\Delta G_{j \rightarrow i}^\ddagger$  equates principally to the resolution cost of H-bond depleted solvation positions from which water was expelled during entry to state j. We define  $\Delta G_{j \rightarrow i, max}^\ddagger$  as the maximum possible cost of re-solvating H-bond depleted solvation positions if all such positions were expelled during the  $i \rightarrow j$  state transition. An additional cost is incurred in the  $j \rightarrow i$  state transition when H-bond enriched solvation is generated de novo in state j.

A



B

Non-equilibrium scenario (costs of entry and exit to/from each sterically accessible state)

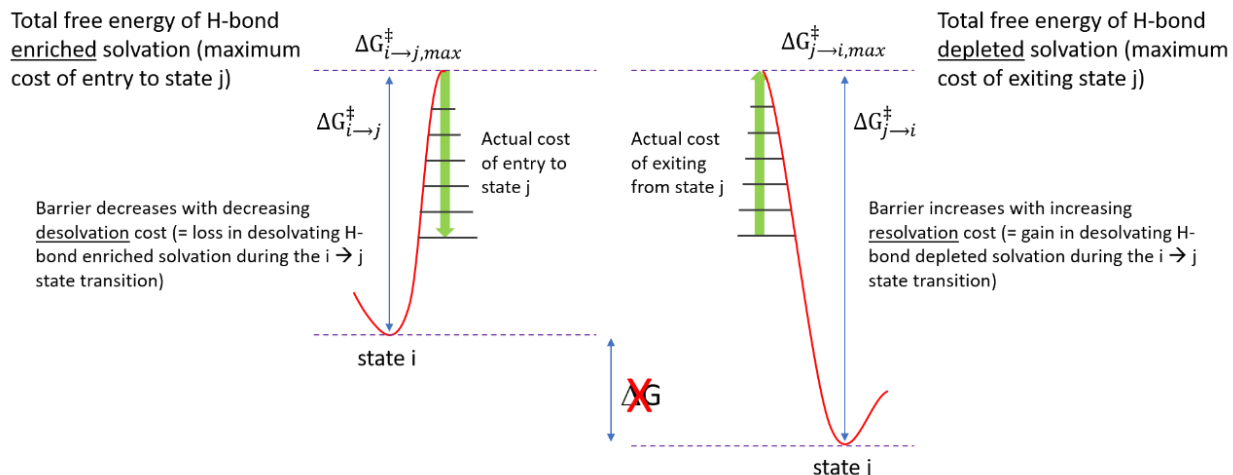


Figure 9. The sterically-accessible intra- and intermolecular structural states of a given solute equilibrate between states  $i$  and  $j$  via the expulsion of H-bond enriched and depleted solvation to bulk solvent upon dissolution of the solute at  $t = 0$ . (A) The equilibrium states at  $t = \infty$  are distributed according to the net free energy of favorable H-bond enriched plus unfavorable depleted solvation in each state ( $\Delta G_{i \leftrightarrow j}$ ). (B) Prior to equilibration, the states are distributed solely based on their rates of entry and exit (proportional to  $\Delta G_{i \rightarrow j}^{\ddagger}$  and  $\Delta G_{j \rightarrow i}^{\ddagger}$ , respectively), which in turn, are contributed principally by the cost of expelling H-bond enriched and depleted solvation to bulk solvent. As such, the cost of desolvating H-bond enriched solvation during the  $i \rightarrow j$  state transition equates to the cost of accessing H-bond depleted solvation, which together, convey specificity (where the H-bond replacements are tuned for the cognate partners) and kinetic stability (where the rate of decay of state  $j$  is proportional to the cost of re-solvating H-bond depleted solvation positions that were desolvated during the  $i \rightarrow j$  transition).

The energetically accessible states of a given species exhibit low  $\Delta G_{i \rightarrow j, max}^\ddagger$  and/or high  $\Delta G_{j \rightarrow i, max}^\ddagger$ .  $\Delta G_{i \rightarrow j}^\ddagger$  decreases as the H-bonds of H-bond enriched solvation are increasingly replaced by iso-energetic intra- or inter-solute H-bonds (the stringency of which grows with increasing magnitude of  $\Delta G_{i \rightarrow j, max}^\ddagger$ ). Net free energy gains due to such replacements are unlikely due to:

- 1) Entropy-enthalpy compensation, equating to the entropic cost accompanying large gains in H-bond enthalpy.
- 2) The typically lower free energy of optimal water-solute H-bonds that promotes H-bond enriched solvation compared with intra- and -inter-solute H-bonds (noting that increasing propensity of solute-solute H-bonds is mirrored by increasing propensity of water-solute H-bonds, resulting in a zero sum game at best).

Intramolecular transitions between states i and j (with the rate constant denoted as  $k_{in}$ ) are guided by the cost of desolvating H-bond enriched solvation in cis, as follows:

- 1)  **$k_{in} \approx 0$  ( $s^{-1}$ ) (collisions)**: high desolvation cost of H-bond enriched solvation within the rearrangement interface during entry to state j in the absence of intra-solute replacements for H-bonds lost during expulsion of H-bond enriched water to bulk solvent. The solvation fields across the interface effectively bounce apart.
- 2) **Maximal  $k_{in}$  ( $s^{-1}$ )**: low cost desolvation of H-bond enriched solvation within the rearrangement interface during entry to state j in the presence of intra-solute replacements. The solvation fields across the interface collapse due to the expulsion of H-bond enriched solvating water to bulk solvent, thereby allowing access to H-bond depleted solvation positions by non-polar/weakly polar solute groups.

- 3)  **$0 < k_{in} < \text{maximal}$** : a range of desolvation costs of H-bond enriched solvation within the rearrangement interface during entry to state j associated with partial replacement of the H-bonds that were lost during water transfer to bulk solvent.

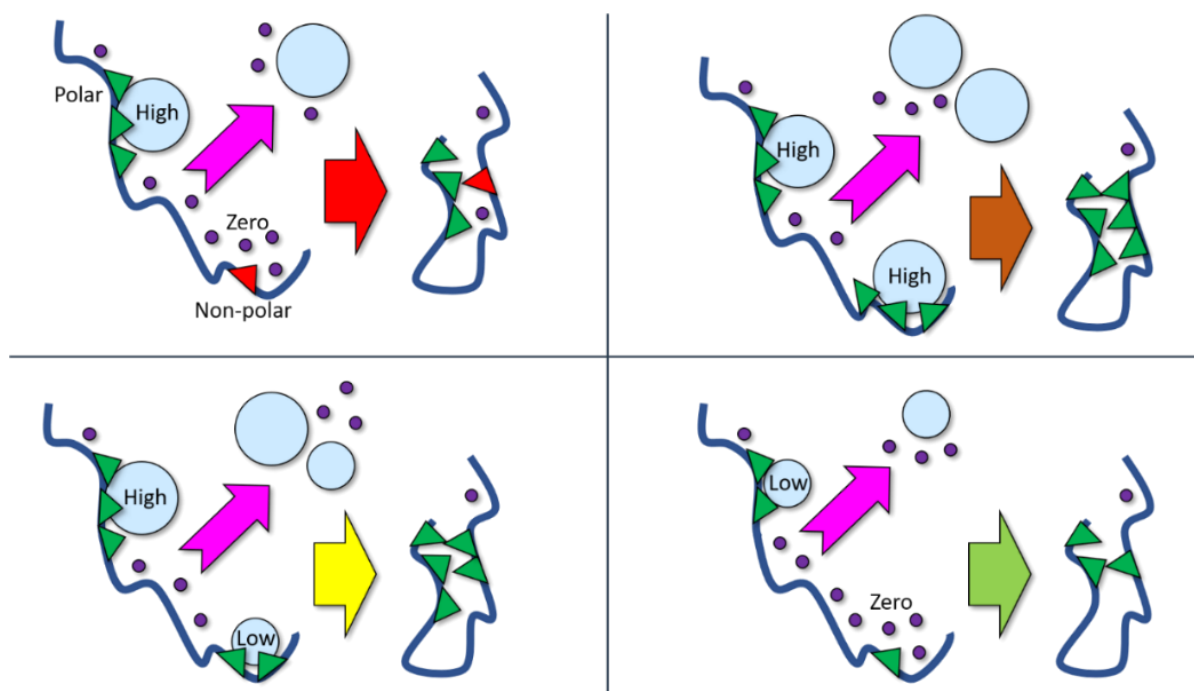
Intramolecular states are kinetically stabilized (with the rate constant denoted as  $k_{out}$ ) by H-bond depleted solvation, as follows (Figure 10A):

- 1) **Maximal  $k_{out}$  ( $s^{-1}$ )**: corresponding to no resolution cost at H-bond depleted positions during exit from state j due to the lack of such positions or non-polar/weakly polar solute groups that are capable of replacing them at a low desolvation cost.
- 2)  **$0 < k_{out} < 1$  ( $s^{-1}$ )**: corresponding to low-high resolution costs at H-bond depleted positions during exit from state j.

Entry to state j is slowed by the de novo generation of H-bond depleted solvation in state j in excess of that in state i (which destabilizes state j), whereas exit from state j is slowed by the de novo generation of H-bond enriched solvation in state j in excess of that in state i (which stabilizes state j). In the absence of de novo H-bond enriched solvation, the lifetime of state j is determined principally by the resolution cost of returning to state i or another state. Intermolecular state transitions are powered in a similar fashion as their intramolecular counterparts, except that desolvation of the interface occurs in trans (Figure 10B), and  $k_{in}$  and  $k_{out}$  are denoted as  $k_{on}$  and  $k_{off}$  (or  $k_1$  and  $k_{-1}$  in the case of enzymes), respectively. Intra- and intermolecular state transitions frequently occur in combination, reflecting the merger of the solvation fields of the binding partners (which are not simply the sum of the original fields of the partners due to non-linearity of the merging process).



A



B

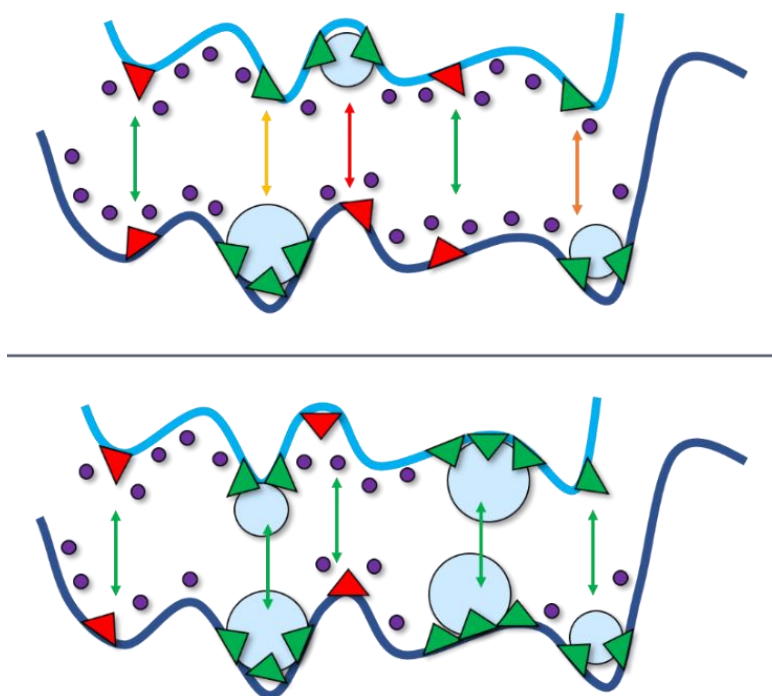


Figure 10. (A) The distribution of sterically accessible intramolecular states under both static equilibrium and dynamic non-equilibrium conditions is governed by the interfacial desolvation cost (enthalpic cost) and resolvation cost (enthalpic cost + entropic cost) incurred during entry to and exit from each state, respectively (noting that desolvation is entropically favorable and resolvation is entropically unfavorable, where entry and exit equate to the gain and loss of interfacial interatomic contacts, respectively). The rate of entry is proportional to the cost of desolvating H-bond enriched positions in cis, where slow entry equates to repulsion between unreplaced H-bonds of the interfacial solvating water. The rate of exit is proportional to the cost of resolving H-bond depleted positions within the dissociating interface. We showed previously that little or no resolvation free energy cost corresponds to state lifetimes on the order of  $\mu\text{s}$ <sup>27</sup>. **H-bond enriched solvation thus serves as a “gatekeeper” for accessing H-bond depleted solvation**, relegating intermediate states to ultrashort lifetimes. The spectrum of desolvation costs is governed case-by-case by: 1) the magnitude of the excess H-bond free energy of H-bond enriched solvation relative to that of bulk solvent (denoted as  $\Delta G^{enriched}$ ); and 2) the magnitude of the excess H-bond free energy of bulk solvent relative to that of H-bond depleted solvation (denoted as  $\Delta G^{depleted}$ ). Four hypothetical scenarios are exemplified for a single pair-wise intramolecular desolvation reaction. Top left: a high  $\Delta G^{enriched}$  solvating water (light blue filled circle) at a polar solute position (green triangle) and two high  $\Delta G^{depleted}$  solvating water molecules (purple filled circles) at non-polar solute positions (red triangles). Top right: two high  $\Delta G^{enriched}$  solvating waters. Bottom left: high and low  $\Delta G^{enriched}$  solvating waters. Bottom right: a high  $\Delta G^{enriched}$  solvating water and a high  $\Delta G^{depleted}$  solvating water located at a weakly polar solute position. The positions of H-bond enriched solvation on polypeptide chains evolved to guide the folding process, whereas depleted positions evolved to stabilize folding-relevant sub-states

along the folding pathway. (B) Same as A, except for the intermolecular case. Top: non-cognate partners collide, rather than associate, due to the high desolvation cost of H-bond enriched solvation within the binding interface. Bottom: cognate partners associate due to optimal solvation field complementarity, corresponding to lower desolvation cost of H-bond enriched solvation, and higher resolution cost at H-bond depleted solute positions.

The most fundamental structure-free energy relationships therefore consist of structure- $\Delta G^{depleted}$  and structure- $\Delta G^{enriched}$ , rather than structure- $\Delta G$  relationships (which may be derived from the solvation free energy contributions under equilibrium conditions). The general mechanisms by which solvation free energy is transduced into the conformational distributions of HMW and LMW species, protein folding, and binding between HMW-LMW (including drug-target) and HMW-HMW species under non-equilibrium conditions are further explored and exemplified in the remainder of this work.

## **Solvation-powered protein folding, structure, and function**

Proteins fold and undergo specific intra- and intermolecular state transitions throughout their lifetimes. Non-linearity of multi-protein fluxes necessitates consideration of protein structure and function in a holistic fashion. However, for the sake of simplicity, we consider protein folding in isolation, as a non-linear stepwise desolvation-resolution-powered process (Figure 11A). Folding begins near the upper boundary of the Goldilocks zone of H-bond depleted solvation free energy in the unfolded/random coil state (in which the protein is fully soluble), and proceeds toward the lower boundary throughout the folding process via the stepwise auto-expulsion of H-bond depleted solvation (nature's form of energy minimization). Under non-equilibrium conditions, folding follows a strictly kinetic/dynamic (versus thermodynamic) pathway, in which the most populated states: 1) form the fastest; and 2) persist the longest. The timescale of protein synthesis/folding (or

activation in the case of fast cellular responses), rearrangement (including binding), and deactivation or degradation is necessarily  $\leq$  that of the corresponding function. The protein folding rate is governed by:

- 1) The rate of entry into each sub-state along the folding pathway, where the fastest rates (corresponding to the lowest cost of desolvating H-bond enriched solvation via optimal intra-protein replacement of the H-bonds of this water) reside along the evolutionarily-determined folding pathway.
- 2) The rate of exit from each substate along the folding pathway. Rate-limiting steps during folding are circumvented by short-lived sub-states. Exit rates slow as the random coil state condenses to secondary structure elements (SSEs) and tertiary structure powered by the cumulative replacement of H-bond depleted solvation by non-polar or weakly polar groups. The resolution cost of total or partial unfolding or exit to a different folded substate determines the lifetimes of the folded states, respectively.

Folding is impeded by:

- 1) Excessively favorable solvation free energy in the unfolded state due to high polar amino acid content, which instead promotes extended, maximally solvent-exposed structural states.
- 2) Excessively unfavorable solvation in the unfolded state due to highly non-polar amino acid content, which promotes poor solubility, and in extreme cases aggregation/precipitation.
- 3) The absence of SSE formation in disordered regions (including loops and the N- and C-termini), in which auto-desolvation of H-bond depleted solvation is hampered due to:

- a) Sterically constrained intra-protein H-bonding.
- b) High desolvation cost of H-bond enriched solvation.

Disorder manifests as the protein's attempt to alternately expel H-bond depleted solvation in multiple regions (commonly referred to as “conformational frustration”).

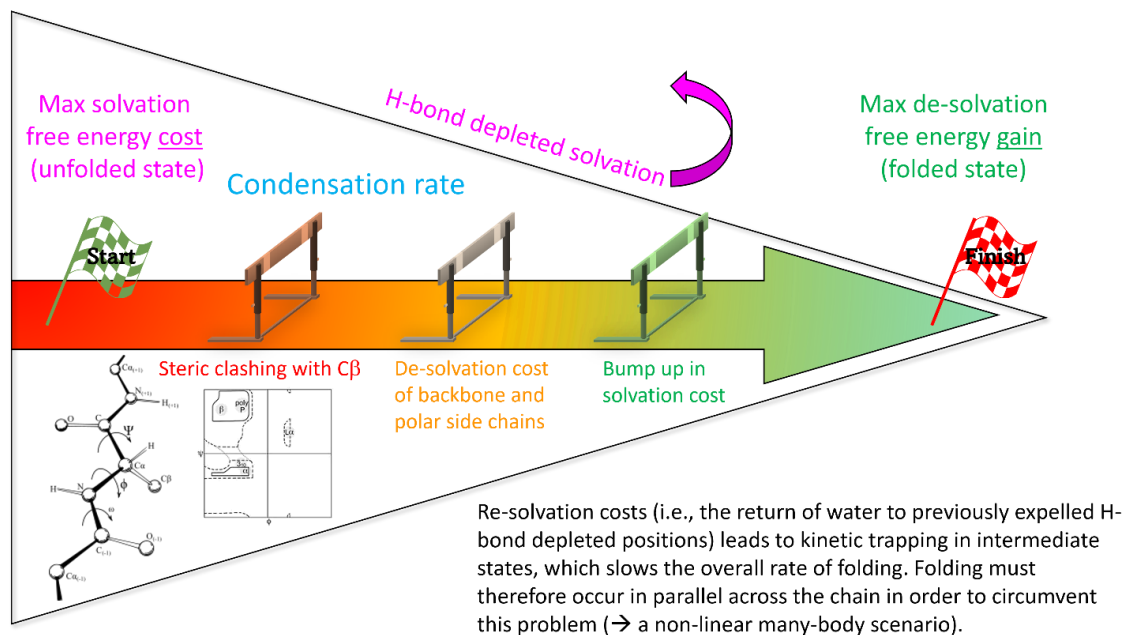
Rearrangeability of the folded state depends on a Goldilocks zone of counter-balanced residual H-bond depleted and enriched solvation, as follows:

- 1) Residual H-bond depleted solvation powers intra- and intermolecular rearrangements.
- 2) Residual H-bond enriched solvation stabilizes the folded state and counter-balances the destabilizing effect of H-bond depleted solvation, thereby limiting state transitions to those in which the H-bonds of expelled H-bond enriched solvation are replaced optimally by intra- or intermolecular H-bonds.

Intra-protein H-bonds present in experimentally observed protein structures represent the final cumulative replacements of H-bond enriched solvation desolvated throughout the folding process. **As such, intra-protein H-bonds correspond to positions at which H-bond enriched solvation “gatekeepers” controlling co-localization of the participating donors/acceptors have been optimally replaced** (Figure 11B), but do not contribute directly to the stability or persistence of the folded structure. It is further apparent that each intramolecular interface from which H-bond depleted solvation was expelled during the folding process decays as the interface resolvates (the lifetime of which depends on local resolvation costs), and reforms at a rate governed by the local desolvation costs. This putatively coincides with the transient formation of buried solvation (apparent in many of our time-averaged SD trajectories) that is capable of storing and releasing

potential energy (as we demonstrated previously for COVID M<sup>pro</sup>,<sup>4</sup> and as we demonstrate for cereblon below).

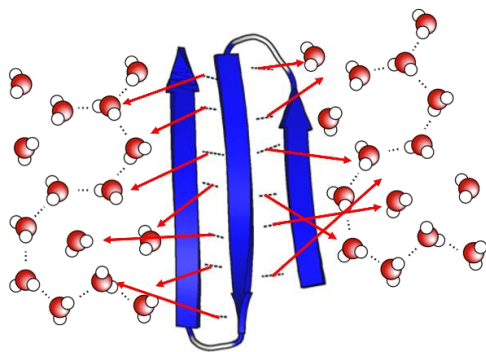
A



B

The rate of SSE formation (and chain condensation) depends on the de-solvation cost of H-bond enriched solvation at backbone NH and C=O and polar side chains. Loops may correspond to regions of higher de-solvation cost (or lower solvation cost).

The rate of  $\beta$ -sheet formation depends on the cost of expelling H-bond enriched solvation from the inter-strand interface.



The rate of helix formation depends on the cost of expelling H-bond enriched solvation from the inter-turn interface.

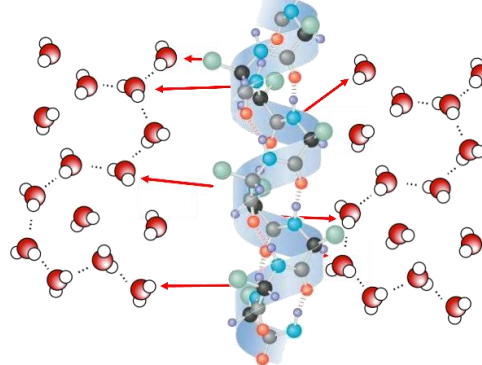


Figure 11. (A) Protein folding is driven by the step-wise desolvation of H-bond depleted solvation, from the maximal H-bond depleted solvation free energy in the unfolded state (i.e., release of the positive potential energy stored in such solvation) to the minimum in the folded state (noting that rearrangeability of the folded state depends on residual H-bond depleted solvation in the folded state within a Goldilocks zone of global stability). The folding pathway is guided by H-bond enriched solvation, in which the desolvation entry cost is lowest at each correct sub-state of the pathway compared with all other sterically accessible substates. (B) The fastest folding trajectories (necessarily corresponding to functionally folded proteins) consist of those in which H-bond enriched solvation is expelled by intra-protein, or in some cases intermolecular, H-bond replacements.

Sterically allowed  $\phi$ - $\psi$  backbone torsion angles, which are constrained largely by their common  $C_{\beta}$ -backbone interactions, are similar for all isolated amino acids except Gly and Pro. Preferences for particular  $\phi$ - $\psi$  angles in polypeptide chains are therefore determined largely by side chain-side chain and side chain-backbone interactions between flanking and remote residues, as follows:

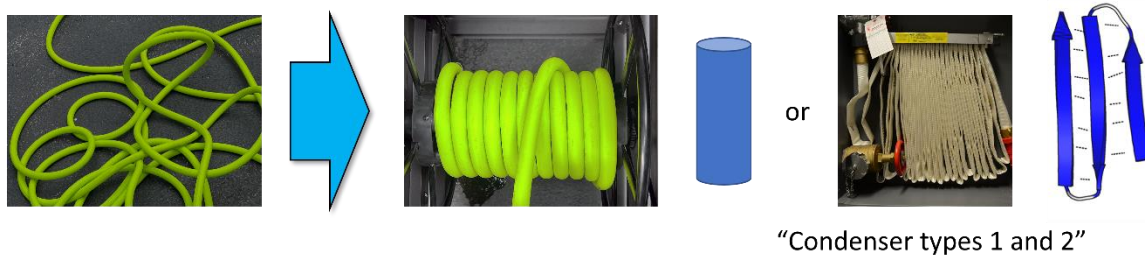
- 1) Side chain-specific and backbone solvation fields in the unfolded state.
- 2) SSE-specific solvation fields.
- 3) Side chain-side chain and side chain-backbone packing (i.e., the extent of packing-mediated backbone and side chain desolvation).
- 4) Side chain-SSE, backbone-SSE, and SSE-SSE packing (i.e., the extent of packing-mediated desolvation of secondary structure elements).

SSE formation is the means by which local polypeptide regions are condensed into compact substructures, analogous to the winding of a garden hose around a reel (helices) or the folding of

a firehose (parallel and anti-parallel  $\beta$ -sheets) (Figure 12A).  $\alpha$ -helices and  $\beta$ -sheets undergo further packing to form helical bundles and  $\alpha/\beta$  structures (Figure 12B).

A

The chain is condensed by “winding” into helices and anti-parallel  $\beta$ -sheets, which depends on intra-chain H-bond formation (SSEs are the primary means of replacing backbone H-bond enriched solvation, which governs the rate of intra-chain H-bond formation).



SSE formation as a function of chain position (corresponding to the lowest de-solvation free energy cost) evolved to maximize expulsion of H-bond depleted solvation via intra- and inter-SSE packing.

B

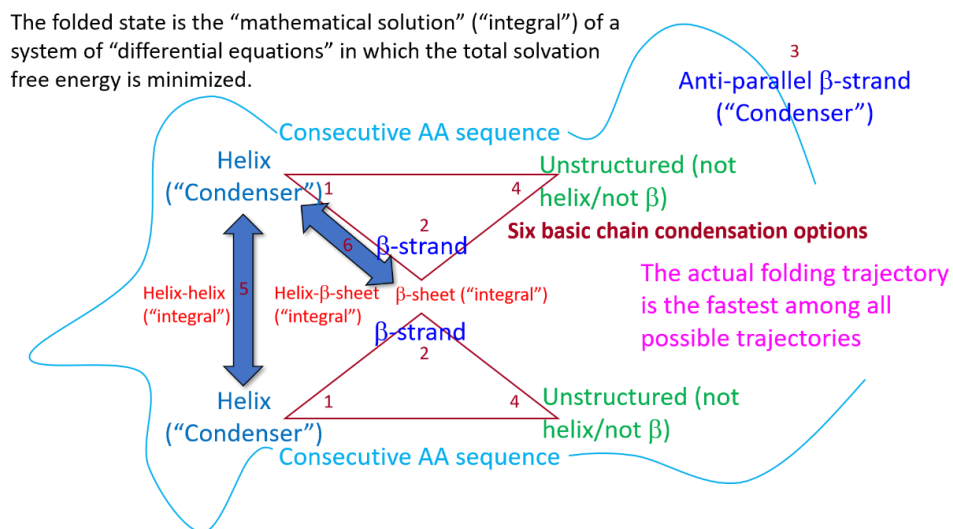




Figure 12. (A) Polypeptide chains condense into compact substructures via helix and  $\beta$ -sheet formation. (B) The chain further condenses via packing between SSEs.

Each folding step depends on the mutual relationship between the transient protein geometry and corresponding solvation field and many-body interactions between the remote quasi-independent parts of the chain (e.g.,  $\beta$ -sheet,  $\alpha/\beta$ , and helical bundle formation), resulting in non-linearity of the folding process. Whereas isolated SSEs can be loosely predicted based on sequence-structure relationship models, first principles prediction of the SSEs within tertiary structures is hampered by the non-linearity of their formation (Figure 13) (noting that data models like AlphaFold<sup>28</sup> are useful for predicting folded structures rather than folding pathways per se).

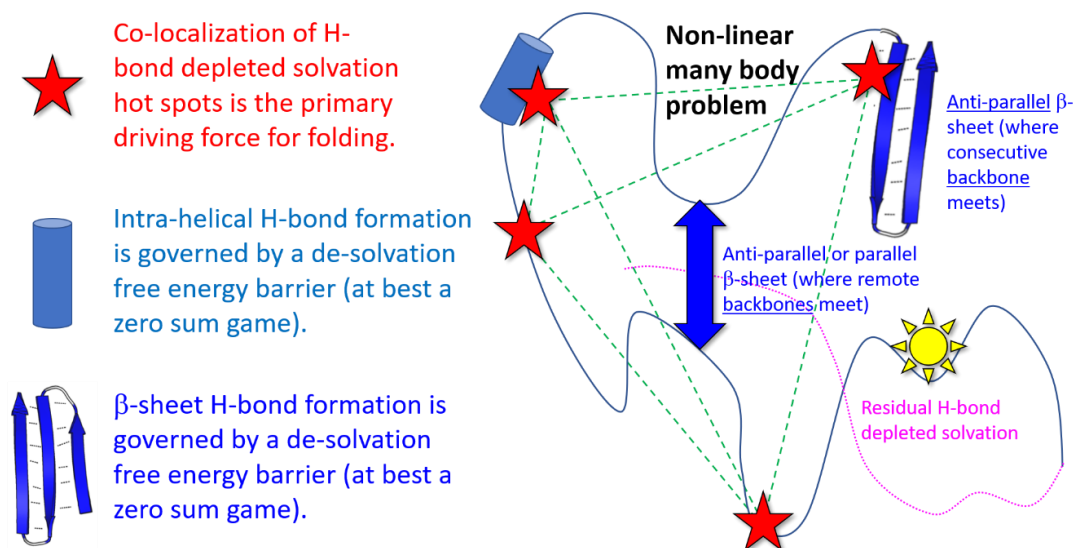


Figure 13. Each substructural element of a protein undergoing folding depends on all other elements, and vice versa, consistent with NLD behavior (i.e., the folded state of a protein is  $>$  the  $\Sigma$ (states of the constituent residues)).

Protein folding and the state transitions of folded proteins are both powered and facilitated by solvation, as follows:

- 1) Protein fluxes operate in a non-equilibrium source (synthesis/translocation)-sink (degradation translocation) fashion, in which the distribution of populated states is tipped transiently to one or more specific states in response to perturbations/signals.
- 2) States build and decay on the basis of local packing and unpacking of SSEs and/or unstructured regions (e.g., surface loops or buried regions) according to their desolvation and resolvation costs, respectively.
- 3) The external surfaces of proteins are solvated, as are rearrangeable buried SSEs and non-structured elements (NSEs), the latter in the form of transient internal water-filled cavities/pockets/channels (which we refer to as “internal solvation”) (Figure 14).
- 4) Internal solvation may be trapped (i.e., non-exchangeable with bulk solvent or other buried cavities), impeded (i.e., exchange more slowly with bulk solvent than external solvation), or freely exchangeable (i.e., exchange at a similar rate with bulk solvent). Internal solvation exists within a Goldilocks zone of enthalpic and entropic stability (i.e., overtly unstable and stable cavities/pockets/channels do not form in the first place, and not rearrange, respectively). The upper boundary of the Goldilocks zone consists of fully trapped water that is partially stabilized by water-water and water-protein H-bonds (at an entropic cost). The lower boundary consists of water that exchanges freely with bulk solvent via surface openings.

- 5) Internal NSEs rearrange transiently into water-filled cavities/pockets/channels, resulting in trapping or expulsion of the resident solvation to bulk solvent. The rearranged structural elements are necessarily resolvated by newly formed cavities/channels. **As such, buried water-filled cavities/pockets/channels are signatures of rearrangeable structural elements.**
- 6) Protein states build and decay at rates governed by the local interfacial resolution and desolvation costs of their rearrangeable internal structural elements (where the rate of buildup exceeds the rate of decay and vice versa, respectively).
- 7) Desolvation of deep concave pockets via ligand association may depend on buried water channels that connect the binding interface to bulk solvent, or between the ligand and binding site lumen when the ligand volume is sufficiently less than the pocket volume.

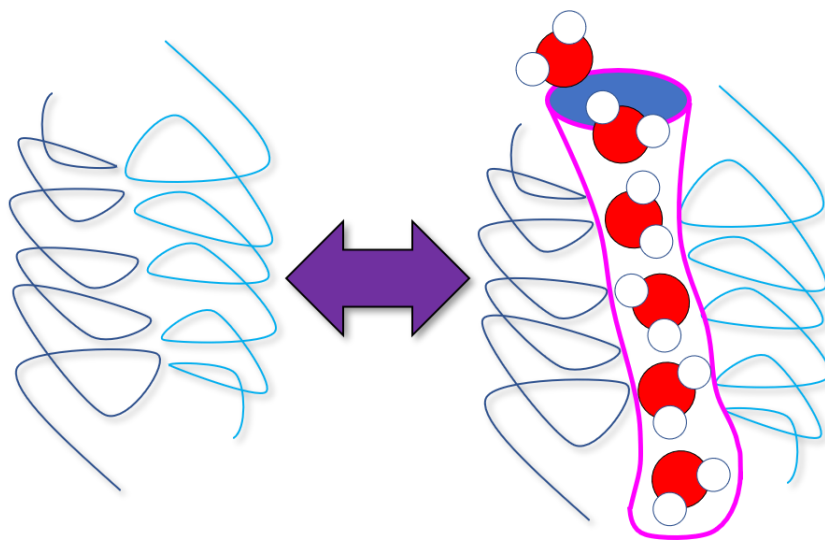


Figure 14. Intra-protein rearrangements are powered by the release of high energy solvation comprised of trapped/impeded water occupying transiently formed buried channels and cavities that we refer to as “internal solvation”.<sup>4</sup>

## **The solvation field is the target of many post-translational modifications**

The solvation fields of many solutes are modulated by post-translational modifications (e.g., phosphorylation/dephosphorylation, lysine acetylation/deacetylation, glycosylation), in which the solvation free energy and desolvation costs of the targeted solute groups are increased or decreased to varying degrees. Electrostatic interactions between charged groups that are separated spatially by  $\geq$  the Debye length are heavily screened by solvating water, the reason that co-localized positively charged side chains are commonly observed on solvent exposed protein surfaces. Rather, charge-charge (as well as charge-neutral) interactions are governed principally by the additional cost of desolvating the stronger water-solute H-bonds arising in such cases, compared with that of neutral groups. Extended conformations of targeted groups and the local substructures to which they are attached result from increased H-bond propensity and solvent exposure (affected substructures are effectively dragged into bulk solvent). Furthermore, the overall molecular density/radius of gyration is putatively governed by counter-balancing between the drive toward the maximal desolvation cost stabilizing the folded state (an inwardly-directed force) and maximal solvation of charged and polar surface groups (an outwardly-directed force). As such, the energy balance to/from folded versus extended solvent exposed conformations of the affected group(s) (e.g., the activation loop of protein kinases in the unphosphorylated and phosphorylated states) is tipped toward one state or the other via reversible post-translational modifications, contrary to the widely assumed electrostatic switching mechanism.

## **Overview of E3 ligase-mediated protein degradation**

Protein degradation serves as the ultimate sink for protein fluxes (the “terminal Yang”), many of which are degraded via the proteasomal pathway. Proteasome-mediated protein degradation in humans is initiated by poly-ubiquitylation of targeted proteins by a broad family of E3 ligases,

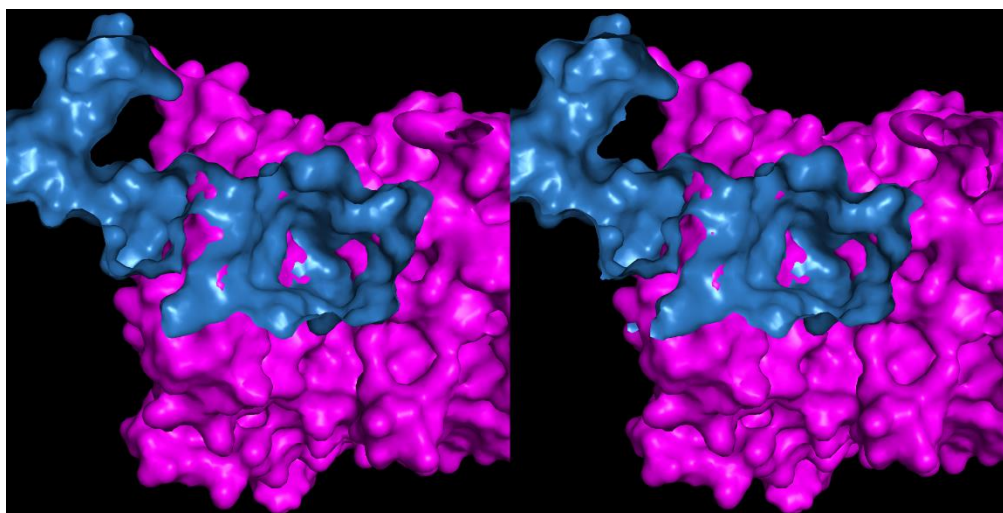
including cullin RING ubiquitin complexes assembled from the family members of cullin (Cul), RING-Box (RBX), adapter (e.g., DDB), and substrate recruitment (e.g., DCAF) subunits.<sup>29</sup> Cereblon, which consists of separate N- and C-terminal (CULT) domains, serves as the DCAF subunit of the CUL4A-DDB1-RBX1-CRBN E3 ligase.<sup>29</sup> Targeted substrates of this ligase bind exclusively to binary complexes comprised of the CULT domain of cereblon and LMW compounds (commonly referred to as “degraders”). The function(s), native degrader(s) and substrate(s) of the CUL4A-DDB1-RBX1/CRBN E3 ligase are poorly understood. Glutamine synthase has been proposed as a native substrate, and uridine as a native degrader (suggesting a possible role of this ligase in repairing DNA containing mis-incorporated uracil bases).<sup>30–32</sup> Cereblon is also the target of drugs (e.g., thalidomide and analogs thereof) that recruit specific non-native (“neo”) substrates (e.g., CK1 $\alpha$  and BRD4) that are poly-ubiquitylated by the CUL4A-DDB1-RBX1-CRBN E3 ligase and subsequently degraded (commonly referred to as “targeted protein degradation” (TPD)). Neo substrates are recruited by two classes of LMW compounds referred to in this work as Type I degraders (widely known as “glues” or IMiDs) and Type II degraders (widely known as “bifunctionals” containing cereblon- and substrate-binding ends). Type I degraders bind in an order-dependent fashion, first to cereblon, followed by substrate binding to the binary cereblon-degrader complex, whereas Type II degraders form binary complexes with cereblon and their substrate targets independently, such that only the fraction of degrader bound to both proteins simultaneously (one per end) is catalytically viable and the remaining fraction is dead. Neo substrates can be classified, as:

- 1) Those mimicking the native substrate-cereblon footprint (Figures 15A and B), which includes the degrader pocket of cereblon (denoted in this work as pocket L) and a  $\beta$ -turn on the target protein (recruited by Type I degraders).

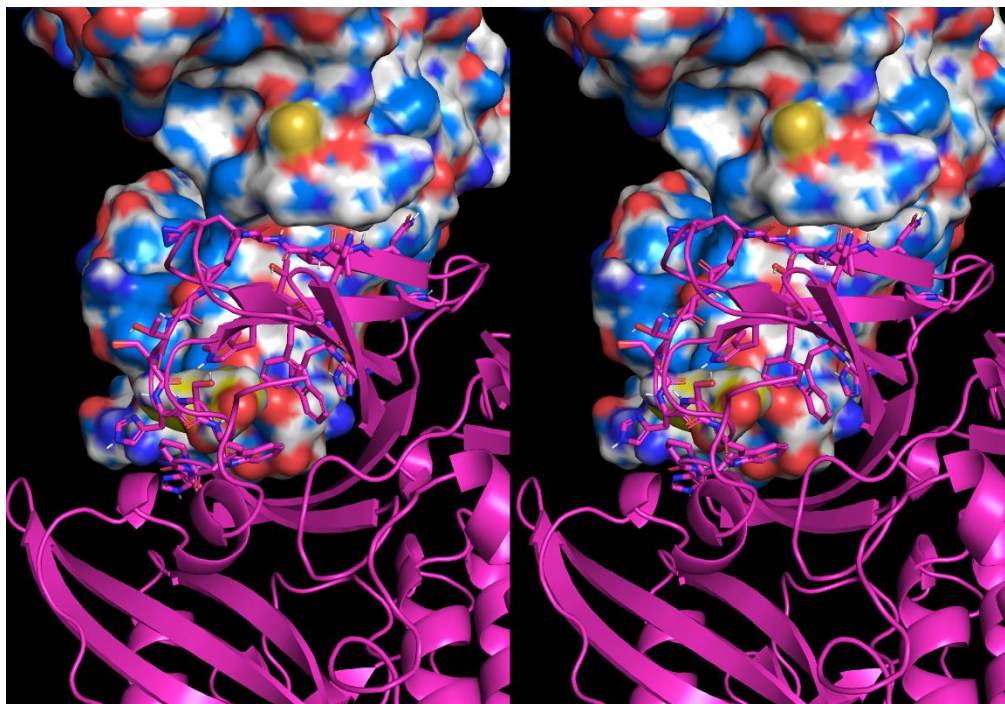
- 2) Those exhibiting smaller footprints that bypass the degrader pocket (Figures 15C and D) (recruited by Type II degraders). We hypothesize that the  $k_{-1}$  (the dissociation rate constant) of neo substrates is slowed by Type II degraders, which otherwise lack sufficient H-bond depleted and/or de novo H-bond enriched solvation residing on the substrate binding surface.

TPD is a proven clinical strategy, and the structural features governing recruitment are known for several target proteins.<sup>33–38</sup> However, the mechanistic, kinetic, and dynamic aspects of recruitment and degradation in the native cellular context addressed in this work are poorly understood, and typically treated using equilibrium models<sup>39</sup> that inconsistent with time-dependent substrate degradation under native conditions.

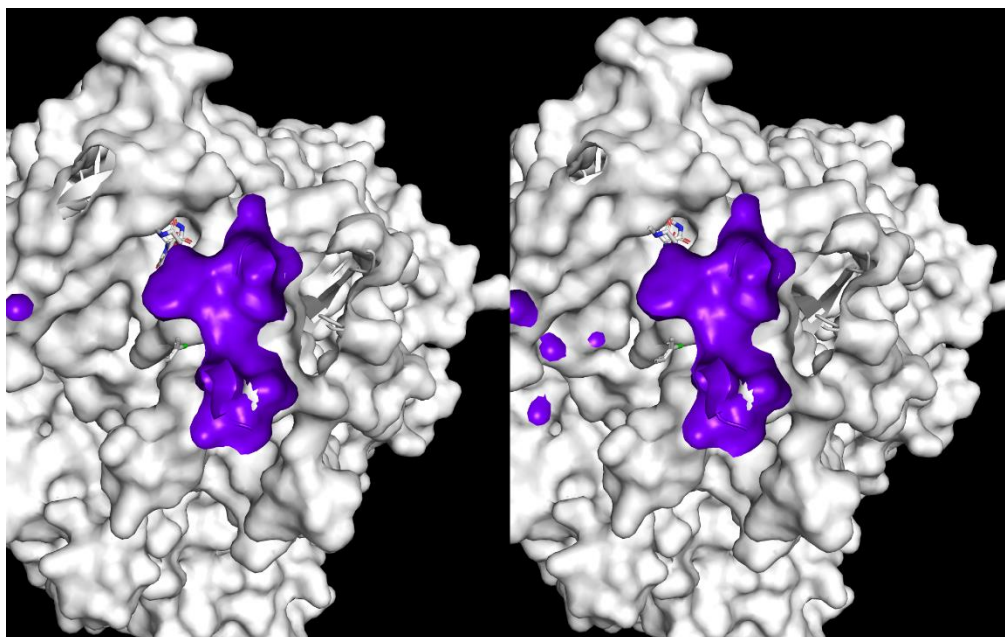
A



B



C



D

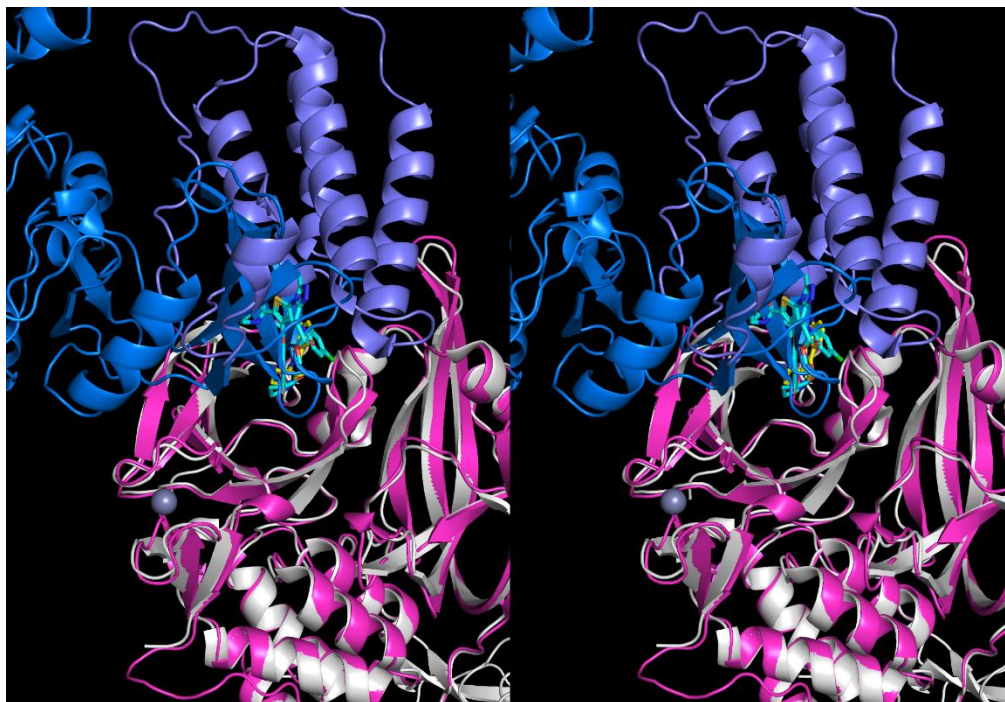


Figure 15. (A) Stereo view of the approximate contact footprint of CK1 $\alpha$  (blue) on cereblon (magenta) in the ternary cereblon-lenalidomide-CK1 $\alpha$  complex. (B) The saddle-shaped CK1 $\alpha$ -cereblon interface in the ternary structure looking from the bottom of the degrader pocket in cereblon (magenta ribbon) toward CK1 $\alpha$  (blue surface). The  $\beta$ -hairpin of CK1 $\alpha$  fits within the circular  $\beta$ -sheet containing the degrader binding site (visible as a cluster of aromatic side chains). (C) Stereo view of the approximate contact footprint of BRD4 on cereblon in 6BN7, which is far smaller than that of CK1 $\alpha$ . Unlike CK1 $\alpha$ , pocket L resides outside of the BRD4-cereblon interface. (D) Stereo view of CK1 $\alpha$ -cereblon from the ternary complex (blue and magenta, respectively) overlaid on BRD4-cereblon (purple and white, respectively) (PDB code = 6BN7). BRD4 and CK1 $\alpha$  bind adjacent to and opposite to the N-terminal domain of cereblon, respectively. The Type II bifunctional degrader dBET23 plays a different role than the Type I degrader lenalidomide in



recruiting substrates. Type II degraders are predicted in this work to speed the substrate  $k_1$  and slow the  $k_{-1}$  by augmenting the small cereblon-substrate footprint, whereas Type I degraders are predicted to primarily speed the substrate  $k_1$ .

## Materials and methods

We used a top-down approach to study TPD using multi-scale Biodynamics principles. First, we explored the generic TPD pathways for Type I and II degraders using ordinary differential equation-based models of the constituent multi-flux systems (noting that we used these models to deduce the optimal kinetic relationships among the species but did not simulate the time-dependent behaviors of the pathways). We then used WATMD-calculated solvation fields to deduce the solvation-powered structural dynamics governing non-equilibrium binary cereblon-degrader and ternary cereblon-degrader-substrate complex formation, including large-scale rotation of the C-terminal (CULT) domain relative to the N-terminal domain observed in 5FQD versus 6H0F (which we speculate facilitates access of the E2-ubiquitin (Ub) subunit to the terminal Ub of the growing poly-Ub chain during processive ubiquitylation). We calculated the solvation fields of cereblon (focused on the degrader and substrate binding sites), LMW degraders, and the neo substrate CK1 $\alpha$  (focused on the cereblon binding site) separately and in combination using the structures summarized below and in Tables 1 and 2:

- 1) Cereblon +/- DDB1 (denoted as structures A\_closed and A\_closed', respectively).
- 2) Cereblon-CK1 $\alpha$ , from which lenalidomide was removed (denoted as structure B).
- 3) Cereblon-lenalidomide, from which CK1 $\alpha$  was removed (denoted as structure C).
- 4) Cereblon-lenalidomide-CK1 $\alpha$  (the original 5FQD structure, denoted as structure D).

- 5) CK1 $\alpha$ , from which cereblon and lenalidomide were removed (denoted as structure E).
- 6) Cereblon + DDB1 (derived from 6H0F), the C-terminal (CULT) domain of which is rotated relative to the N-terminal domain (denoted as structure A<sub>open</sub>).
- 7) Mouse cereblon, in which the N-terminal domain is absent.

The software and methods are summarized briefly here, and in detail in our previous work.<sup>4</sup> The simulations were performed using AMBER 20 (University of California, San Francisco) in explicit TIP3P water to simulate the solvation dynamics (which we refer to as “SD simulations” due to the focus on water) of both the LMW and HMW structures and complexes. We used WATMD V9 was to analyze the dynamic water exchanges between bulk solvent and a three-dimensional grid of 1 Å<sup>3</sup> voxels proximal to the solute surfaces. The accumulated occupancy of water oxygen (O) and hydrogen (H) atoms within each voxel was calculated over the last 10 ns of a 100 ns trajectory (Figure 16). Such occupancies are governed by the water exchange rates to/from each voxel, including water donor versus acceptor preferences, as described in the Introduction. H-bond depleted solvation manifests as ultra-low occupancy voxels (denoted as ULOVs) represented by tiny white spheres exhibiting little or no preference for water H or O. Such solvation resides in the left tail of the Gaussian distribution of water occupancies reported in our previous work.<sup>4</sup> H-bond enriched solvation manifests as high occupancy voxels residing in the right tail of the Gaussian distribution (denoted as HOVs) represented by larger spheres (the radii of which are proportional to water occupancy) exhibiting a range of preferences for water H or O, including:

- 1) High preference for H (bright blue).
- 2) Moderate preference for H (light blue/purple).
- 3) No preference for either (white).

4) Moderate preference for O (pink/light red).

5) High preference for O (bright red).

Table 1. The HMW solutes (proteins) included in our study (noting that 5FQD is the human ortholog of the protein complex).

Name	PDB code	Reference code
Apo-like* cereblon (closed, -DDB1)	Time-averaged (from 5FQD <sup>40</sup> )	A_closed
Apo-like cereblon (closed, +DDB1)	Time-averaged (from 5FQD)	A_closed'
Apo-like cereblon (closed, +DDB1 with restraints applied in SD simulation)	Time-averaged (from 5FQD)	A_closed''
Apo-like cereblon (open, +DDB1)	Time-averaged (from 6H0F <sup>33</sup> )	A_open
Cereblon-CK1 $\alpha$ (closed)	Time-averaged (from 5FQD)	B
Cereblon-lenalidomide (closed, -DDB1)	Time-averaged (from 5FQD)	C
Cereblon-lenalidomide (closed, +DDB1)	Time-averaged (from 5FQD)	C'
Cereblon-lenalidomide (closed, -DDB1, -N-terminal peptide)	Time-averaged (from 5FQD)	C''
Cereblon-lenalidomide-CK1 $\alpha$ (closed)	5FQD (chains B and C)	D
CK1 $\alpha$	Time-averaged (from 5FQD)	E
Cereblon-lenalidomide-CK1 $\alpha$ (closed)	Time-averaged (from 5FQD)	F
Cereblon-pomalidomide-IKZF1 (open)	6H0F	G
Apo cereblon CULT (mouse) <sup>#</sup>	Time-averaged (from 3WX2 <sup>35</sup> )	H
Lenalidomide-cereblon CULT (Ms-CI4)	4V30 <sup>30</sup>	I
BRD4-cereblon (closed)	6BN7 <sup>41</sup>	J
Uridine-cereblon docking model	Time-averaged (from 5FQD)	K

\*Apo-like cereblon refers to the substrate and degrader unbounded state of the protein modeled from 5FQD. <sup>#</sup>Mouse cereblon (which lacks the N-terminal domain present in the human protein) is the only experimentally determined apo cereblon structure reported as of this writing.

Table 2. The LMW solutes included in our study.

Name	Chemical structure	PDB code
Lenalidomide	<a href="https://pubchem.ncbi.nlm.nih.gov/compound/Lenalidomide">https://pubchem.ncbi.nlm.nih.gov/compound/Lenalidomide</a>	Extracted from 5FQD
Thalidomide	<a href="https://pubchem.ncbi.nlm.nih.gov/compound/Thalidomide">https://pubchem.ncbi.nlm.nih.gov/compound/Thalidomide</a>	Extracted from 4CI1 <sup>42</sup>
Uridine	<a href="https://pubchem.ncbi.nlm.nih.gov/compound/Uridine">https://pubchem.ncbi.nlm.nih.gov/compound/Uridine</a>	Modeled

The solvation fields were superimposed on the corresponding time-averaged solute structures (generated by WATMD) and visualized using PyMol 2.4.2 (Schrodinger, LLC).

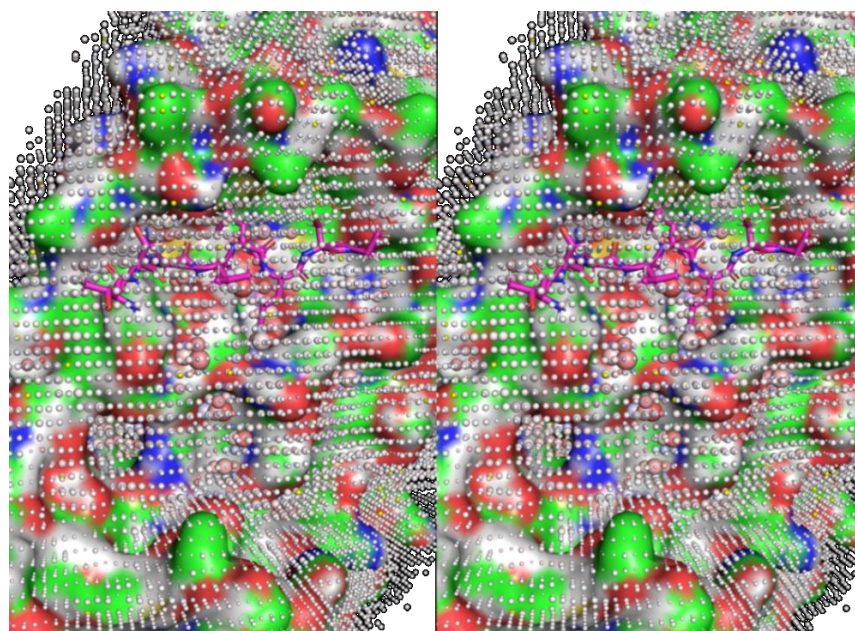


Figure 16. Stereo view of predicted water occupancies within a three-dimensional grid of voxels around a solute surface exemplified for COVID M<sup>PRO</sup> (surface colored by atom type)<sup>4</sup> generated using PyMol 2.4.2 (Schrodinger, LLC) (noting that all WATMD results reported in this work were generated using PyMol). Bulk-like solvation (centered around the mean of the Gaussian curves shown in reference <sup>4</sup>) is represented by small white spheres, H-bond enriched solvation by larger spheres, and H-bond depleted solvation by tiny spheres (non-discernable in this figure). Bulk solvent, residing above the outermost voxel layer shown, was omitted from the drawing (and all

other drawings in this work) for the sake of clarity. The spheres are color-coded according to the preference of the corresponding voxel for water H or O in shades of blue and red, respectively (see text). Water occupancy and H versus O preference within each voxel are determined by the local the solute environment.

### **HMW SD simulation protocol**

Briefly, the structures listed in Table 1 were prepared using the default settings of the PPREP tool in Maestro 12.7 (Schrodinger, LLC). The structures were then solvated in explicit TIP3P water using AMBER tools, followed by a 100 ns SD simulation using AMBER 20: PMEMD CUDA and the ff14SB force-field. WATMD calculations were performed on the last 10 ns (40,000 frames) of each trajectory. All of the resulting structures and voxel grids were aligned to a common reference frame consisting of cereblon or CK1 $\alpha$  in structure F (see Table 1) using the align function of PyMol.

### **LMW SD simulation protocol**

The LMW structures listed in Table 2 were extracted from crystal structures or modeled using the Build function of Maestro (as noted in the table) and pre-processed with GAFF in ANTECHAMBER (charges generated using the AM1-BCC method). The structures were solvated in explicit TIP3P water using AMBER tools, followed by 100 ns restrained SD simulation using AMBER 20 (ff99SB force-field). WATMD calculations were performed on the last 10 ns (40,000 frames) of each trajectory. All of the resulting structures and voxel grids were aligned to a common reference frame consisting of lenalidomide in structure F (see Table 1) using the align function of PyMol.

## Uridine docking protocol

We built the uridine structure de novo using Maestro, followed by minimization in MacroModel (default settings). We then docked the structure in the degrader pocket of cereblon taken from 5FQD using the Glide XP tool of Maestro (default settings), which resulted in a single pose that we used as input for LMW WATMD calculations.

## Results

### The general dependence of $\Delta G_{\text{enriched}}$ and $\Delta G_{\text{depleted}}$ on solute surface composition, topology, and curvature

Here, we summarize the general learnings from our previous WATMD calculations on COVID M<sup>pro4</sup> and LMW hERG blockers,<sup>5</sup> together with cereblon and the degrader drugs studied in this work. Preferences among the sterically-accessible conformational states of a given solute are determined principally by solvation H-bond free energy (versus interatomic energy contributions), which varies spatially across solute surfaces, as follows:

- 1) **Type, number, and distribution** of polar/charged acceptor/donor groups versus non-polar groups across non-buried surface regions.
- 2) **H-bond propensity (charge, polarity)** of solute donor/acceptor groups relative to that of water-water H-bond groups.
- 3) **Regional concavity** (depth below the mean surface elevation) versus convexity (height above the mean surface elevation) across the surface. Optimization of the solvating water-water-solute H-bond network geometry depends on ordering-dependent adaptation to the solute surface geometry, resulting in varying degrees of entropic loss.

- 4) **Regional surface curvature magnitude**, which determines the requisite degree of geometric H-bond network adaptation and ordering.
- 5) **Solvent accessibility** of each donor/acceptor group, as reflected in the local solvent exposed surface area, which ranges between uni- and omni-directional at high versus low exposed surface regions, respectively.

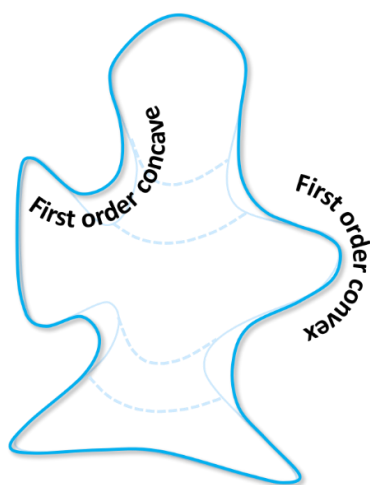
The aforementioned characteristics of solvent exposed surfaces differ fundamentally between LMW and HMW solutes due to the significantly greater folding/side chain packing ability of the latter relative to the former, together with the lack of buried surfaces in LMW solutes, as follows:

- 1) LMW solute surfaces are contributed largely by individual loosely packed/unpacked substituents/R-groups, whose surfaces range from planar (e.g., the core of aromatic rings) to convex (e.g., acyclic groups and the edges of aromatic rings) to concave (e.g., inter-substituent connections). We refer to LMW surface geometries as first order convexity/concavity (Figure 17A). Single strongly polar groups projecting from convex surfaces into solvent promote  $G_{\text{enriched}} < (\text{or } \ll) G_{\text{bulk solvent}}$  (Figure 17B), whereas non-polar convex surfaces may promote  $G_{\text{enriched}} > (\text{or } \gg) G_{\text{bulk solvent}}$  (depending on the magnitude of the surface curvature and degree of entropic loss due to additional ordering at such regions) (Figure 17C).
- 2) HMW solute surfaces consist largely of the outer boundaries of their packed side chains and backbones, which we refer to as second order convexity (Figure 18A) and concavity (Figure 18B). Co-localization of remote H-bond groups in the monomer sequence may occur in concave surface regions due to folding. H-bond enriched solvation is often promoted by such environments (Figure 18C). High convex surface curvature in proteins

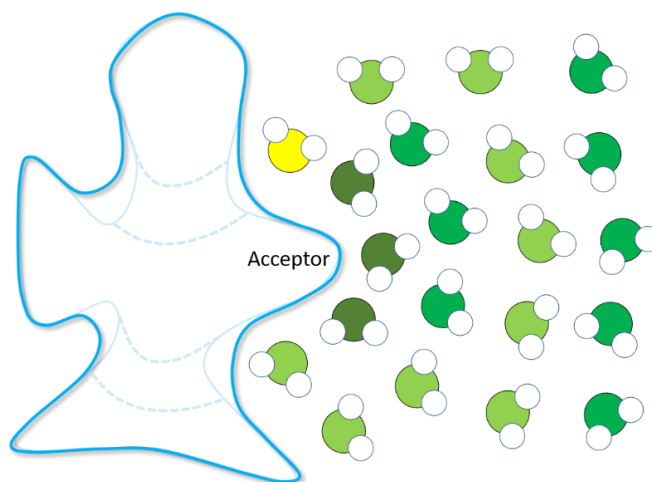
maps largely to loops and  $\beta$ -turns. Buried water channels/cavities may additionally arise in folded structures.

- 3) Foldability/packability of HMW solutes enables far greater solvation-powered rearrangeability of surface convexity/concavity, curvature, and polar/non-polar surface composition compared with LMW solutes.

A



B





C

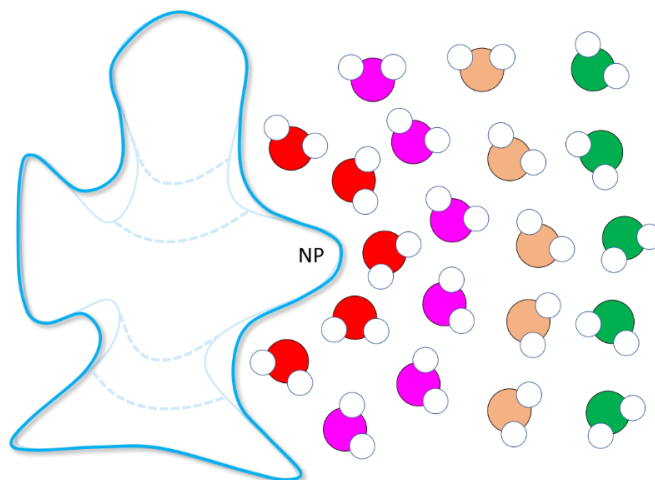
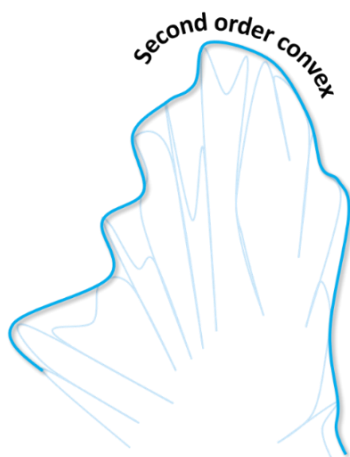


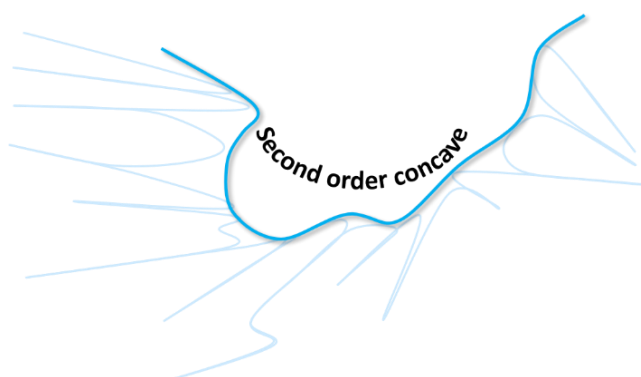
Figure 17. Hypothetical cartoons exemplifying the general properties of the non-isotropic H-bond fields of LMW solutes, which are strongly influenced by solute surface properties, including convexity/concavity, local curvature, and polar/non-polar composition. Solvation fields seek to minimize their free energy in the presence of dissolved solutes (which may reside above or below that of bulk solvent) via intra- and/or inter-solute rearrangements corresponding to optimal solute surface curvature, convexity/concavity, and polar/non-polar group exposure. Depending on the local polar/non-polar surface composition, high surface curvature may result in severe distortion of the water H-bond network geometry, accompanied by the loss or gain of free energy (noting that entropic losses are expected at highly distorted H-bond network regions due to additional ordering in such regions). (A) The surfaces of LMW solutes are contributed largely by individual (i.e., non-packed) chemical groups, which exhibit varying degrees of concave and convex curvature (which we refer to as first order curvature), ranging from planar to severely curved (e.g., at the core versus edges of aromatic substituents, respectively). Convex and concave surfaces map largely to substituents and inter-substituent connections, respectively. (B) Polar convex surfaces incur a tradeoff between favorable H-bond enrichment and unfavorable warping of the water H-

bond field. (C) Non-polar convex surfaces (denoted as NP) of LMW solutes induce varying degrees of strain on the H-bond field, which are both H-bond depleted and warped.

A



B



C

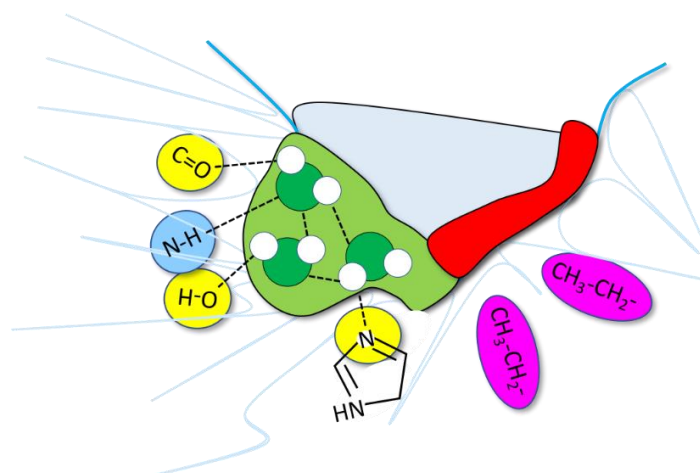


Figure 18. As for LMW solutes, the non-isotropic solvation fields of HMW solutes are strongly influenced by local surface curvature and polar/non-polar composition. The surfaces of HMW solutes are contributed largely by packed side chain and backbone moieties, where curvature is a collective property of the packed moieties (which we refer to as second order curvature). (A) Highly curved convex surfaces are typically associated with loops and  $\beta$ -turns. (B) Highly curved concave surfaces typically map to the interfaces between secondary structure elements and domain interfaces. (C) Folded concave surfaces may result in co-localization of remote H-bond groups in the monomer sequence, promoting H-bond enriched solvation of such regions (light green background) interspersed with H-bond depleted solvation (red background).

### **Case study: A top-down analysis of E3 ligase-mediated protein degradation**

We set about to qualitatively analyze:

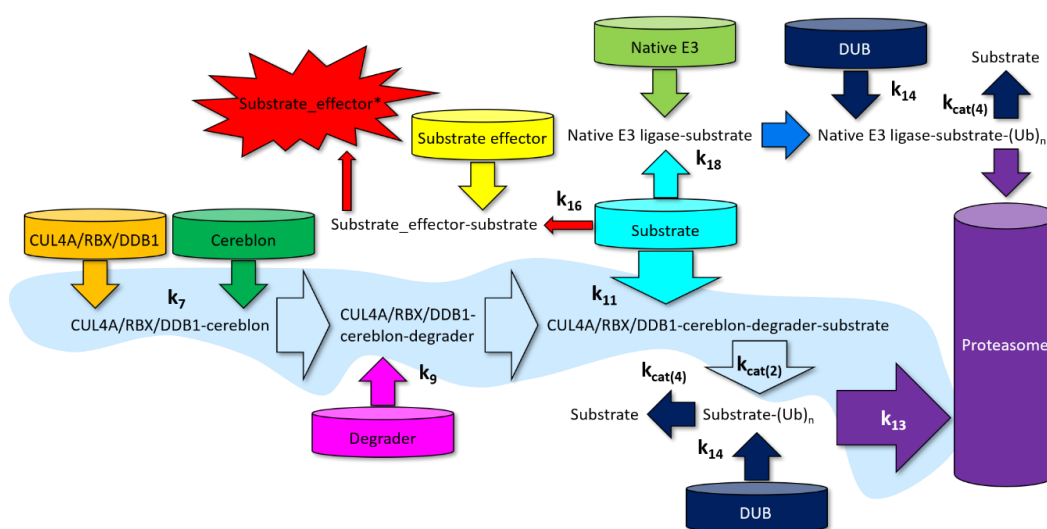
- 1) Proteasomal protein degradation via a symbolic mathematical model of a representative subset of the pathway in terms of:
  - a) The extrinsic and intrinsic rates of assembly of the substrate-degrader-bound E3 complex, buildup of the poly-ubiquitylated substrate species and decay thereof via deubiquitylation (the Yang of degradation) and proteasomal degradation (the Yin of degradation) (noting that Clocks were omitted for the sake of simplicity).
  - b) The optimal binding kinetics of Type I and II degraders needed to achieve the maximum rate of ternary complex formation (yielding the degradation rate  $\gg$  rate of substrate binding to its downstream effector) at the lowest possible degrader concentration.

- 2) The following solvation properties that we used to deduce the functionally important non-equilibrium state transitions of cereblon, including binary and ternary complex formation:
  - a) Complementarity between the solvation fields of cereblon and Type I degraders governing binary complex formation analyzed using WATMD (a solvation field perspective on structure-activity relationships).
  - b) The role of cereblon-Type I degrader complexes (e.g., cereblon-lenalidomide) in promoting substrate binding.
  - c) Complementarity between the binary degrader-cereblon solvation field and that of CK1 $\alpha$ , which governs both ternary complex formation and the lifetime thereof (chosen based on the availability of a published crystal structure of the ternary complex).
  - d) The solvation field of cereblon-degrader-CK $\alpha$  vis-à-vis stability of the ternary complex.
  - e) The role of buried water channels in powering the intra- and intermolecular state transitions of cereblon.
  - f) Rotation of the CULT domain relative to the N-terminal domain (denoted as the “open” state) (6H0F).

## Analysis of proteasome-mediated protein degradation pathway dynamics

The cereblon-mediated protein degradation pathway examined in this work is outlined schematically in Figure 19 and formulated as a symbolic mathematical model provided in the Supplementary Information. The model includes native and non-native (“neo”) substrate recruitment by endogenous LMW degraders and drugs.

A



B

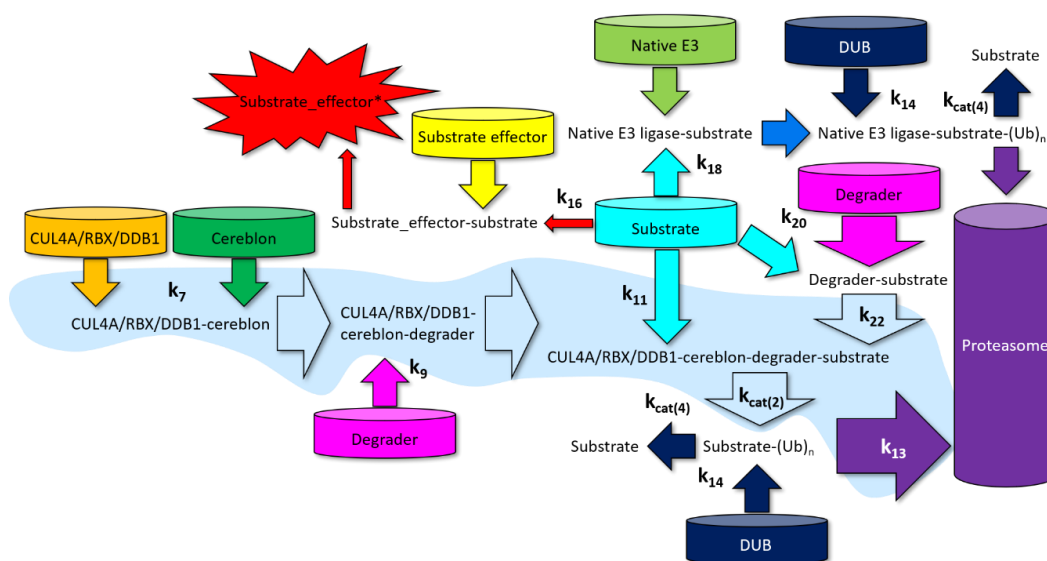


Figure 19. Cartoon of the cereblon-mediated degradation pathway schema detailed in the text (including DUB counter-balancing), with the main trunk highlighted in light blue. Some reactions and rate constants have been condensed and Clocks omitted for the sake of simplicity. (A) Type I degraders (e.g., lenalidomide, pomalidomide) and their targeted substrates (e.g., CK1 $\alpha$ ) and downstream effector of the target substrate. The substrate degradation rate ( $k_{cat(4)}$ ) is optimally  $\gg$  the rate of downstream substrate-mediated effector activation ( $k_{16}$ ) and the rate of binary cereblon-degrader formation ( $k_9$ ) is optimally  $\approx$  the rate of cereblon buildup ( $k_7$ ). (B) Type II degraders (e.g., dBET23) and their targeted substrates (e.g., BRD4). The rate of degrader-substrate binding ( $k_{20}$ ) is ideally  $>$  the rate of degrader-cereblon binding ( $k_9$ ), which promotes substrate inhibition and degradation, while limiting cereblon inhibition.

**Achieving efficacious cereblon-degrader-substrate dynamic occupancy (corresponding to a threshold inhibition of the downstream neo substrate effectors) at the lowest possible degrader concentration depends on kinetic tuning between the degrader and the overall pathway, and in particular:**

**For Type I degraders (Figure 19A)**

- 1) Degradation rate ( $k_{13}$ )  $\gg$  substrate + substrate\_effector  $k_{on}$  ( $k_{16}$ ).
- 2) Substrate + CUL4A/RBX1/DDB1-cereblon-degrader  $k_1$  ( $k_{11}$ )  $\approx$  [substrate](t).
- 3) Fast degrader + CUL4A/RBX1/DDB1-cereblon  $k_{on}$  ( $k_9$ ).
- 4) CUL4A/RBX1/DDB1-cereblon-[degrader]  $k_{off}$   $<$  CUL4A/RBX1/DDB1-cereblon-degrader + substrate on-rate ( $k_1 \cdot$  [substrate](t)).
- 5) Fast degradation via the native ligase (i.e., fast  $k_{18}$ ).

### **For Type II degraders (Figure 19B)**

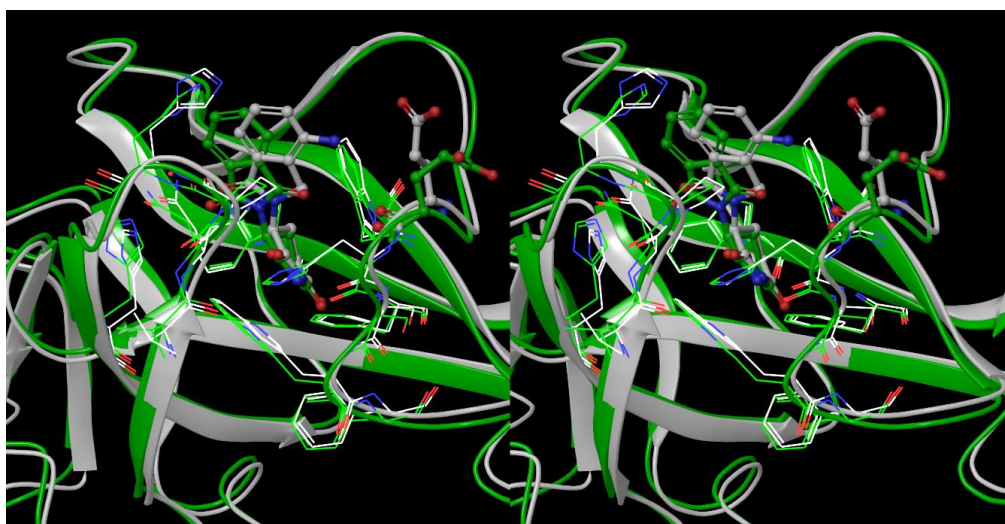
- 1) Degradation rate ( $k_{13}$ )  $\gg$  substrate + substrate\_effector  $k_{on}$  ( $k_{16}$ ).
- 2) Degradation rate  $k_{on}$  ( $k_{20}$ )  $\approx$  [substrate](t).
- 3)  $CUL4A/RBX1/DDB1$ -cereblon-[degrader-substrate]  $k_{off} < CUL4A/RBX1/DDB1$ -cereblon-degrader-substrate  $k_{cat}$ .
- 4) Degradation rate  $k_{on}$  ( $k_{20}$ )  $>$  degradation rate  $k_{on}$  ( $k_9$ ), given that  $CUL4A/RBX1/DDB1$ -cereblon is inhibited in the latter case, and the substrate is both inhibited and degraded in the former.
- 5)  $CUL4A/RBX1/DDB1$  + degrader-substrate  $k_{on}$  ( $k_{22}$ )  $>$   $CUL4A/RBX1/DDB1$  + degrader  $k_{on}$  ( $k_9$ ).
- 6)  $k_9$  is rate-limiting when slower than  $k_{20}$  but must be balanced against  $k_{16}$ .
- 7) Fast ternary complex formation in both directions (i.e., fast  $k_{22}$  and  $k_{11}$ ).
- 8) Degradation is faster than deubiquitylation (i.e.,  $k_{13} \gg k_{14}$ ).
- 9) Fast degradation via the native ligase (i.e., fast  $k_{18}$ ).

### **Thalidomide-cereblon-CK1 $\alpha$ complex formation is precluded by steric incompatibility**

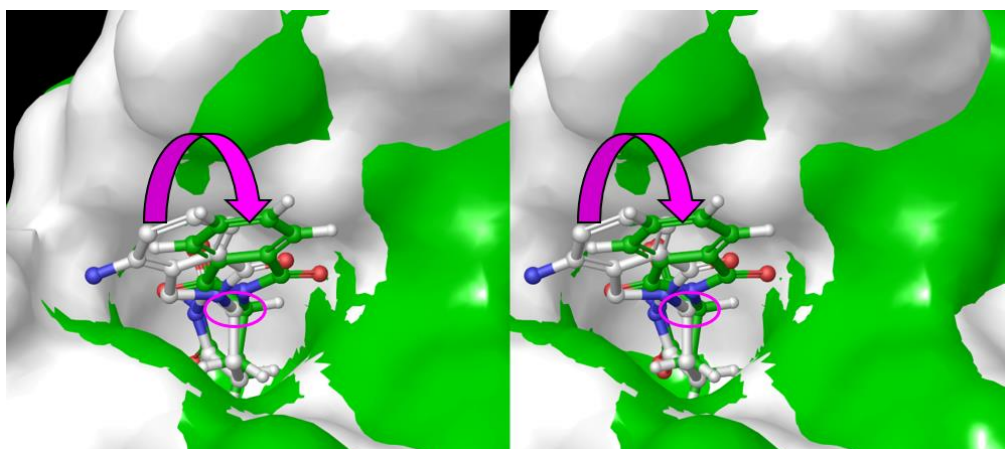
CK1 $\alpha$  is recruited by lenalidomide-cereblon, but not thalidomide-cereblon, which can be explained largely on the basis of steric incompatibility between the phthalimide ring of thalidomide and substrate  $\beta$ -hairpin.<sup>43</sup> The backbones and side chains of thalidomide- and lenalidomide-bound cereblon are well-aligned, ruling out the possibility of binding mode differences per se. Rather, the phthalimide ring of thalidomide is rotated about the glutarimide nitrogen (N) toward the  $\beta$ -

hairpin of CK1 $\alpha$  (Figure 20A), stemming largely from puckering differences in the glutarimide ring in the presence of a second unique carbonyl oxygen (O) (noting that the glutarimide ring binding geometry is highly similar in both cases) (Figure 20B). The phthalimide ring position in lenalidomide is well-packed with the  $\beta$ -hairpin of CK1 $\alpha$  (Figure 20C), which clashes in thalidomide (Figure 20D).

A

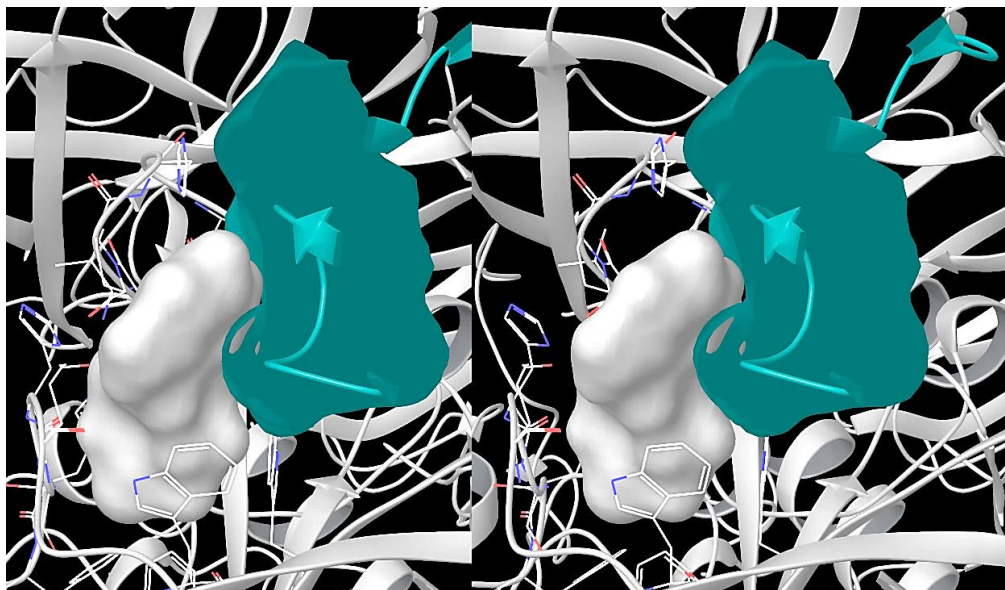


B





C



D

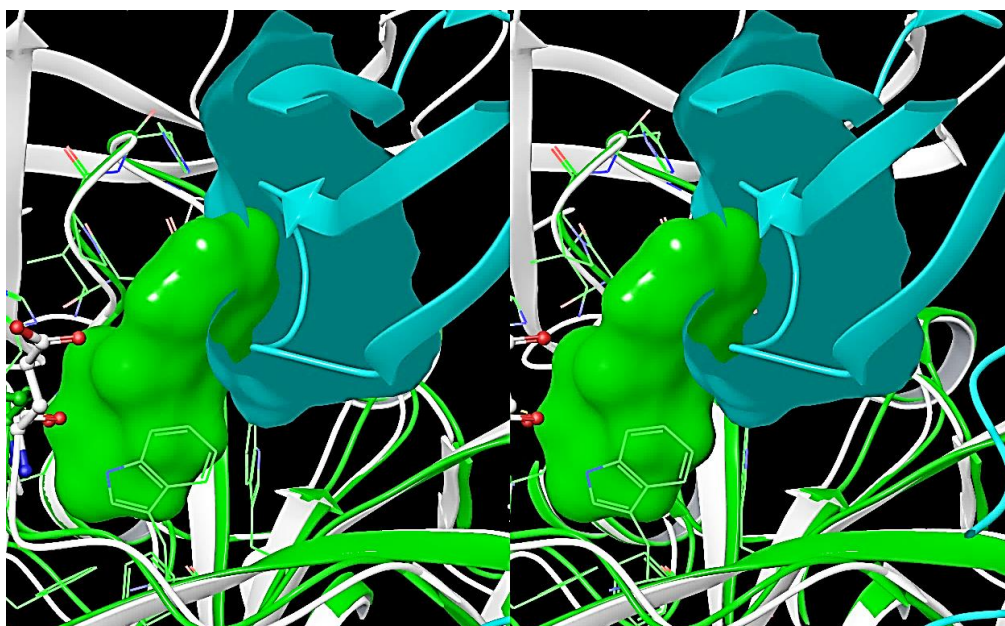


Figure 20. Stereo views of cereblon-thalidomide (S isomer) extracted from PDB code = 4CI1 (green) overlaid on cereblon-lenalidomide-CK1 $\alpha$  (S isomer) extracted from 5FQD (white). (A) A large rotation in the phthalimide group position between the two structures is apparent, resulting

largely from puckering differences in the glutarimide ring. (B) Overlay of thalidomide and lenalidomide based on the CULT domain of cereblon, demonstrating high conservation of the glutarimide ring binding mode, despite rotation of the phthalimide group (magenta arrow) about the ring N (magenta circle). Binding of thalidomide to cereblon is accommodated by the openness of the substrate-facing side of the ligand pocket in the absence of CK1 $\alpha$ . (C) The phthalimide ring position in lenalidomide is well-packed with the  $\beta$ -hairpin of CK1 $\alpha$ , which explains the selectivity of this compound for this substrate. (D) The phthalimide ring position in thalidomide is sterically incompatible with the  $\beta$ -hairpin of CK1 $\alpha$  (teal blue) which precludes cereblon-thalidomide-CK1 $\alpha$  ternary complex formation.

### **The structural dynamics of cereblon and its binding interfaces**

Next, we set about to study the structural dynamics of cereblon based on the 100 ns SD simulations described in Materials and methods, the prerequisite for WATMD solvation field predictions (noting that only high frequency protein rearrangements occur on this timescale). Our preference is to use experimental rather than predicted structures for this purpose, given that solvation-dependent free energy contributions are grossly underweighted in conventional MD simulations and other force-field-based approaches, as noted above. However, local rearrangements may be predicted from conventional MD simulations on a case-by-case basis in the absence of experimental structures. Our study focused on the mechanism by which cereblon is switched to its substrate binding-competent state by bound degraders (focused qualitatively on the intrinsic rates), the possible scenarios of which consist of either increased substrate  $k_1$  via stabilization of the binding-competent form of the substrate binding surface, or decreased  $k_{-1}$  via degrader-contributed augmentation of the H-bond depleted solvation of the substrate binding surface. The small fractional increase in the substrate-binding surface area and H-bond depleted solvation free energy

contributed by bound degraders is inconsistent with the total absence of substrate binding in the degrader-unbound state (Figure 21) (the H-bond depleted solvation of unbound and bound lenalidomide and thalidomide is addressed in a later section).

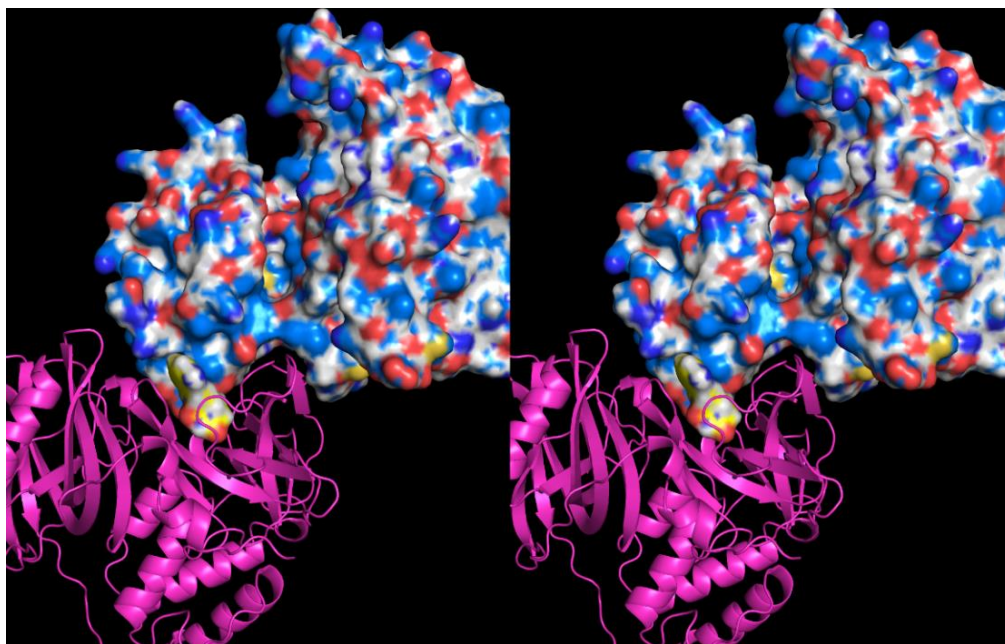


Figure 21. Stereo view of the CK1 $\alpha$ -cereblon interface (blue surface and magenta cartoon, respectively) in the ternary structure F. Overlay of structures F and A, looking from CK1 $\alpha$  (blue) into pocket L. The small contributions of lenalidomide (yellow) to the H-bond depleted solvation and overall interfacial footprint are inconsistent with the complete absence of substrate binding in the degrader-unbound state of cereblon.

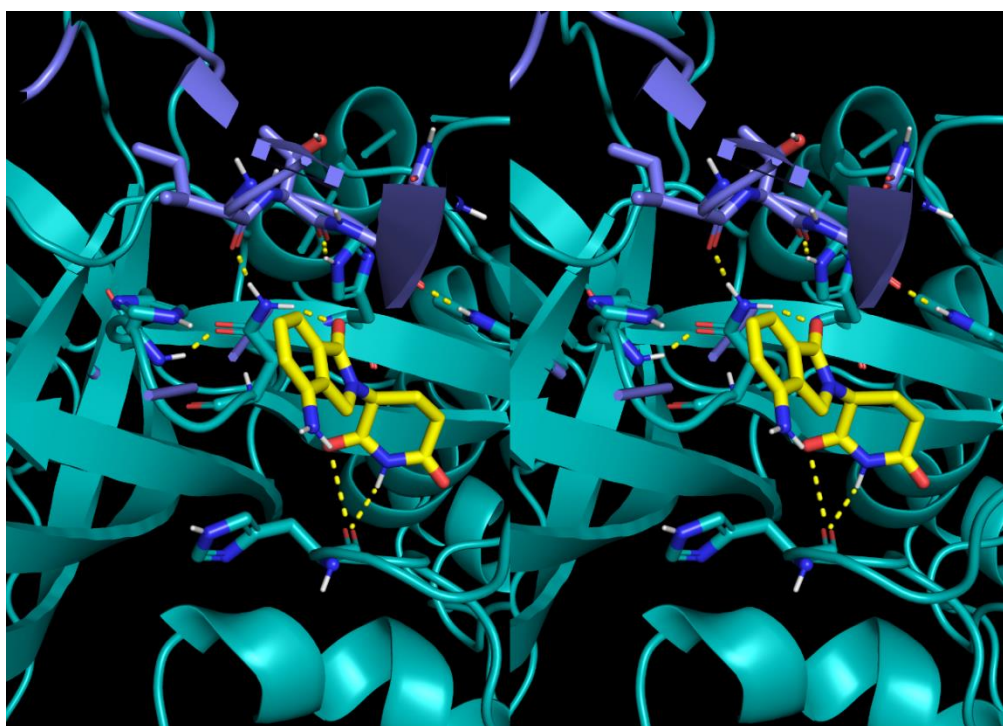
We therefore postulate that association between cereblon and its cognate substrates depends on degrader-mediated stabilization of the substrate binding site, resulting in dramatic speeding of substrate  $k_1$ , as follows:

- 1) The empty form of the degrader pocket (widely referred to as the Trp-box, and denoted in this work as pocket L) cycles between degrader binding-competent and incompetent states

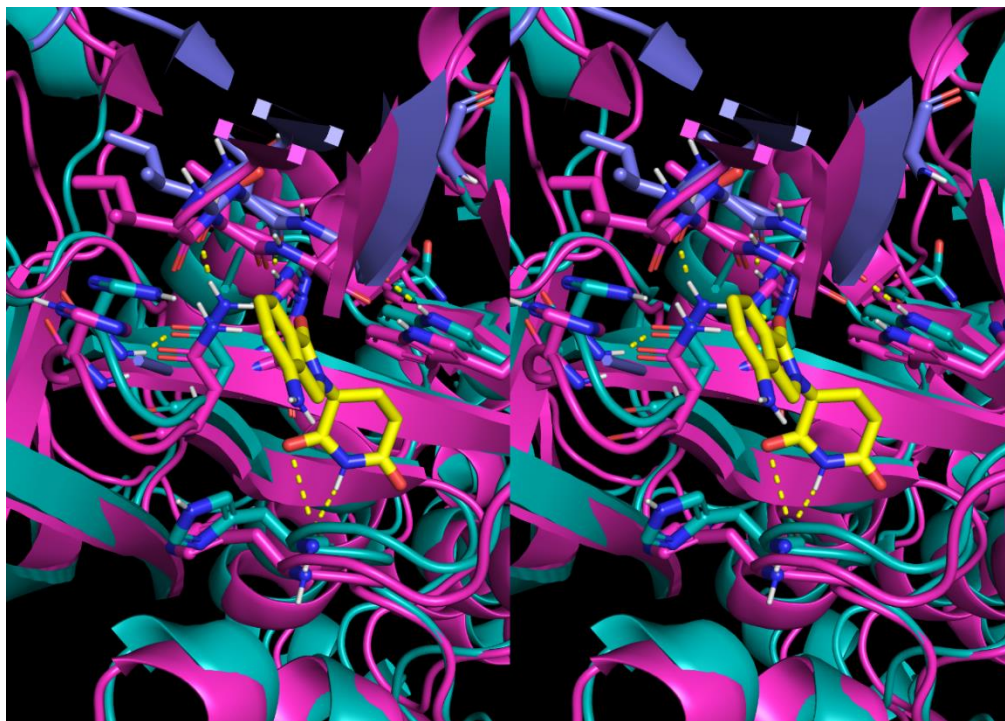
in our simulations, which correspond to fully formed and collapsed states of the pocket, respectively (noting that the fully formed empty pocket in human cereblon has not been observed experimentally as of this writing).

- 2) The substrate binding-competent state of cereblon corresponds to the degrader-bound state of the protein, in which key polar groups contributed by Asn305, His311, Arg327, and Trp354 are positioned for H-bonding to the  $\beta$ -hairpin and other residues of CK1 $\alpha$  (Figure 22A) (noting that these H-bonds are absent in the starting ternary structure D (Figure 22B), but present in the time-averaged structure (Figure 22C)). We simulated, but did not perform WATMD calculations on the cereblon-palomalidomide-IKZF1 structure (6H0F), which likewise forms interfacial H-bonds during the simulation, including between the  $\beta$ -hairpin and Asn305.

A



B



C

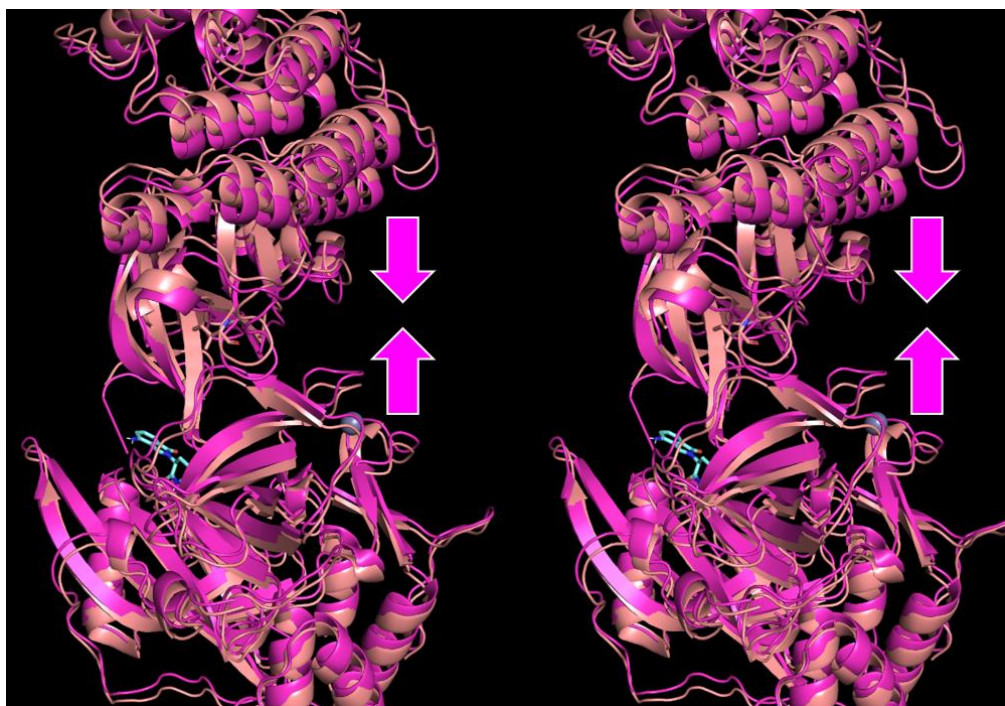
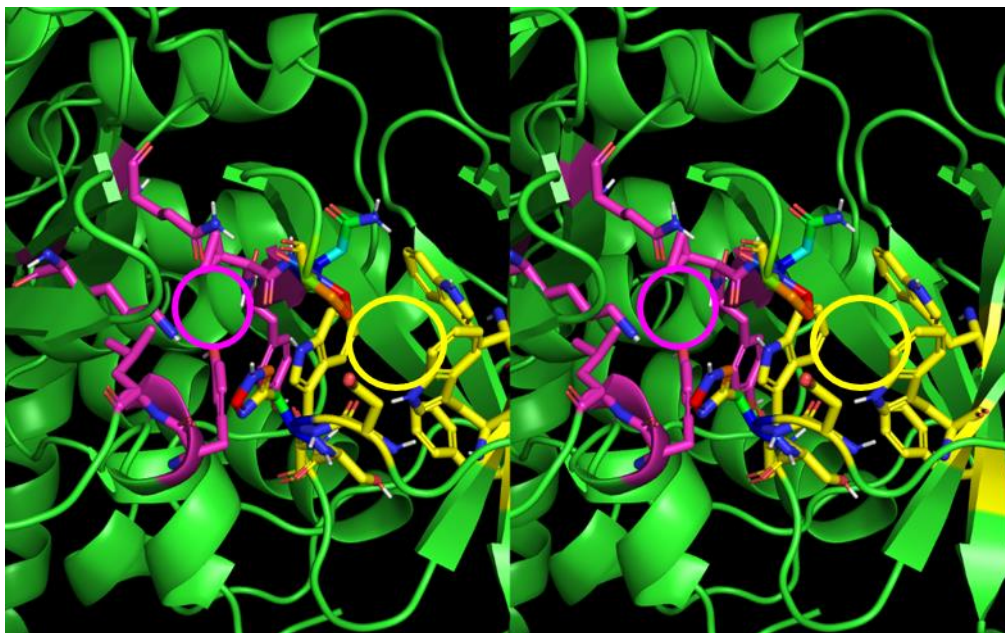


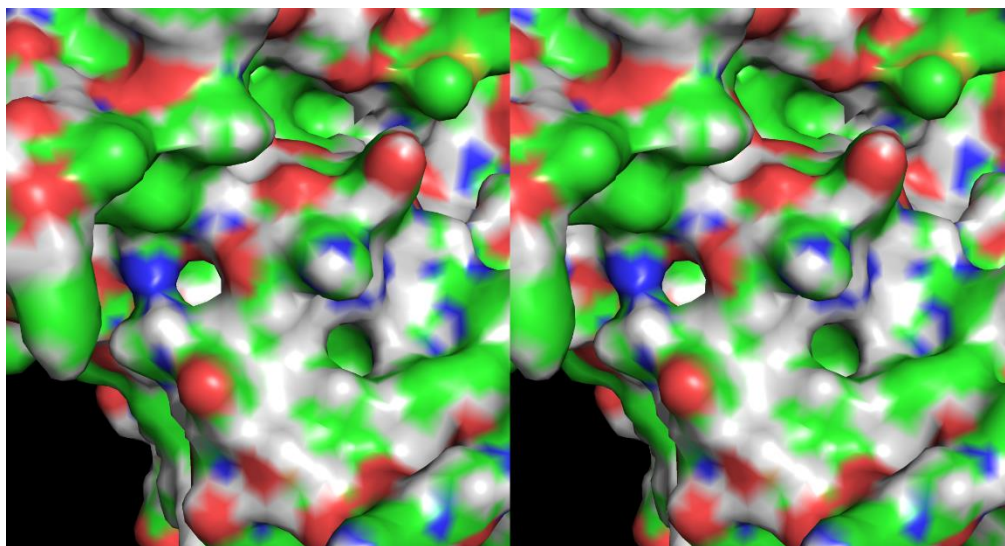
Figure 22. (A) The interfacial H-bonds (dotted yellow lines) between CK1 $\alpha$  (purple), cereblon (cyan), and lenalidomide (yellow). Simultaneous alignment of these H-bond groups among the two partners permits close approach of the partners during ternary complex formation (noting that desolvation of H-bond depleted water from the interface, which is located along the protein surfaces, is dependent on tight surface-surface complementarity). Water H-bond replacement partners consist of cereblon His311 ( $\epsilon$ -NH) to CK1 $\alpha$  Thr419 (backbone O), cereblon Trp354 (indole NH) to CK1 $\alpha$  Asn420 (backbone O), cereblon Asn305 (NH<sub>2</sub>) to CK1 $\alpha$  Ile418 (backbone O), cereblon Arg327 (guanidium NH) to CK1 $\alpha$  Glu423 (COO), lenalidomide phthalimide (O) to Asn305 (NH<sub>2</sub>), lenalidomide glutarimide (NH) to cereblon His332 (backbone O). We hypothesize that complementarity among these H-bond partners depends on the presence of a bound degrader. (B) Overlay of structures C (cyan) and D (magenta), showing the complete absence of interfacial H-bonds in D that are present in the time-averaged structure C. (C) Stereo view of structure F (magenta) overlaid on the initial ternary structure D (copper). The packing between cereblon and CK1 $\alpha$  is further optimized in structures F and B (not shown) via an approximately rigid body translation (magenta arrows).

Next, we analyzed the dynamic behavior of the residues lining pocket L by comparing the time-averaged structures A<sub>closed</sub>, A<sub>closed</sub>', A<sub>open</sub>, C, C', and H (Table 2). Pocket L is connected to a second transient pocket (denoted as pocket T) in the time-averaged structures A<sub>closed</sub>, C, and F, but not C' or D (in which it is occupied by the side chain of His378). Pockets L and T are lined by both unique and common residues (Figure 23A), and the surface openings reside at nearby regions of the protein (Figure 23B). The pockets and surface openings thereof are exemplified in Figure 23C for structure A<sub>closed</sub>.

A



B



C

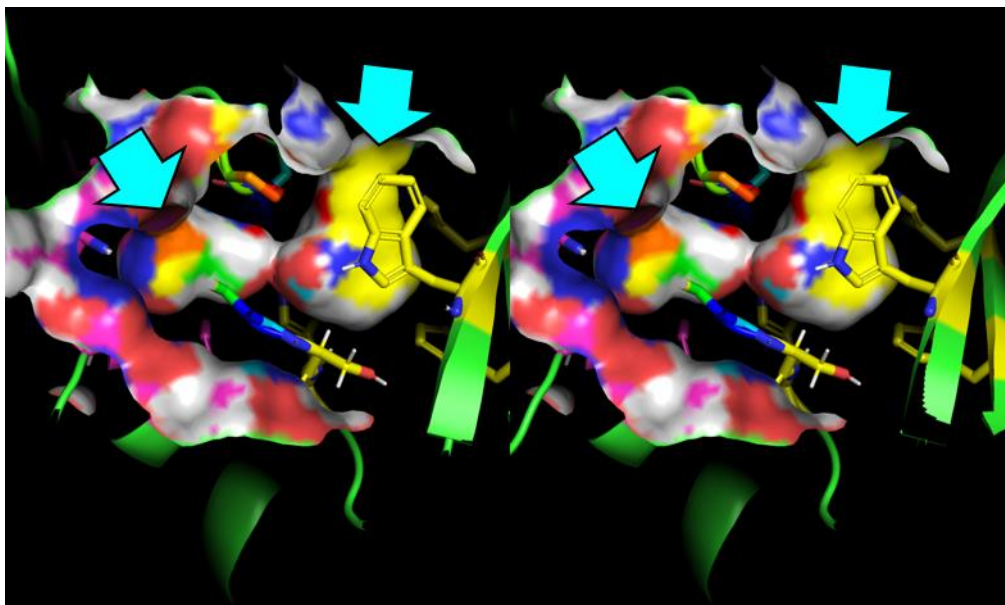


Figure 23. (A) Zoomed-in stereo view of the residues lining pockets L (Glu331, Ser333, Trp334, Trp340, Trp354, and Phe356) (yellow), T (Leu14, Tyr13, Gln54, Lys110, Tyr303, Val304, Pro306, and His332) (magenta), and both L and T (Asn305, Pro306, and His332) (color spectrum) in structure A. (B) Stereo view of the external surface openings of pockets T (northwest) and L (southeast) in structure A. (C) The buried surface of pockets L and T structure A.

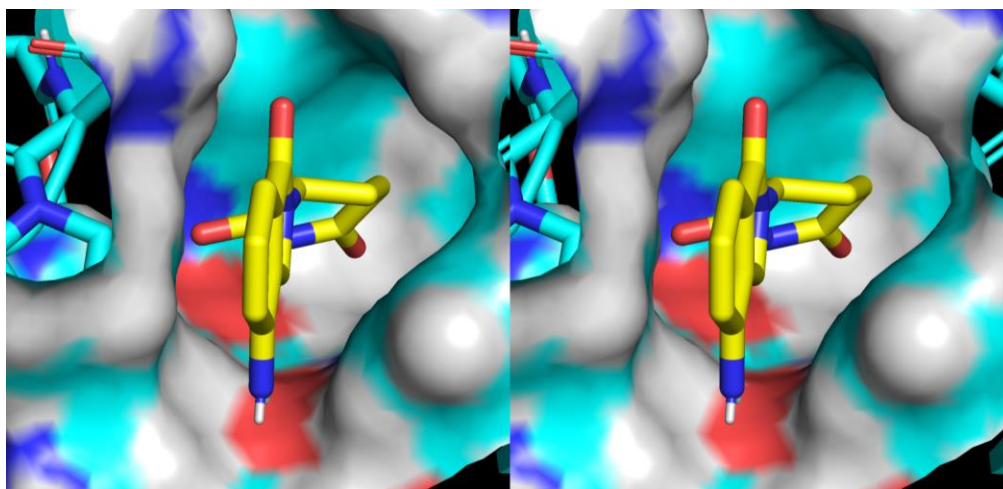
We inferred the dynamic behavior of pocket L based on the time-averaged structures as follows (recalling that the starting structures used in our simulations are biased toward the ternary complex in 5FQD):

- 1) Fully formed in the degrader-occupied structures C, C', and C'' and the empty partially restrained structures H and A\_closed'' (exemplified for structure C with lenalidomide shown for reference in Figure 24A).

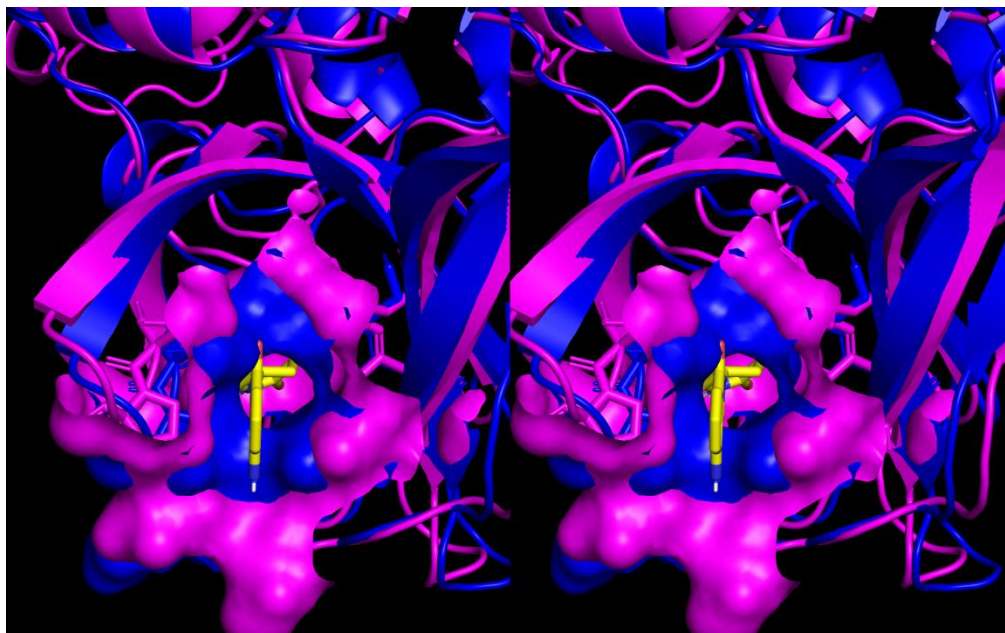


- 2) Partially formed in structure A\_closed (sampled exclusively in the hypothetical DDB1-unbound state) (Figures 24B), which is caused by counterclockwise shifts in the side chains of Trp334, Asn305, and Pro306 driven by twisting of the semi-circular  $\beta$ -strand 308-316 (Figure 24C).
- 3) Fully collapsed in structures A\_open and A\_closed' (exemplified for structure A\_open in Figures 24D and E). The collapsed pocket stems largely from rotation of the Trp340 side chain into the pocket, thereby retaining an H-bond between the indole NH and backbone O of Glu331, whose position shifts away from that in the other structures (e.g., structure F) (Figure 24F). We attribute this rearrangement to overweighting of the H-bond energy contribution in the AMBER force-field, which we postulate is right for the wrong reasons (addressed below).

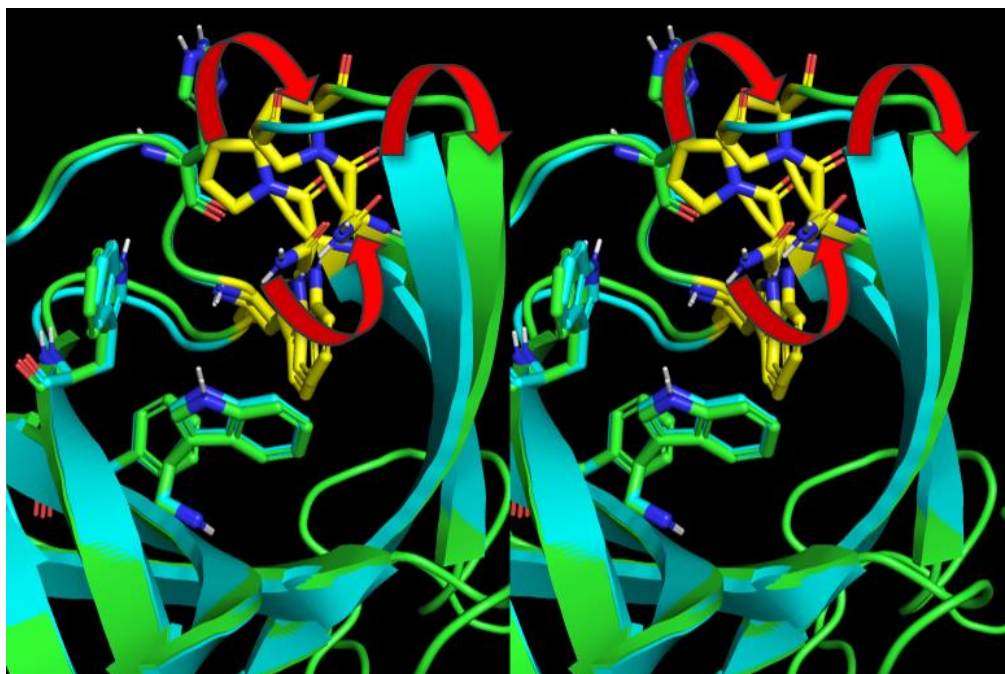
A



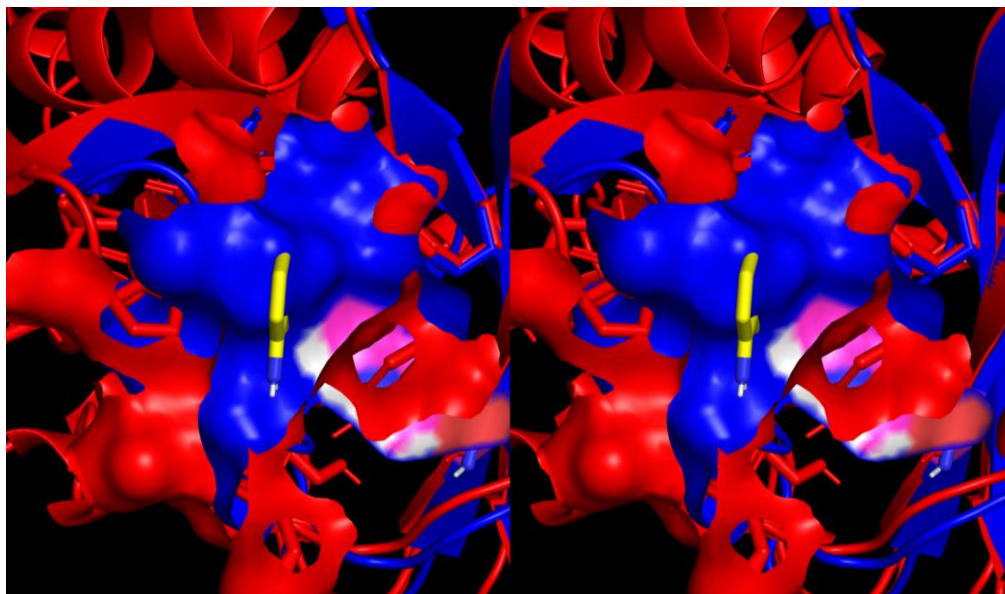
B



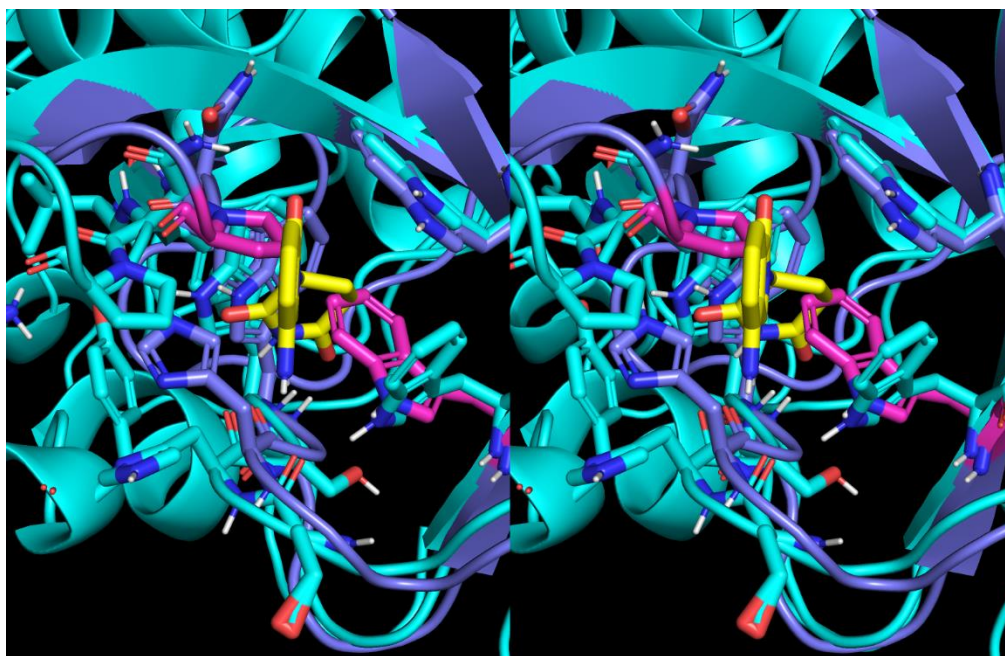
C



D



E



F

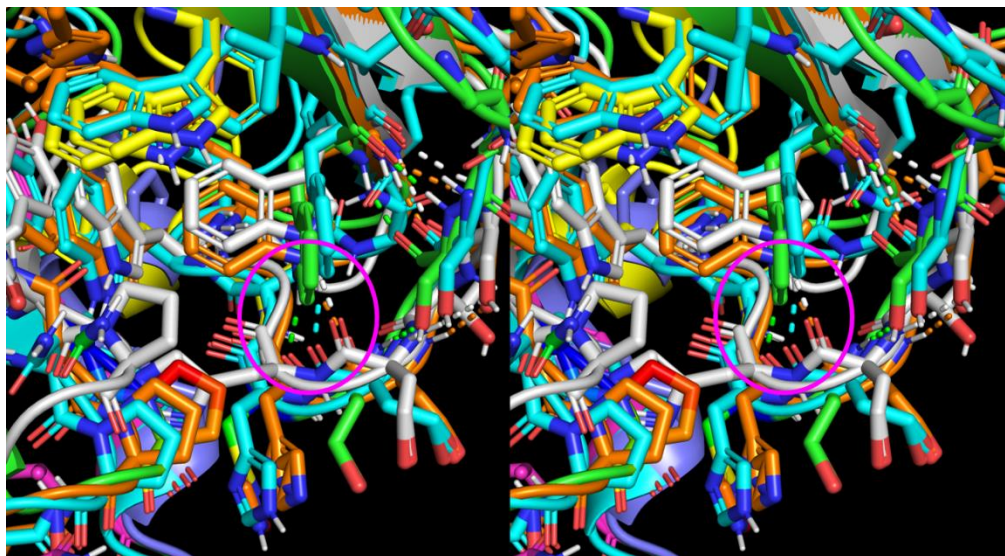


Figure 24. (A) Zoomed-in stereo view of lenalidomide (yellow) bound to pocket L in the binary structure C. (B) Zoomed-in stereo view of the opening of pocket L in structures A (blue) versus C (magenta), showing contraction of the opening in the empty versus lenalidomide-bound state. (C) Zoomed-in stereo view, showing the contracted diameter of the pocket L opening in structure A (green) relative to that of structure C (cyan) due to further twisting of  $\beta$ -strand 308-316 accompanied by shifts in the side chains of Trp334, Asn305, and Pro306 toward the pocket opening. (D) Zoomed-in stereo view of pocket L, which is fully collapsed in structure A<sub>open</sub> (blue) relative to structure C (red) due largely to occupation by Pro306 and Trp340. (E) Zoomed-in stereo view, showing the rearranged residues lining pocket L in structure A<sub>open</sub> (purple) relative to structure C (cyan). Pro306 and Trp340 are shown in magenta. (F) Zoomed-in stereo view, showing retention of the Trp340-Glu331 H-bond (circled in magenta) via Trp340 rotation in structures A<sub>closed</sub>' (orange) and A<sub>open</sub> (white) versus A<sub>closed</sub> (green) and C (cyan).

We assume that:

- 1) A<sub>closed</sub>' (derived from 5FQD with partial restraints, as described in Table 1) and mouse cereblon (3WX2 and the time-averaged structure H with partial restraints) are reasonable models for the fully formed pocket L as exists just prior to degrader binding, which is unobserved experimentally (Figure 25).
- 2) A<sub>closed</sub> (from which DDB1 was omitted) is a reasonable model for the partially formed pocket. The pocket and internal solvation in this predicted state are unstable relative to that of the collapsed and fully formed states, which we attribute to the omission of DDB1 in the time-averaged structure (addressed below).
- 3) A<sub>closed</sub>' (in which DDB1 was included) and A<sub>open</sub> (in which DDB1 was likewise included) are reasonable models for the fully collapsed pocket.

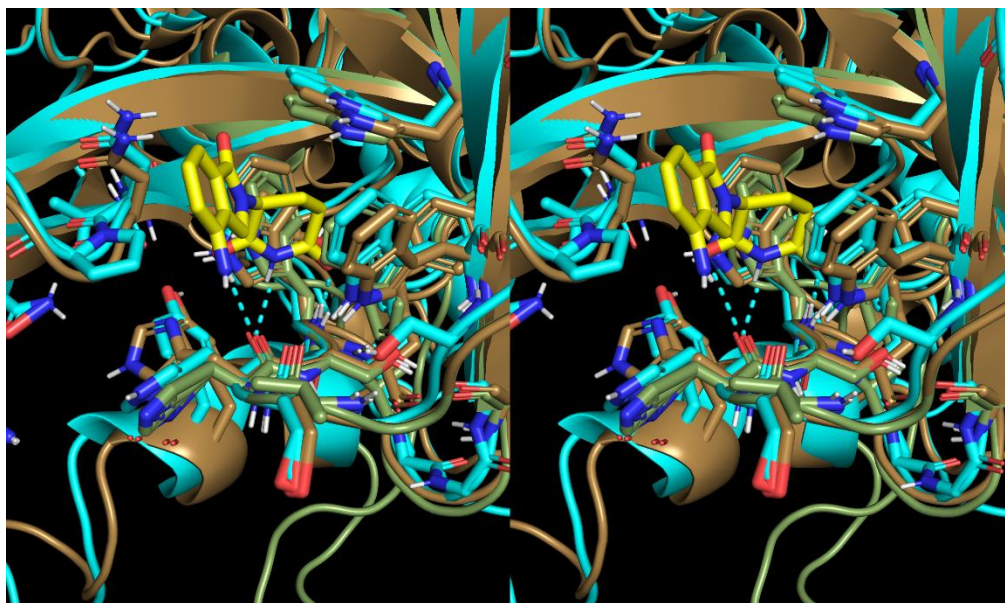


Figure 25. Overlay of the lenalidomide-bound structure C (yellow sticks and cyan cartoon, respectively) on structures A<sub>closed</sub> (tan cartoon) and H (olive green cartoon), which serve as putative models for the empty fully formed pocket L. The H-bonds between lenalidomide and cereblon are shown as cyan dashed lines. Similar structural properties are apparent between the mouse and human orthologs, including high sequence identity among the residues lining the pocket.

Our overall findings are consistent with the following mechanism by which binding between cereblon and its cognate substrates is switched on and off in the presence and absence of bound degraders (Figure 26):

- 1) The cereblon population is tipped toward the substrate binding-incompetent form in the degrader-unbound state, corresponding to the fully collapsed and partially formed states of the degrader pocket, in which the interfacial H-bond groups needed for low cost desolvation of the interface are misaligned (promoting a non-physiologically slow  $k_1$ ) (Figures 27A and B).
- 2) The cereblon population is tipped toward the substrate binding-competent form in the degrader-bound state, in which the aforementioned H-bond groups are fully aligned (promoting a physiologically meaningful  $k_1$ ) (Figures 27C and D).

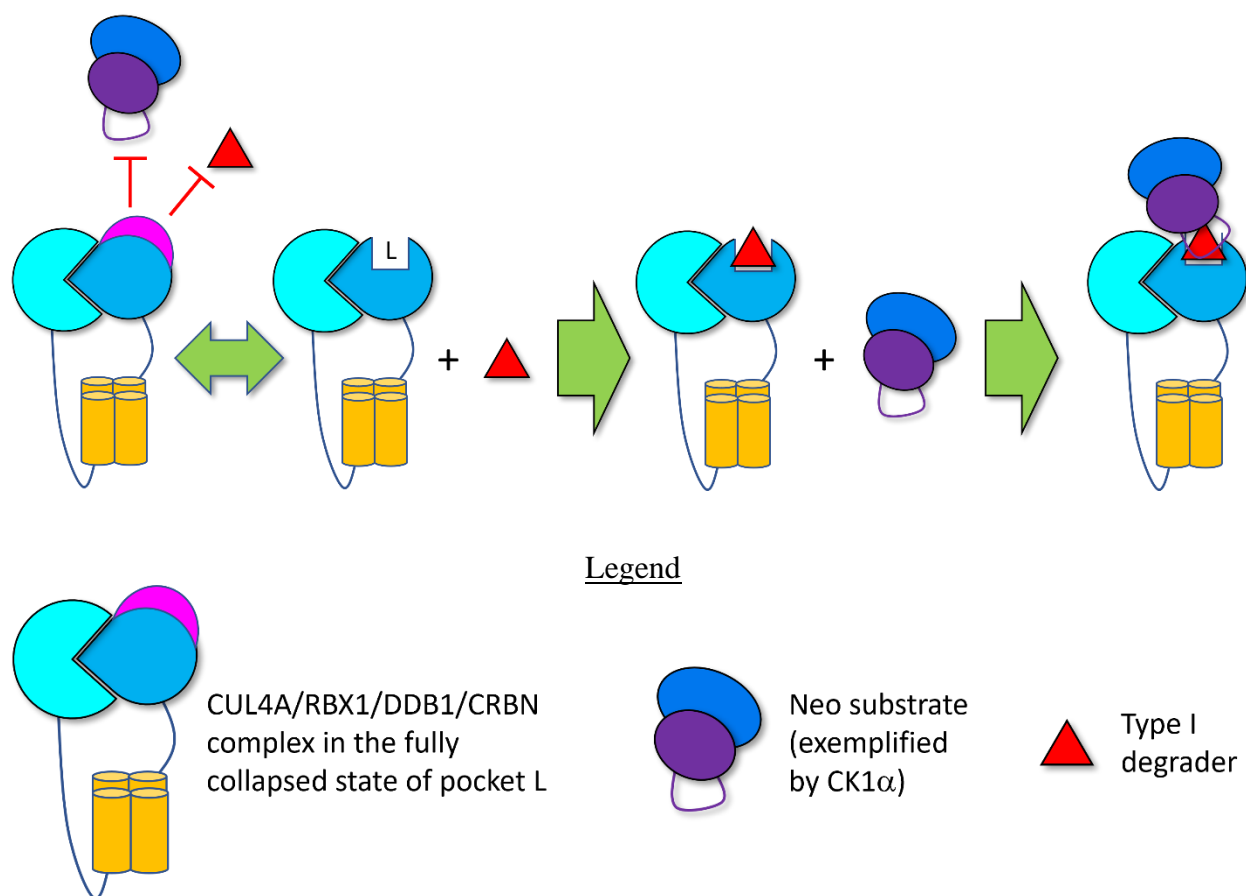
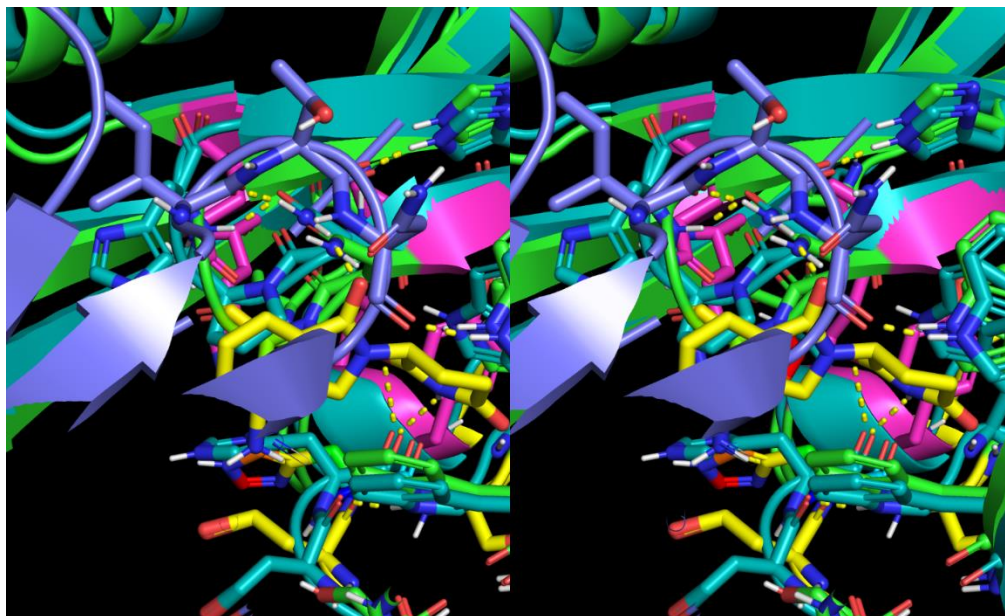
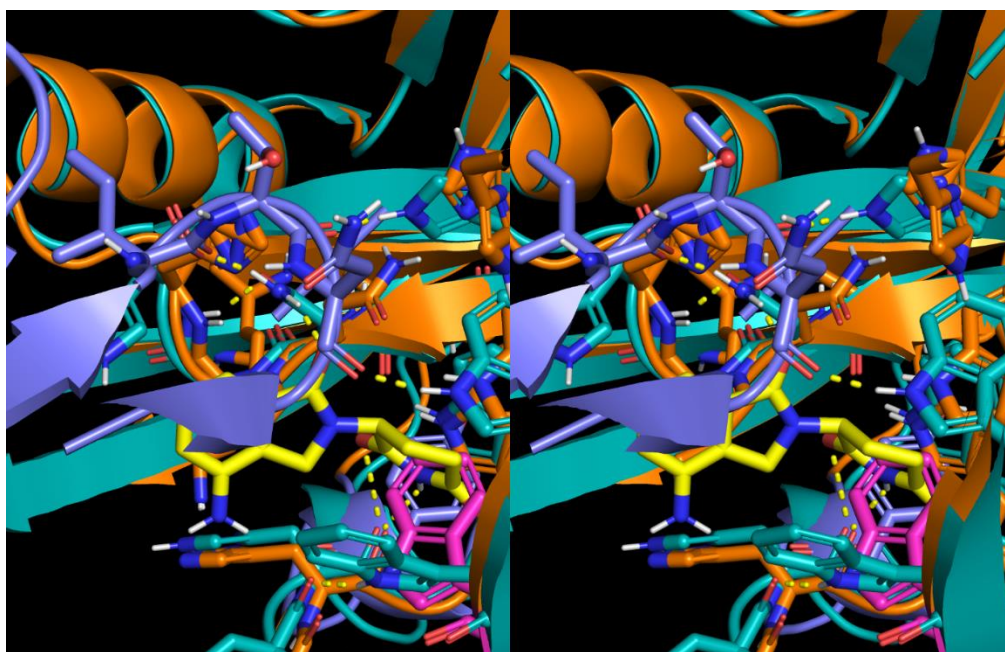


Figure 26. Cartoon depicting cereblon (medium blue tear dropped circle) in the CUL4A/RBX1/DDB1/CRBN complex (described in the Legend), which we postulate is distributed between the fully collapsed and partially formed states of pocket L (double-headed arrow) in the absence of a bound degrader (red triangle). We hypothesize that the substrate binding surface of cereblon (magenta region) is misaligned for substrate binding in the collapsed degrader-unbound state of pocket L and aligned in the degrader-bound state. We further postulate that the major fraction of cereblon exists in the fully collapsed, substrate binding incompetent state of pocket L, which is shifted to the substrate binding competent form upon degrader binding.

A

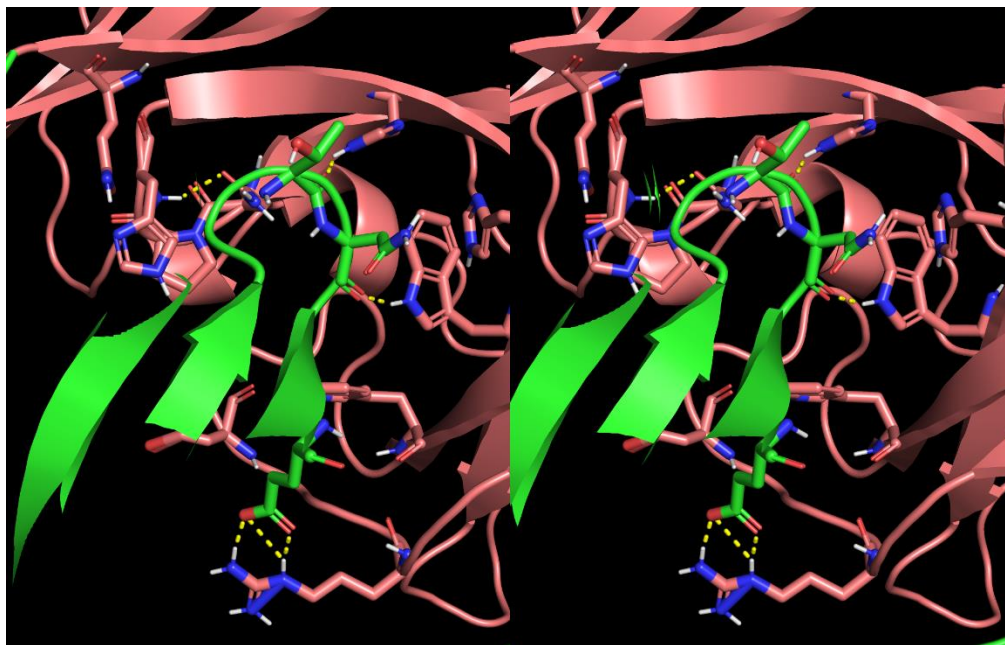


B





C



D

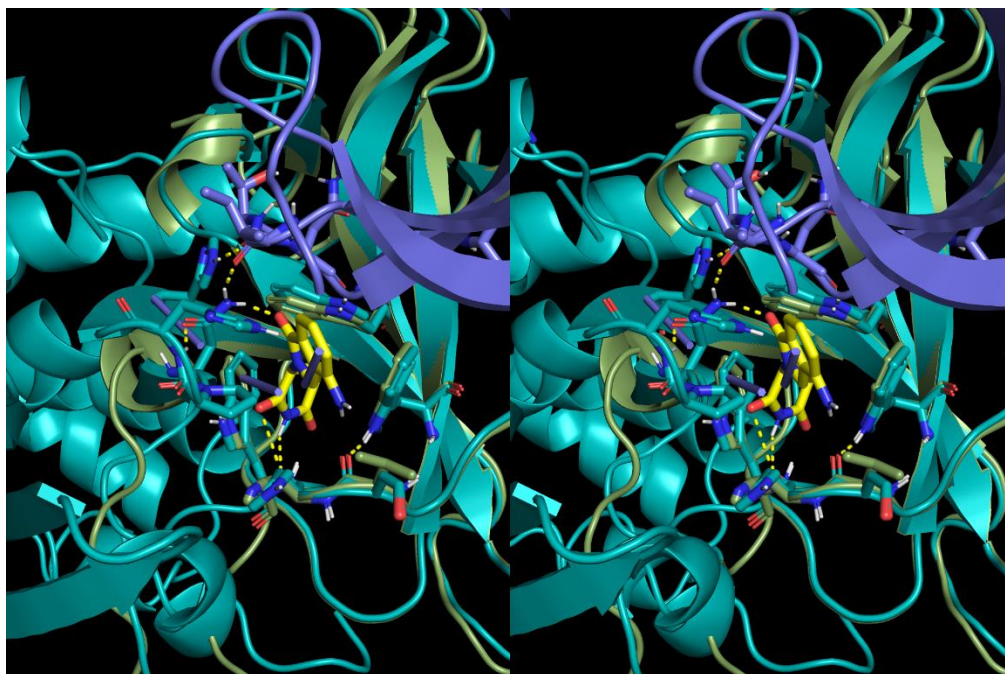


Figure 27. (A) Zoomed-in stereo view of the misaligned interfacial H-bond groups in the degrader-unbound structure A\_closed (green) relative to the aligned groups in the degrader-bound structure F (cyan). (B) Zoomed-in stereo view of the misaligned interfacial H-bond groups in A\_closed' (orange) relative to the aligned groups in structure F (cyan). (C) Zoomed-in stereo view of the aligned interfacial H-bond groups in the degrader-unbound structure B. Pre-existing H-bonds in the starting structure (derived from 5FQD) are retained by AMBER, despite the absence of bound lenalidomide, which further supports our claim that the solute H-bond energy contribution is overweighted in the force-field. (D) Zoomed-in stereo view of the interfacial H-bond groups in the degrader-unbound structure H (olive green) and degrader-bound structure C (cyan). The backbone and aromatic side chains lining the degrader pocket of this structure are well-aligned with those in structure C (cyan) but not the substrate H-bond groups.

We hypothesize that the open/rotated CULT domain of cereblon (structures A\_open and G) plays a key functional role in substrate ubiquitylation, as follows (Figure 28):

- 1) Degrader association triggers the release of the N-terminal peptide from the CULT domain that acts as a putative “latch” needed to stabilize the closed form observed in 5FQD. This is followed by dissociation of the CULT and N-terminal domains, thereby forming the open state observed in 6H0F and addressed below (noting that 5FQD may be stabilized in the closed state by crystal contacts). We made no attempt to simulate this process.
- 2) Substrate association ensues (noting that substrate binding prior to CULT rotation cannot be ruled out, and the two possibilities are non-mutually exclusive).
- 3) Ubiquitylation of the growing poly-Ub chain is facilitated by dynamic rotation of the ternary CULT-degrader-substrate complex.

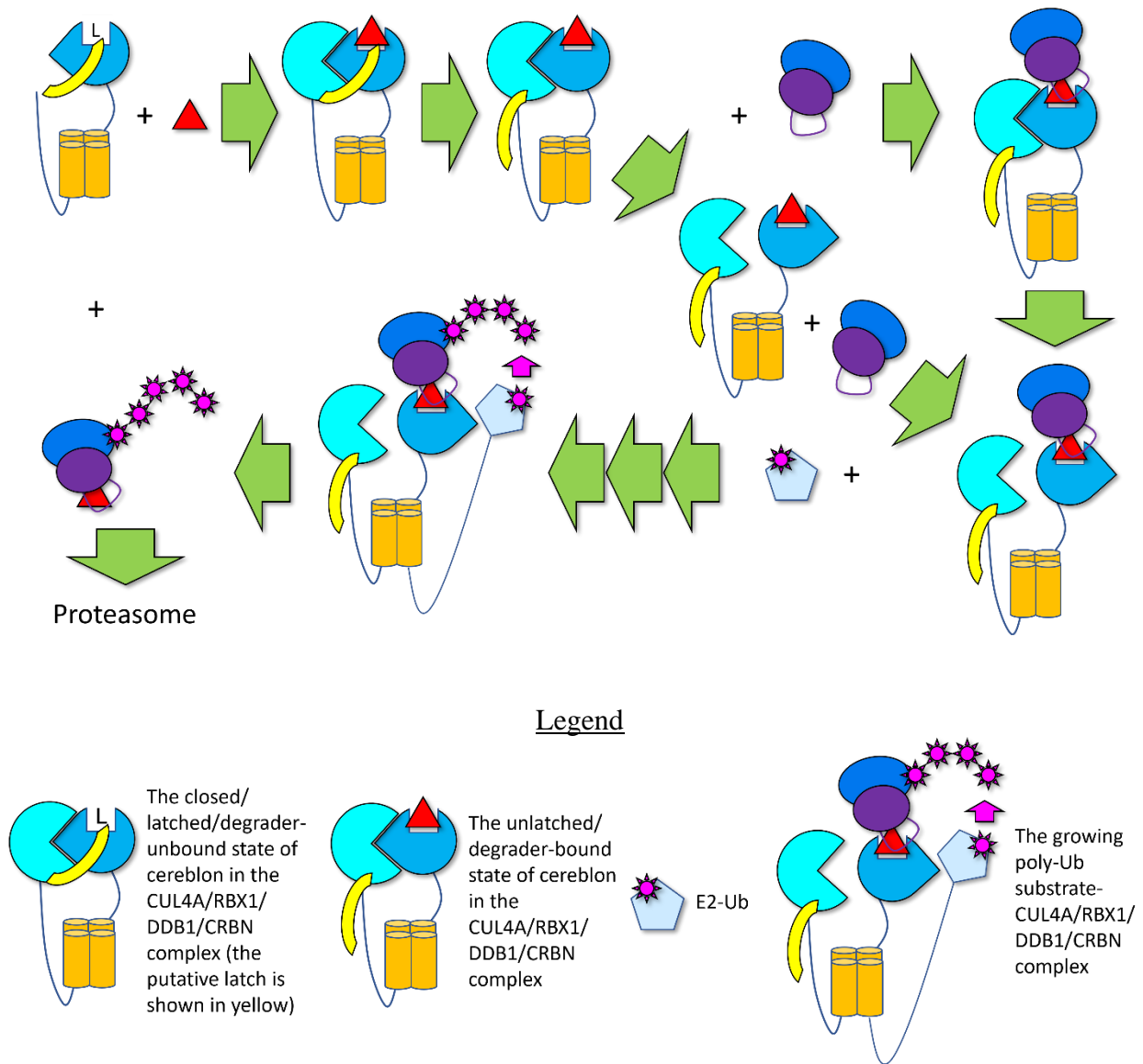
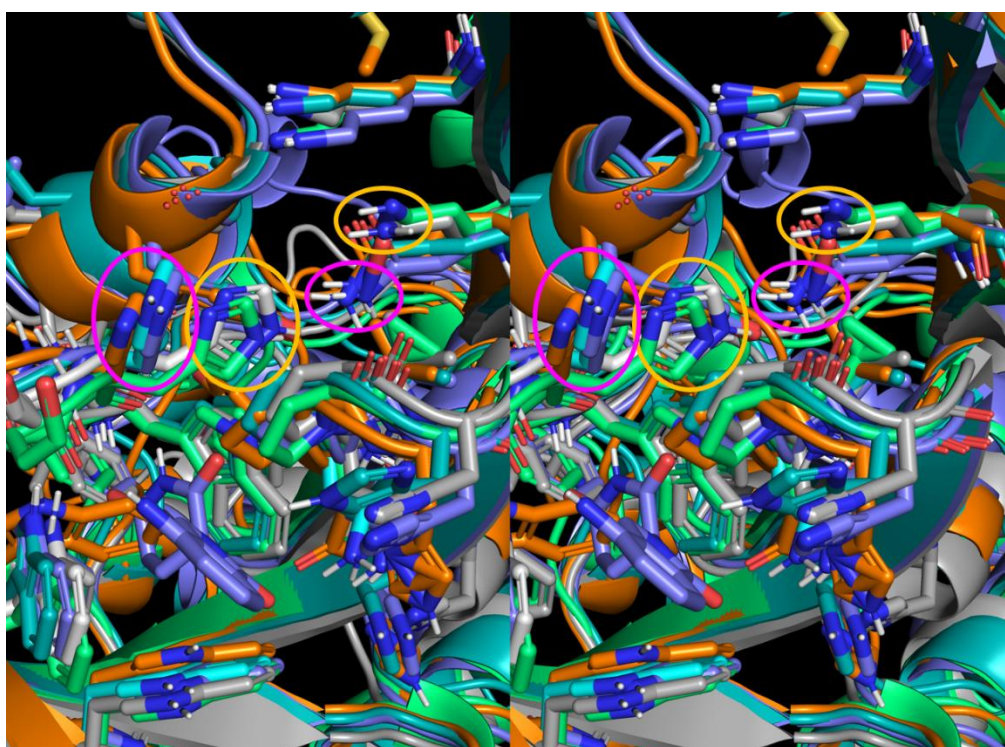


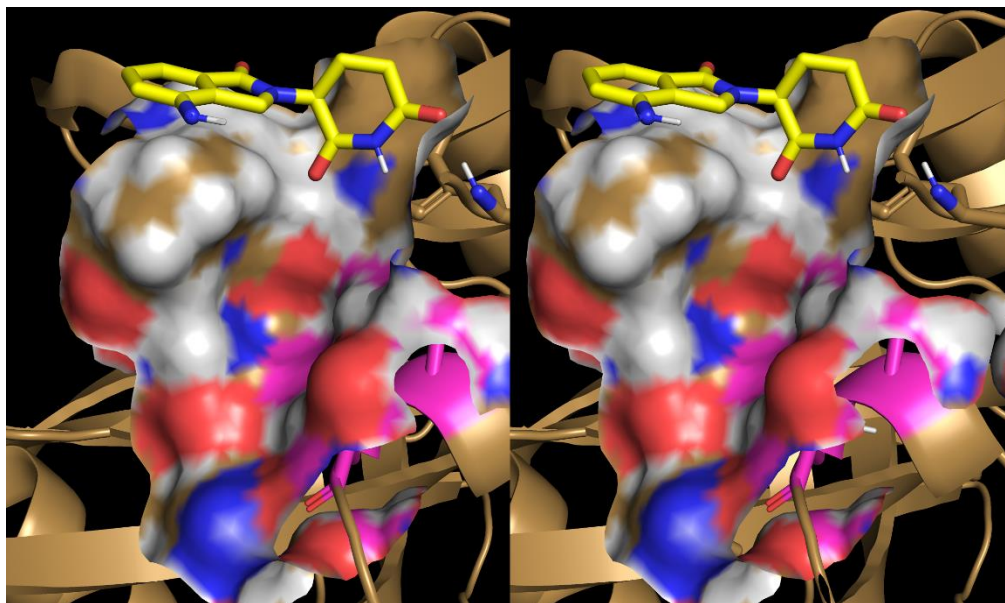
Figure 28. Cartoon depicting the putative CUL4A/RBX1/DDB1/CRBN enzyme cycle (described in the Legend), in which the degrader (red triangle) binds to the closed form of cereblon (the cyan and blue shapes and orange helical bundle), resulting in the release of the N-terminal peptide “latch” (yellow shape), followed by dissociation of the N-terminal and CUL4A domains. Substrates may associate to the closed and/or open forms of cereblon, which are then poly-ubiquitylated (magenta shapes) by the E2 subunit (light blue pentagon), followed by release of the poly-ubiquitylated product.

Pocket T exists solely in the time-averaged structures C, F, and A\_closed (DDB1 omitted), as well as A\_closed' (DDB1 included), but is occupied by His332 in structures D, C' (DDB1 included) and C'' (N-terminal peptide and DDB1 omitted) (Figure 29). We postulate that the closed  $\rightarrow$  open state transition (captured in 6H0F and structure A\_open) depends on degrader-induced release of the N-terminal peptide via desolvation of pocket T by His332.

A



B



C

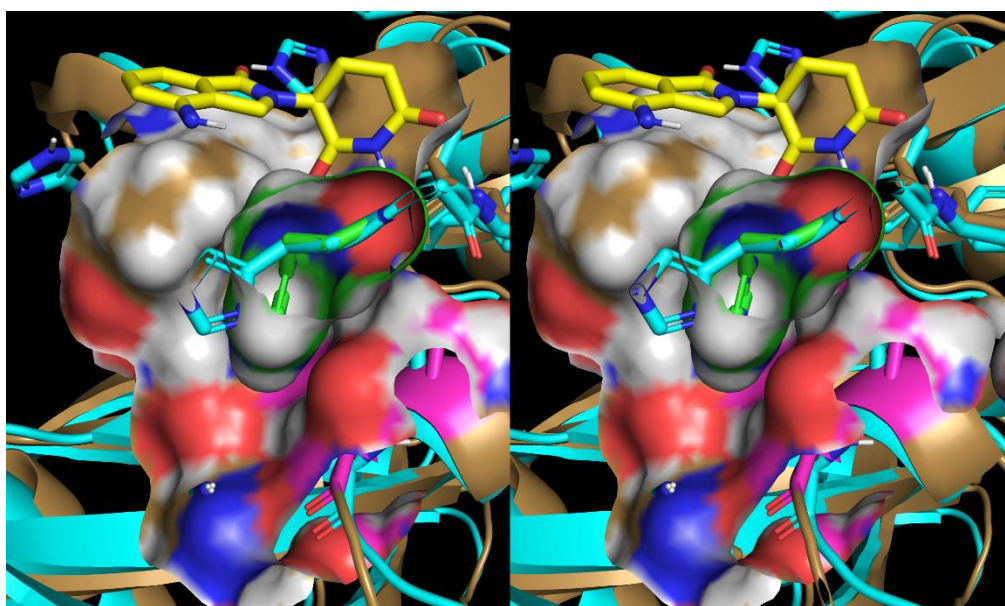


Figure 29. (A) Stereo view of His332, which transitions from the interior of pocket T in structures C' (mint green cartoon) and C'' (white cartoon) to the exterior in structures C (purple cartoon) and F (cyan cartoon). Occluding versus non-occluding His332 side chain positions are circled in yellow and magenta, respectively (where the pocket overlaps with the yellow circled His side chain). The His332 side chain position is correlated with that of Gln54 located on the N-terminal domain (the position correlated with the His332-occupied form is circled in yellow, and that with His332 unoccupied form is circled in magenta). (B) Stereo view of pocket T in the His332-occupied state in structure A\_closed'' with the His332 side chain omitted for clarity and lenalidomide from structure F shown for reference (yellow). Pocket T resides at the three-way interface between the N-terminal peptide (magenta cartoon and surface), N-terminal domain, and CULT domain. (C) Same as B, except in the His332-occupied form of pocket T (green sticks and surface). Lenalidomide (yellow sticks) and His332 (cyan sticks) in pocket L of structure F are shown for reference. We hypothesize that the interface is disrupted when pocket T is vacated by His332 in response to degrader binding, which is driven by desolvation of the pocket by the southern glutarimide O of the degrader.

### **The solvation-powered intra- and intermolecular dynamics of cereblon**

Next, we studied the solvation fields powering the intra- and intermolecular dynamics of cereblon, and specifically:

- 1) The solvation free energy contributions to the substrate and degrader association and dissociation barriers underlying  $k_1$ ,  $k_{-1}$  and  $k_{on}$ ,  $k_{off}$ , respectively, in the context of the time-averaged structures and solvation fields of cereblon complexed with DDB1:
  - a) Modeled unbound cereblon (structures A\_closed, A\_closed', and A\_open, and H) and CK1 $\alpha$  (structure E).

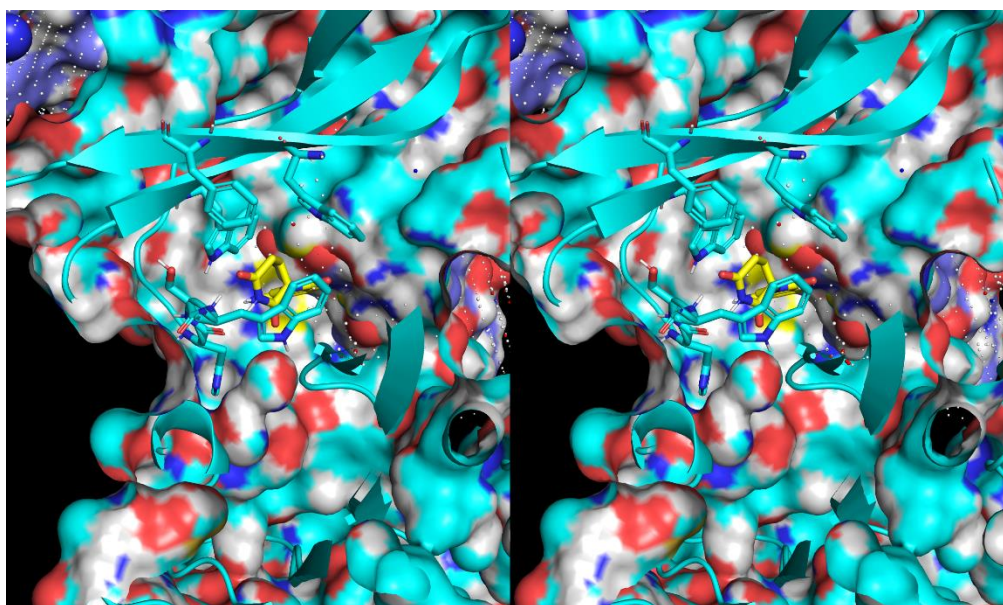
- b) The hypothetical binary cereblon-CK1 $\alpha$  (structure B).
  - c) Modeled binary cereblon-lenalidomide (structures C and C’).
  - d) The ternary cereblon-lenalidomide-CK1 $\alpha$  (structure F) complexes (5FQ1).
- 2) Complementarity between the solvation field of the degrader pocket (structures A\_closed and H) and those of thalidomide, lenalidomide, and uridine.
  - 3) The putative mechanism by which dissociation of the N-terminal peptide (observed in 6H0F) is powered by the degrader-induced release of solvation free energy stored within pocket T.
  - 4) The internal/buried solvation within cereblon and its role in facilitating and powering the intramolecular state transitions of the protein (explained in general terms in the Introduction).

According to Biodynamics theory, association barriers, in general, are contributed principally by the desolvation costs of the associating partners, whereas dissociation barriers are contributed principally by the resolution costs of the dissociating partners (as explained in previous sections of this work and in our previous works<sup>2-4</sup>). Fast association depends on high binding partner concentrations and/or fast  $k_1$ , which in turn, depends on optimal H-bond replacements of H-bond enriched interfacial solvating water by inter-partner H-bonds. Slow dissociation depends on significant desolvation of H-bond depleted solvation during association (i.e., depending on the existence of such solvation on either partner, and the extent to which such solvation is sterically overlapped in the bound state). We assume that the cognate substrates of each specific cereblon-degrader complex:

- 1) Are sterically matched and tightly packed so as to access  $k_{-1}$ -slowing H-bond depleted solvation residing at the protein surface, deep within the solvation field.
- 2) Contain sufficient amounts of H-bond depleted solvation that is mutually expelled from the binding interface (conveying a physiologically meaningful  $t_{1/2}$ ).
- 3) Project spatially matched polar groups into the interface.

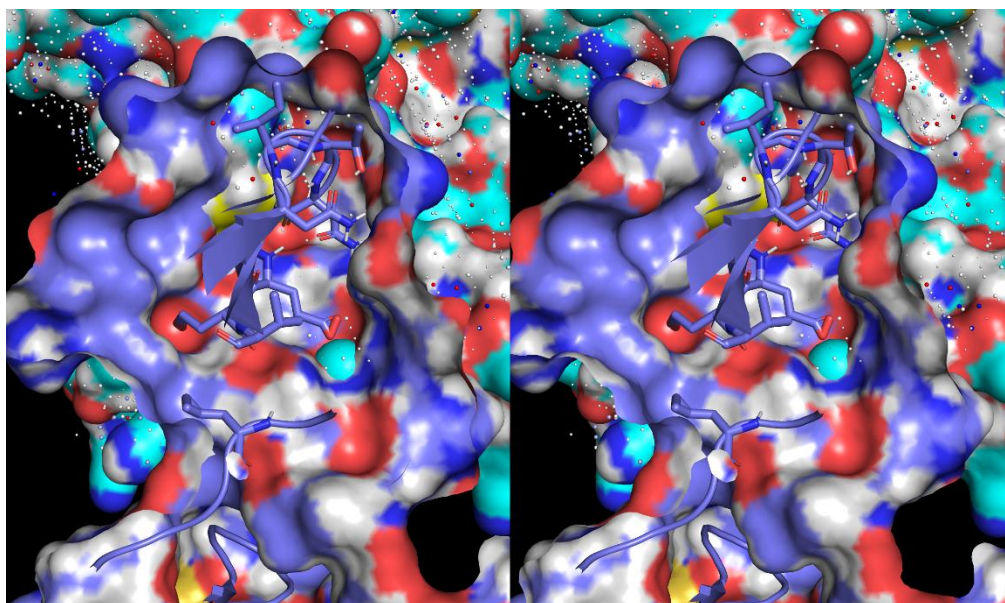
We calculated the solvation fields within the binding interface of CK1 $\alpha$  and cereblon separately, along with those of the hypothetical binary cereblon-CK1 $\alpha$  (structure B) and ternary (structure F) complexes using WATMD. Replacement of water in the ultra-low occupancy voxels (ULOVs) of CK1 $\alpha$  by non-polar groups contributed by the cereblon-lenalidomide complex and of cereblon by non-polar groups contributed by CK1 $\alpha$  is shown in Figures 30A and B, respectively. H-bonds formed among the aligned partners replace those of their solvating water, thereby lowering the desolvation cost of the interface (Figure 30C).

A





B



C

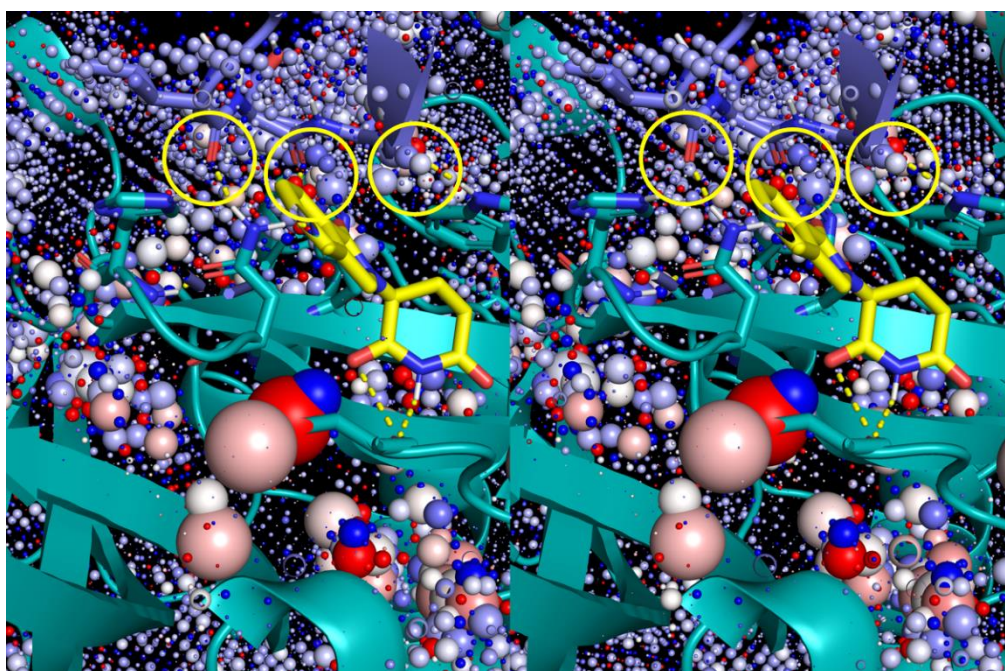
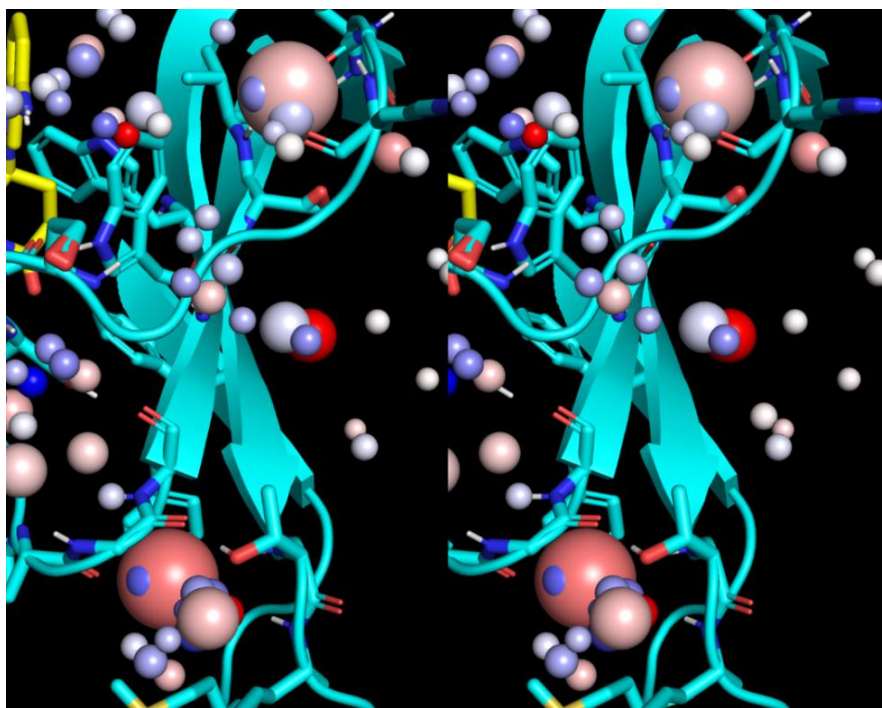


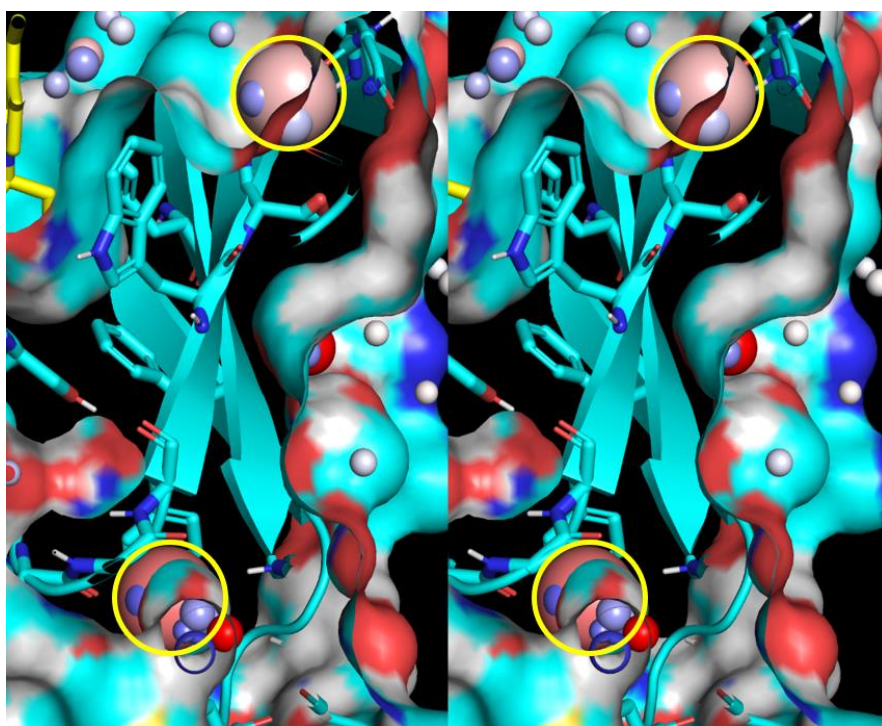
Figure 30. (A) Stereo view of the ultra-low occupancy voxels (ULOVs: small white spheres) on the surface of CK1 $\alpha$  (blue cartoon) corresponding to H-bond depleted solvation that is replaced by non-polar groups during cereblon-lenalidomide (cyan cartoon) association (viewed from the interior of the cereblon-lenalidomide complex looking toward the external surface of CK1 $\alpha$ ). (B) Same as A, except the ULOVs on the surface of the cereblon-lenalidomide complex corresponding to H-bond depleted solvation replaced by non-polar CK1 $\alpha$  groups (viewed from the interior of CK1 $\alpha$ , looking toward the external surface of cereblon-lenalidomide). (C) Stereo view of high occupancy voxels (HOVs: blue-white-red color spectrum coded according to donor-no preference-acceptor preference) corresponding to H-bond enriched solvation of cereblon Asn305, His311, and Trp354 desolvated by CK1 $\alpha$  (circled in yellow) (noting that the intra-cereblon and cereblon-lenalidomide H-bonds were necessarily desolvated prior to CK1 $\alpha$  binding). The small voxel sizes corresponding to this water reflect relatively mild H-bond enrichment (which is potentially underestimated by WATMD), suggesting that the interfacial desolvation cost is relatively low and  $k_1$  is therefore relatively fast. The H-bonds of this water are replaced by the cereblon-CK1 $\alpha$  H-bonds shown in Figure 22A.

It then follows that the substrate binding-competent state of cereblon is stabilized by degrader-induced desolvation of pocket L, and further stabilized by two remote H-bond enriched solvation positions (corresponding to the HOVs in Figures 31A and B), one of which persists in the CK1 $\alpha$ -bound state, that likely further stabilizes that state (Figure 31C).

A



B



C

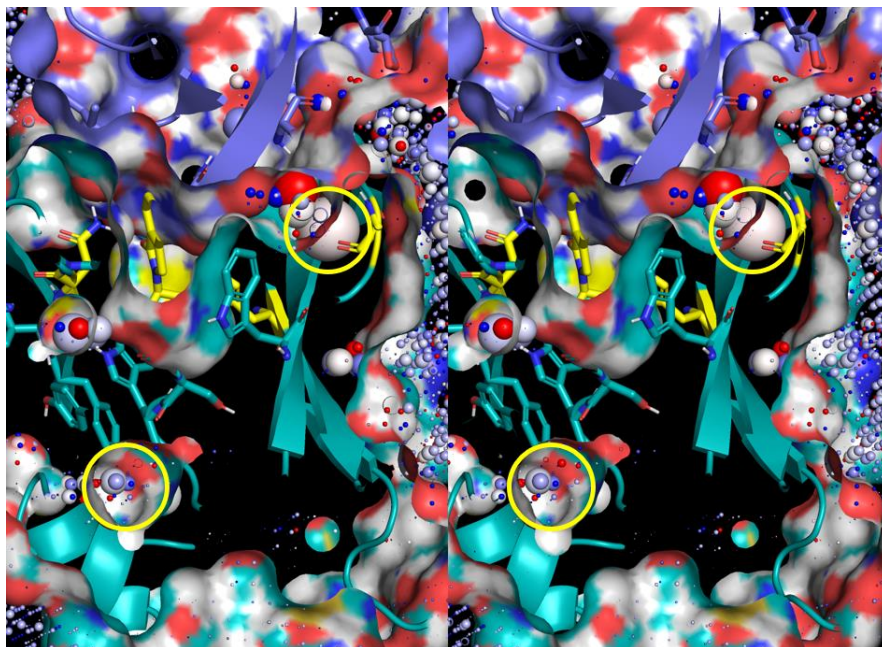


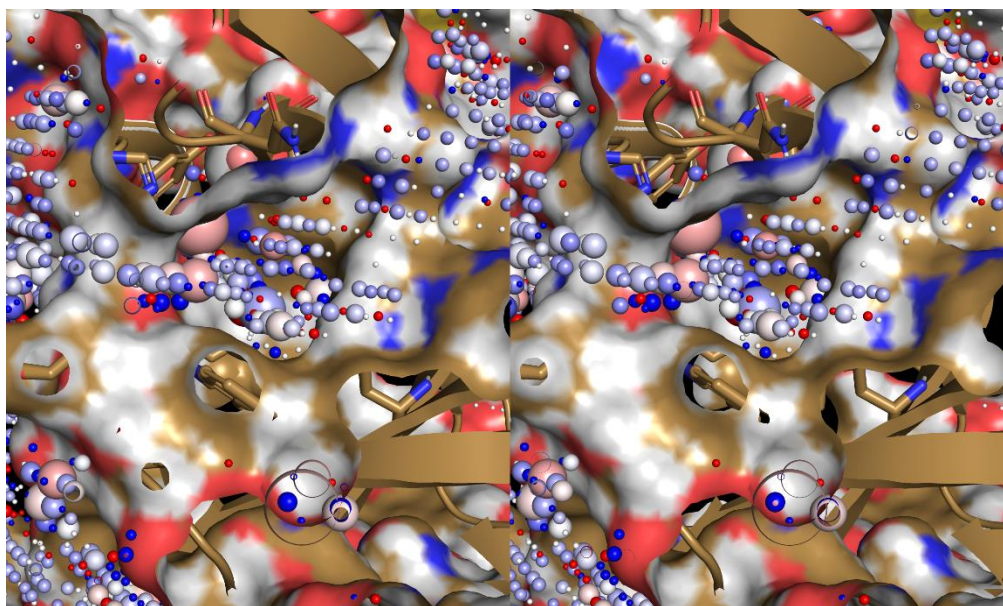
Figure 31. (A) Zoomed-in stereo view of two HOVs (circled in yellow) present in structure C, but not structure A. This water likely helps to stabilize the degrader-bound state of the protein. (B) Same as A, except showing the protein surface. The two HOVs reside at the protein surface, and therefore represent H-bond enriched, rather than buried/trapped water. (C) Same as A, except for the ternary structure F. The upper voxel resides within the CK1 $\alpha$ -cereblon interface (purple and cyan surfaces, respectively), whereas the lower one is entirely absent.

Next, we analyzed the WATMD results for the empty fully formed state of pocket L (structures A<sub>closed</sub> and H). The ability of pocket L to switch between the substrate binding-competent and incompetent forms of cereblon depends on instability of the empty pocket and substrate binding surface, corresponding to disrupted shape and interfacial H-bond group complementarity vis-à-vis cognate substrate(s). We postulate that the solvation of pocket L (Figure 32A) resides within a Goldilocks zone of unfavorability (both enthalpically and entropically), such that the population

is distributed dynamically between the collapsed/auto-desolvated and fully formed/solvated states (states 1 and 2, respectively), as follows:

- 1) State 1 (the degrader and substrate binding-incompetent state): desolvation of the pocket via inward rotation of Trp340 (which is favorable) is coupled directly to the loss of H-bond enriched solvation in the peri-pocket region (which is unfavorable) (Figure 32B).
- 2) State 2 (the degrader and substrate binding-competent state): resolution of the pocket via outward rotation of Trp340 (which is unfavorable) is driven by the accompanying gain of H-bond enriched solvation in the peri-pocket region (which is favorable). As such, the substrate-binding competent/substrate-bound populations build on the coattails of degrader association.

A



B

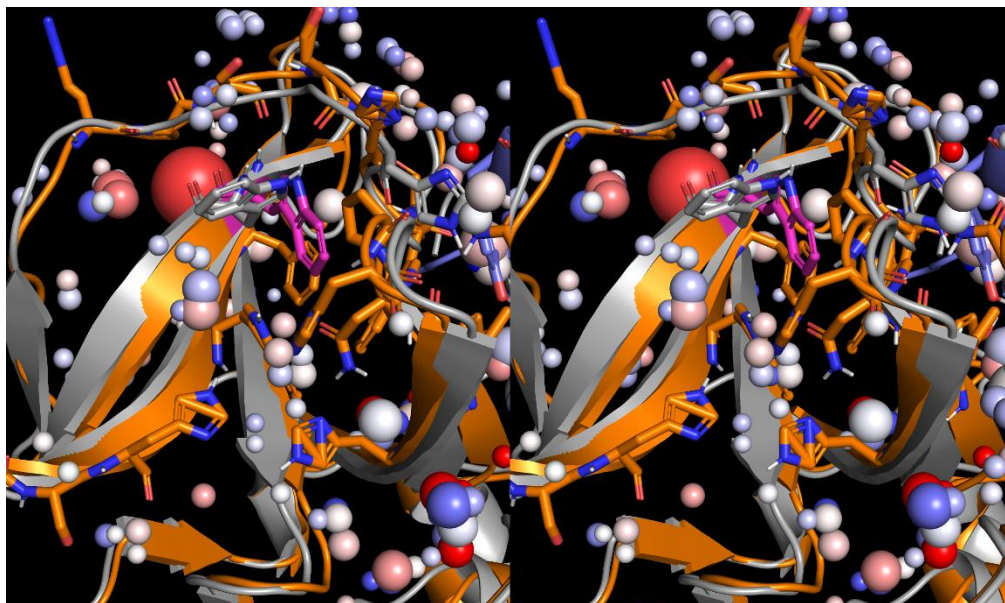
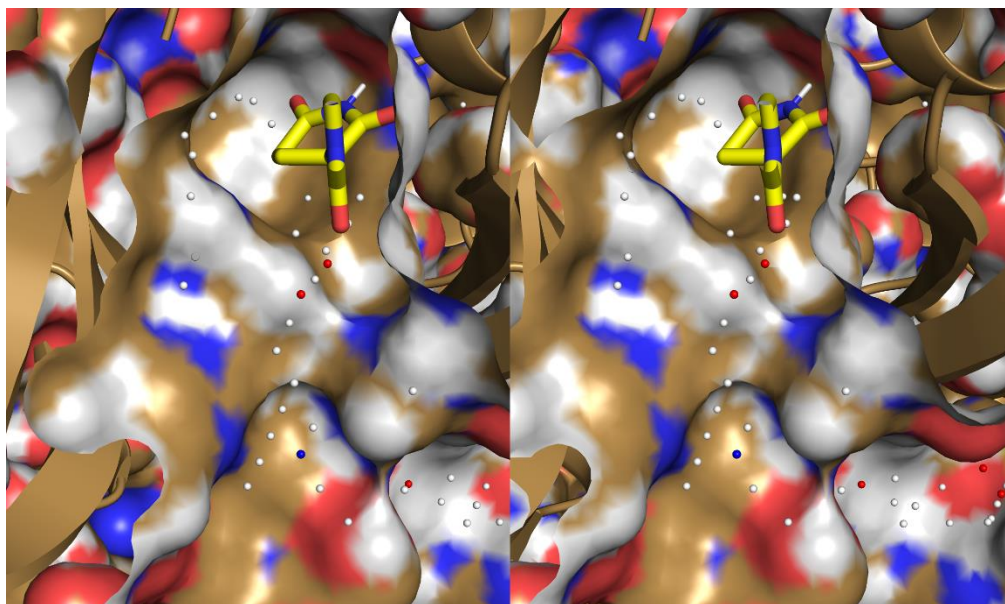


Figure 32. (A) Stereo view of the ULOVs and HOVs within the fully formed pocket L of structure A<sub>closed</sub>'. This solvation exhibits a range of H-bond depletion (small to large spheres) relative to bulk solvent (see text). (B) Favorable auto-desolvation of pocket L in structure A<sub>closed</sub>' (in which pocket L is fully collapsed) is achieved via transient rotation of Trp340 into the pocket (white stick from structure C versus magenta stick from structure A<sub>closed</sub>'). This rearrangement is putatively counter-balanced by the unfavorable loss of H-bond enriched solvation from the voxel positioned directly behind the Trp340 backbone (red sphere), resulting in conformational “frustration”.

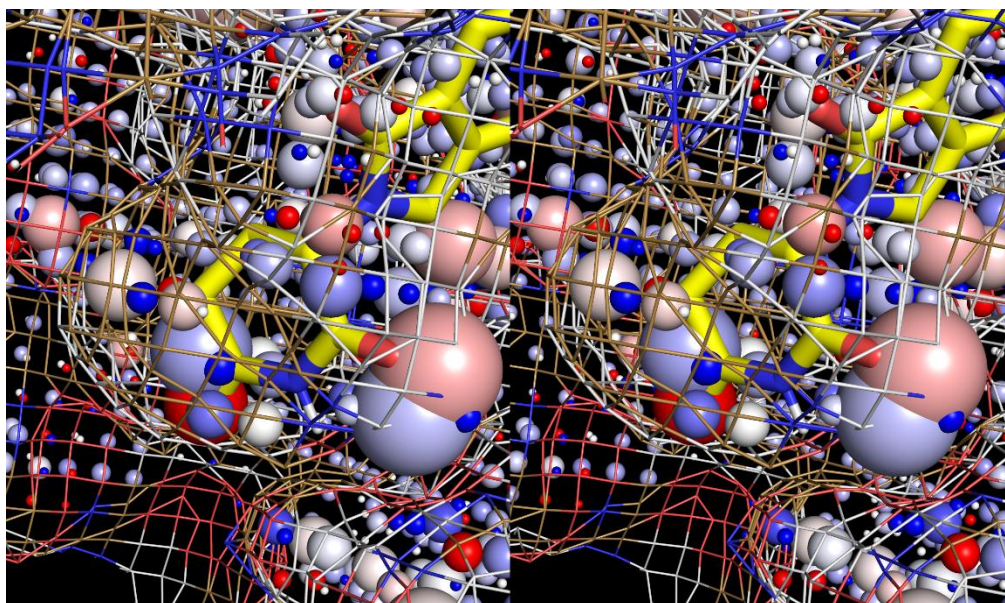
A significant number of ULOVs reside within the fully formed (Figures 33A and 34A) and partially collapsed states of pocket L (structures A<sub>closed</sub>' and A<sub>closed</sub>, respectively), which is consistent with the highly aromatic composition of the pocket. The higher occupancy voxels within the pocket (Figures 33B and 34B) correspond to solvating water that is H-bonded to a small number of protein H-bond groups (Figure 33C) (well below the enrichment threshold, recalling

that bulk solvent contains  $\sim 3.5$  moderately strong H-bonds per water at zero entropic cost). The voxel configuration is consistent with fast  $k_{\text{on}}$  and slow  $k_{\text{off}}$  of lenalidomide under native conditions (corresponding to low desolvation and high resolvation costs, respectively), the correct measurement of which may be hampered by the hypothesized non-continuous availability of the pocket. Fewer HOVs are present in structure H, likely due to the shallower nature of the pocket (Figures 34B and C).

A



B



C

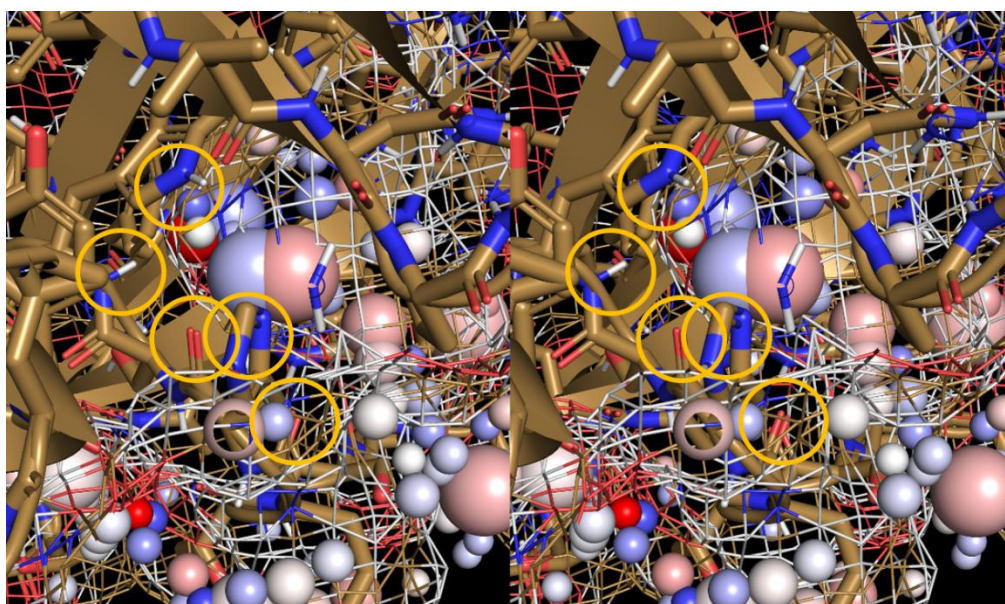
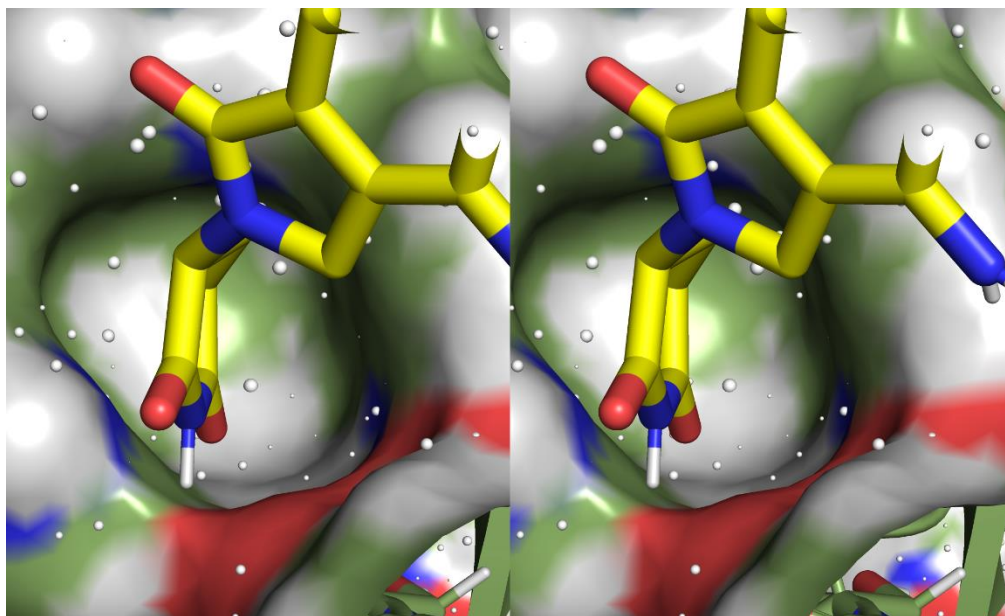


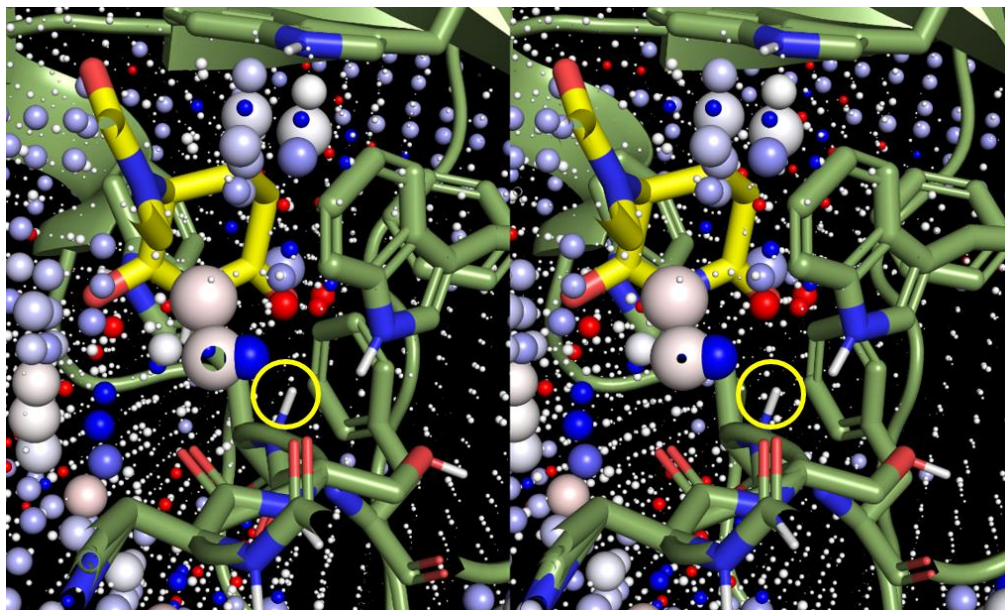


Figure 33. Stereo views of the voxels within the modeled fully formed pocket L in structure A<sub>closed</sub> (tan solid surface) (shown as spheres described in Materials and methods). (A) Zoomed-in view of the ULOVs in the pocket overlaid on lenalidomide from structure C (yellow sticks). (B) Same as A, except showing the higher occupancy voxels which are organized in a ring-like fashion that mirrors the glutarimide ring of the degrader (such that the glutarimidine oxygens overlap the two large voxel clusters, with the NH group residing between). Favorable desolvation of HOVs/H-bond enriched solvation is conveyed strictly by polar atoms, such as the glutarimide moiety of the ring, which is consistent with the lack of H-bond enriched solvation at these positions. (C) Same as B, except zoomed into the HOVs at the base of the pocket, showing the H-bond groups that partially stabilize the solvating water at these positions (circled in yellow).

A



B



C

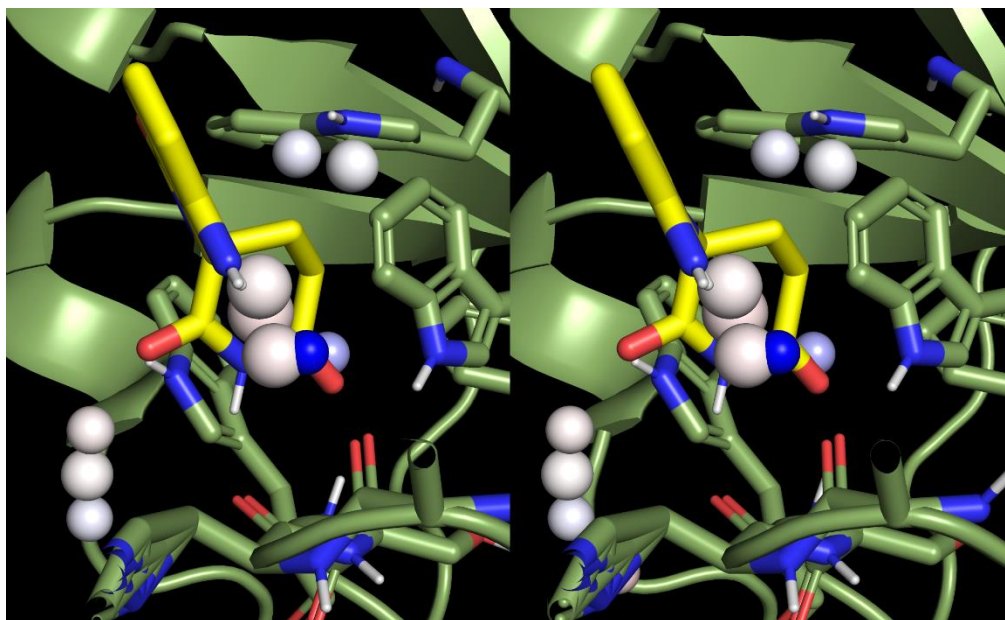
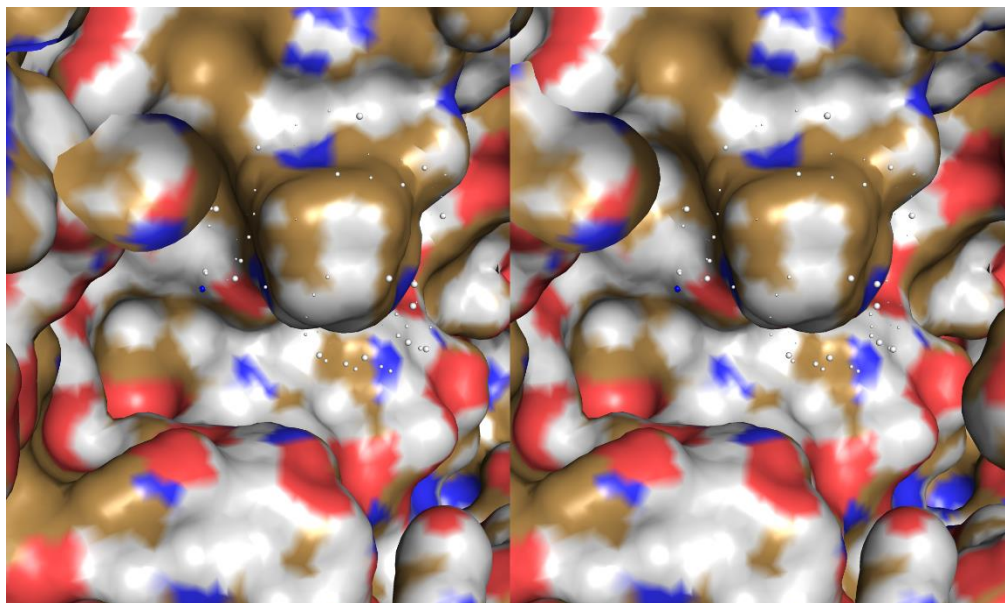


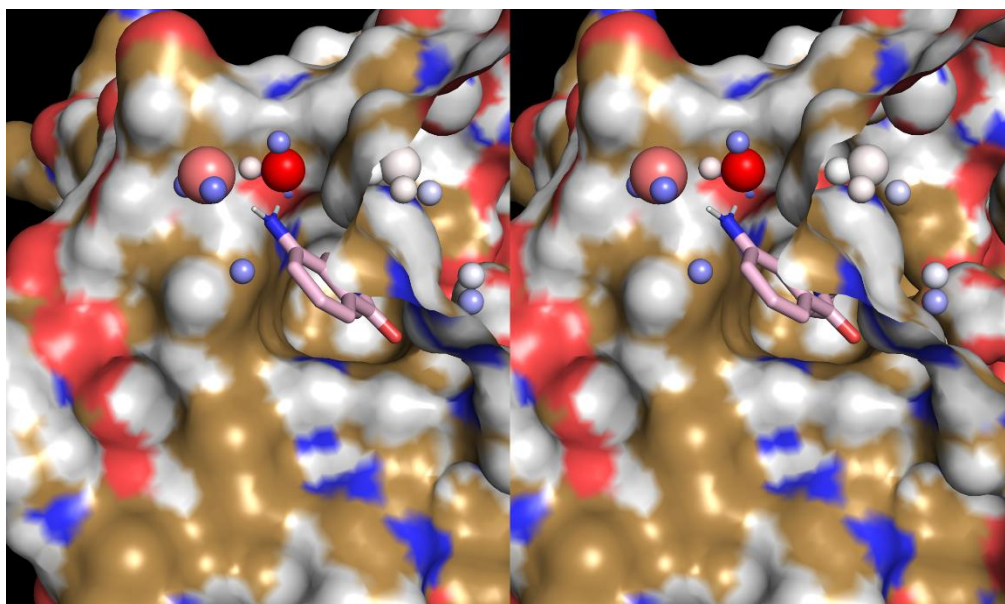
Figure 34. Zoomed-in stereo views of the voxels within the fully available pocket L of structure H. Lenalidomide (yellow) from structure C is shown for reference. (A) ULOVs corresponding to H-bond depleted solvation (white spheres) residing along the pocket surface. (B) Moderate occupancy voxels corresponding to slightly enriched solvation arising from H-bonding with the backbone NH of Trp64 (circled in yellow). The voxels are overlapped by one glutarimide O of lenalidomide, consistent with replacement of the occupying solvation by this group during association. (C) HOVs corresponding to H-bond enriched solvation arising from H-bonding with the backbone Os of Val61 and His62. The voxels are approximately overlapped by the exocyclic NH<sub>2</sub> of lenalidomide, consistent with replacement of the occupying solvation by this group during association.

Next, we analyzed the LMW WATMD results for lenalidomide (Figure 35), thalidomide (Figure 36), and uridine (based on a docked model) (Figure 37) and assessed the voxel complementarity with those of the fully formed pocket L in structure A<sub>closed</sub>” (the results are described in the figure captions).

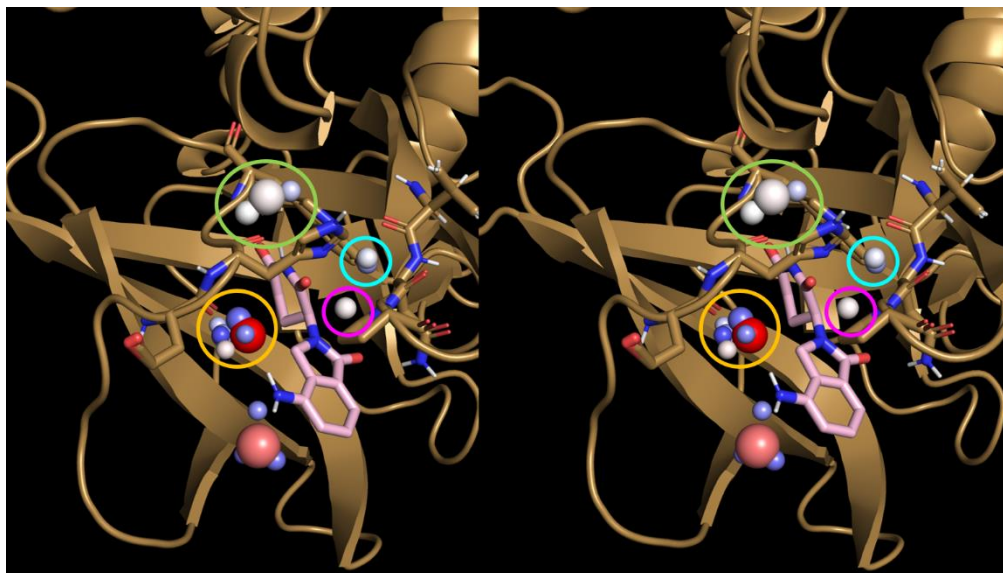
A



B



C



D

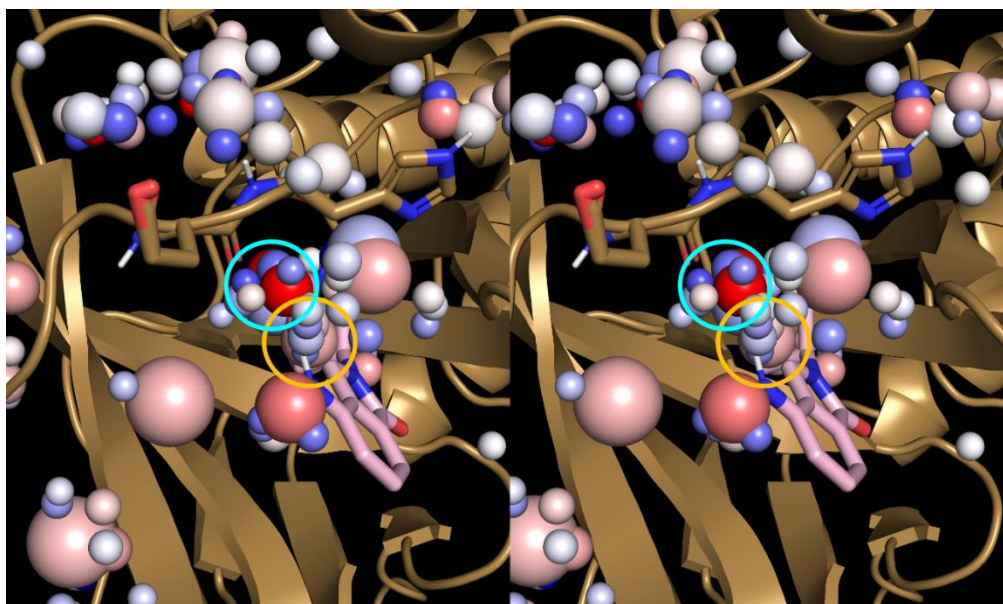
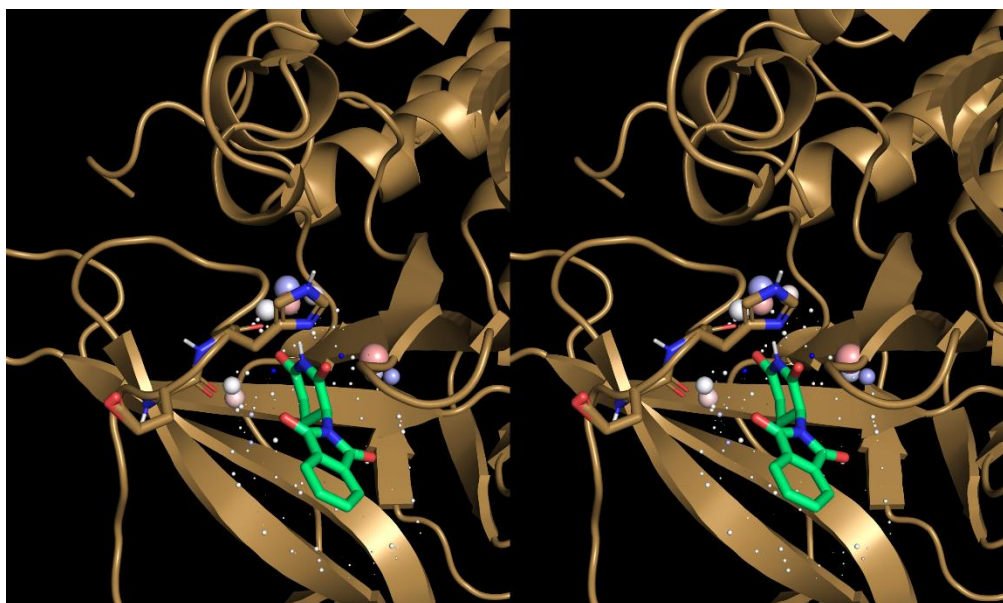


Figure 35. (A) Stereo view of the the LMW time-averaged lenalidomide structure (not visible) and ULOVs (white spheres) overlaid on structure A\_closed". The interior surface of structure A\_closed" (looking from the DDB1 interface toward pocket L) is visible in the center of the figure.

The solvation in these voxels is necessarily expelled by cereblon during lenalidomide association, contributing to slowed  $k_{\text{off}}$  of the degrader. (B) Stereo view of the LMW time-averaged lenalidomide structure (pink sticks) and HOVs (colored using a blue-white-red spectrum, as described in Materials and methods) overlaid on structure A<sub>closed</sub>". All but the large red voxel in the northeast corner are expelled during lenalidomide association, which putatively governs the degrader  $k_{\text{on}}$ . (C) Stereo view of the LMW time-averaged lenalidomide structure and HOVs thereof, showing desolvation of those positions by the backbone O of Glu331 (circled in yellow), side chain of Trp334 and backbone O of His332 (circled in light green), side chain of Pro306 (circled in magenta), and side chain of Trp334 (circled in cyan). (D) Same as C, except including the HOVs around the backbone O of Glu331 of structure A<sub>closed</sub>" (circled in yellow), together with those around the exocyclic amine of lenalidomide (circled in cyan). These positions are mutually desolvated by the protein and degrader during association.

A



B

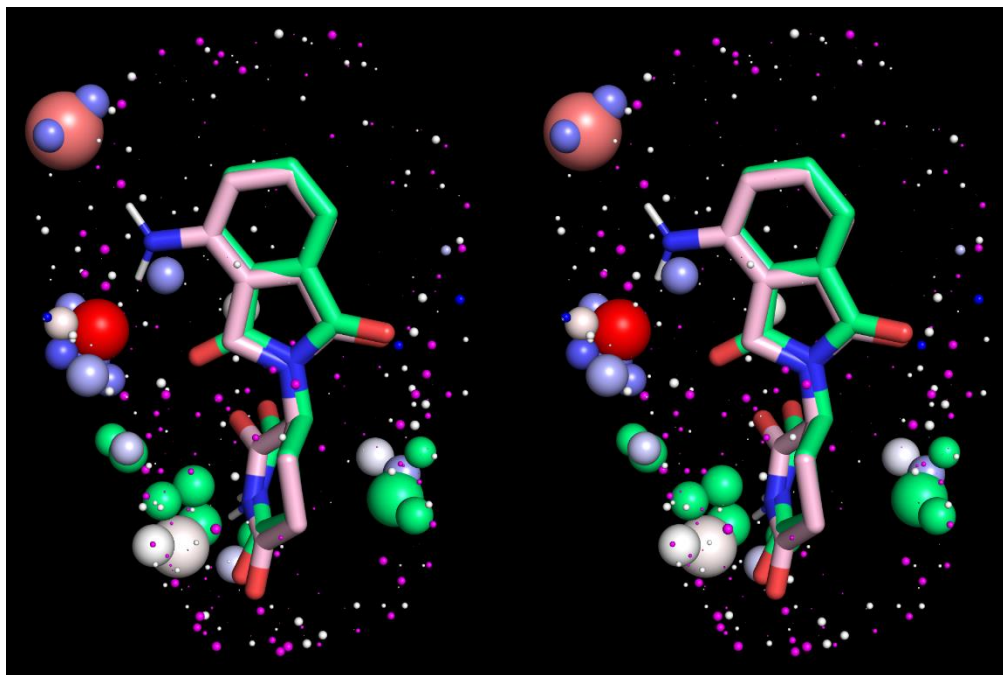
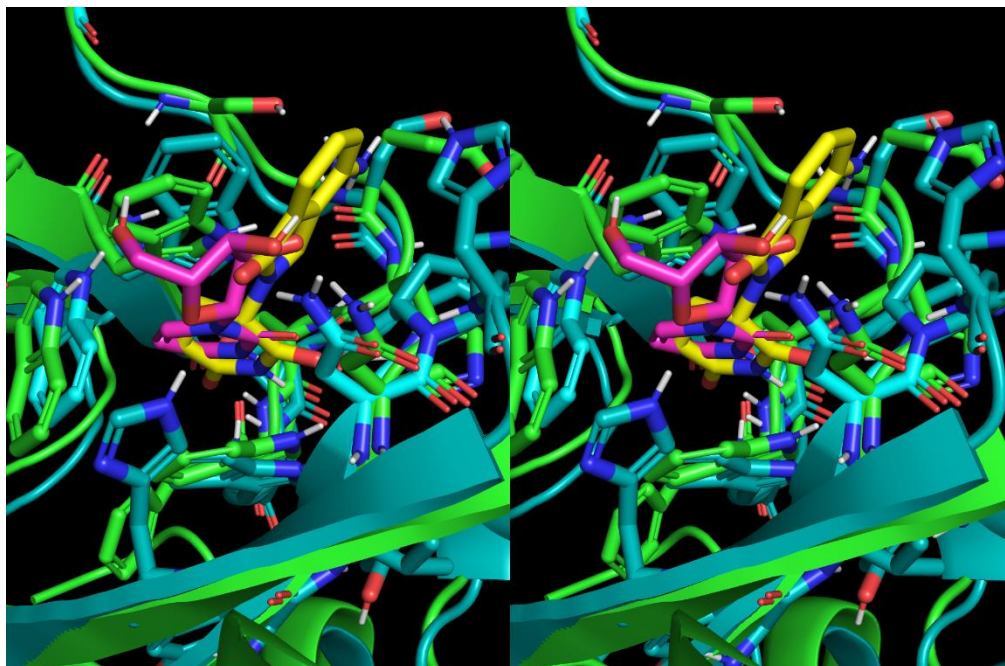


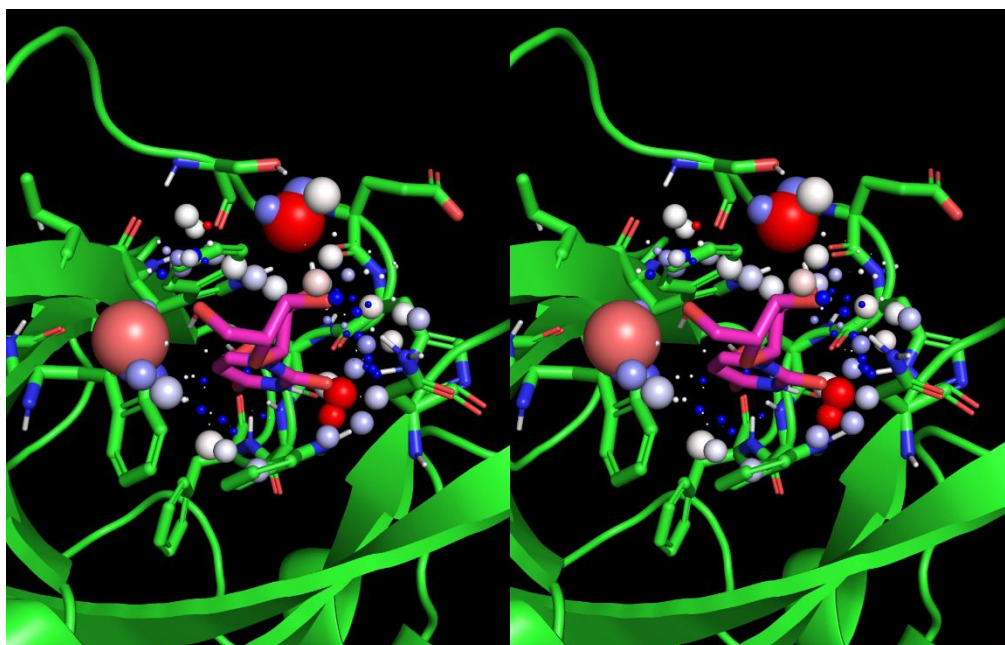
Figure 36. (A) Stereo view of the LMW time-averaged lenalidomide structure and the ULOVs and HOVs overlaid on structure A\_closed". (B) Stereo view of the time-averaged lenalidomide structure (pink sticks) and voxels (colored as in A) overlaid on the voxels (HOVs and ULOVs colored in mint green and magenta, respectively) and time-averaged thalidomide structure (mint green sticks). The voxels of the two degraders differ mainly around the exocyclic amine position.

We docked uridine into the pocket L of structure D as described in Materials and methods, followed by a 10 ns SD simulation (resulting in time-averaged structure K). The uracil moiety of the docked degrader is well-overlaid on the glutarimide ring of lenalidomide, with the ribose ring projecting outward toward the binding surface of the substrate (Figures 37A and B). CK1 $\alpha$ , and likely all  $\beta$ -hairpin containing substrates, appear to be highly incompatible with the binary cereblon-uridine complex (the lenalidomide-bound surface of cereblon is shown in Figure 37C for comparison).

A

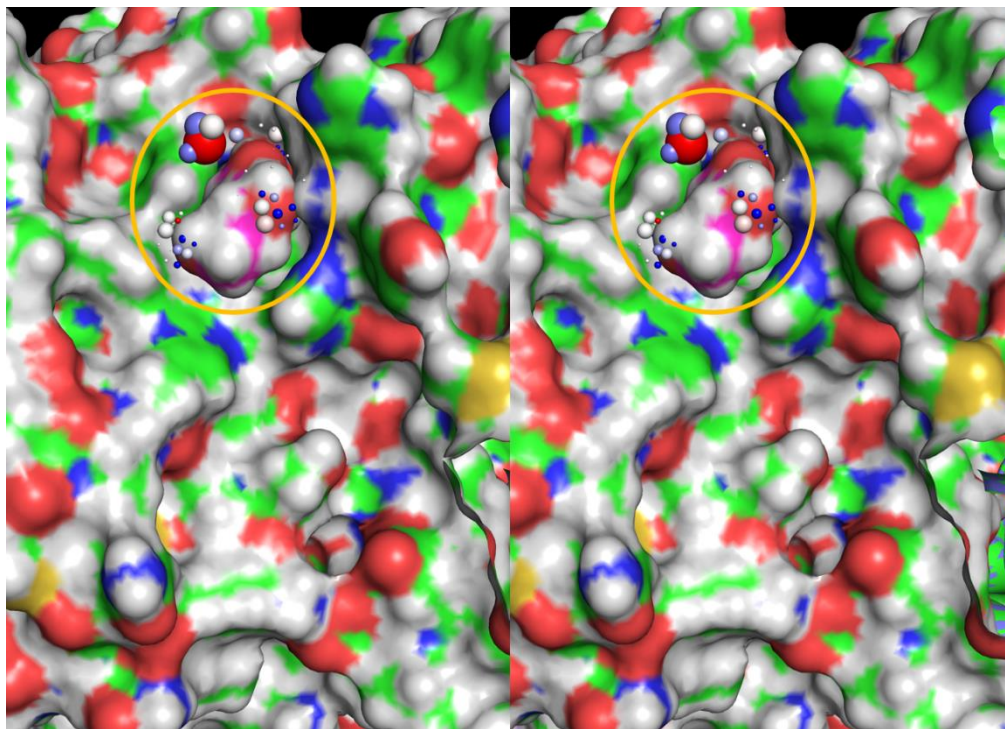


B





C



D

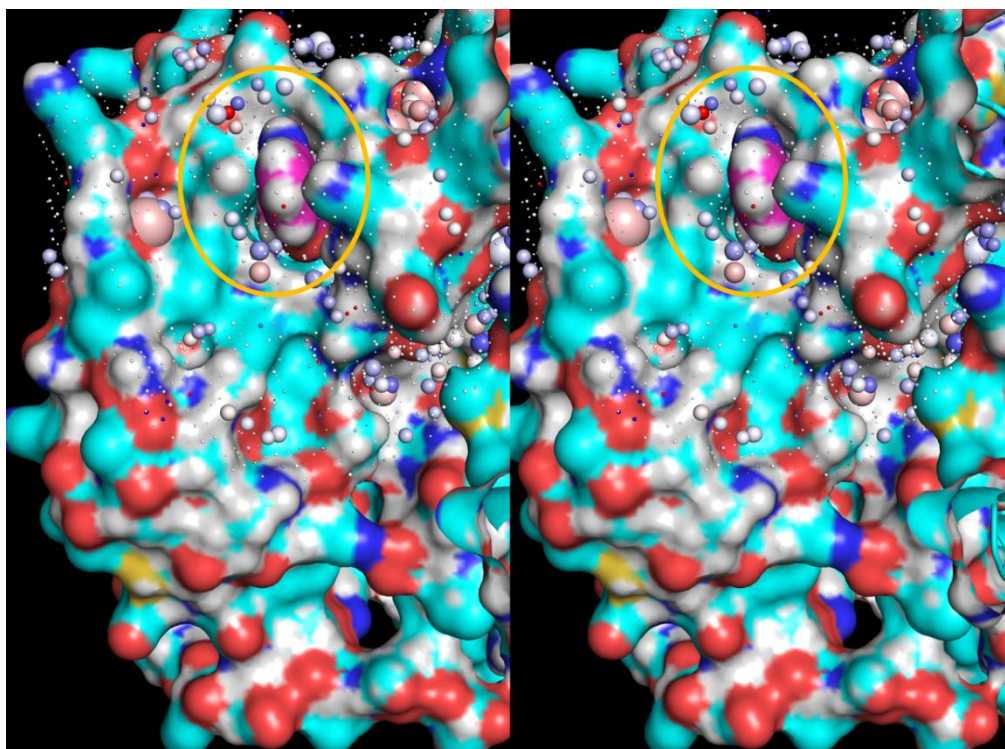
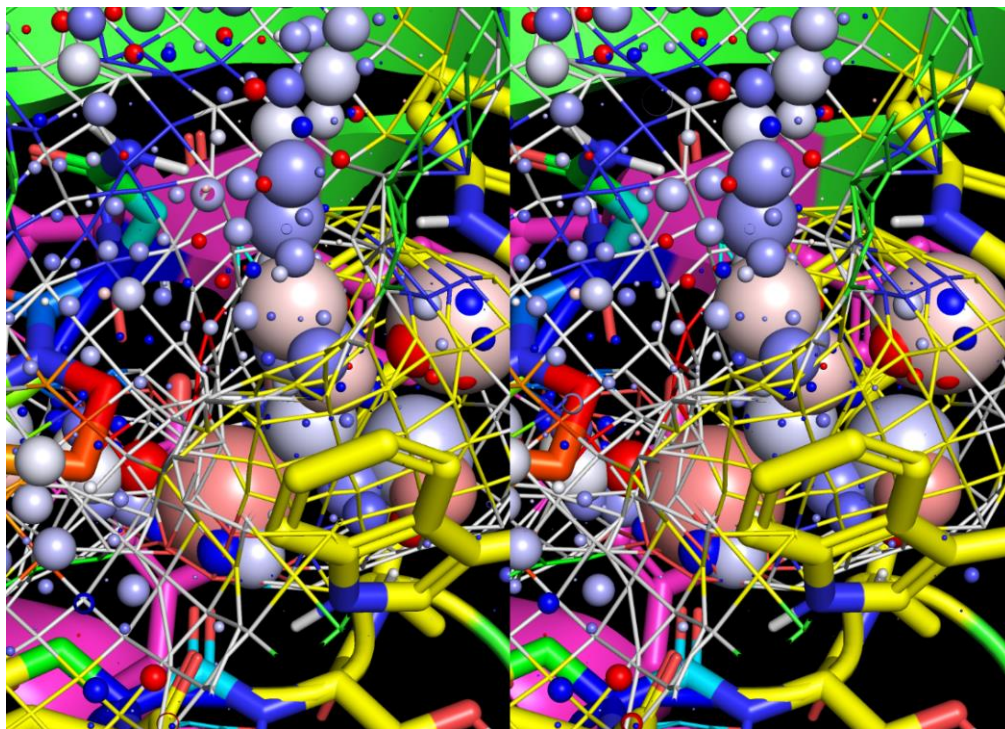


Figure 37. (A) Overlay of structure F (cyan cartoon) with lenalidomide (yellow sticks) and structure K (green cartoon) with uridine (magenta sticks) docked in 5FQD and subjected to a 10 ns MD simulation. (B) Same as A, except showing the HOVs and ULOVs of uridine (magenta sticks). (C) Same as A, except showing the exposed ribose moiety on the substrate binding surface (circled in yellow), together with residual ULOVs and HOVs. The substrate binding surface differs greatly from that of phthalimide-type degraders, suggesting that the native substrate(s) of uridine-bound cereblon differ greatly from neo substrates like CK1 $\alpha$ . We postulate that native substrates bind to an extended interface encompassing the deep butterfly-shaped depression visible in the figure (with uridine residing in the top left “wing”). (D) Same as C, except showing the lenalidomide-bound substrate binding surface (circled in yellow) of structure C (magenta and cyan surfaces, respectively) and the residual ULOVs and HOVs.

Next, we analyzed the partially formed (structure A\_closed) state of pocket L, which contains significantly larger voxels than the fully formed state (structure A\_closed”) (Figure 38) (noting that this state is observed in our simulations solely in the DDB1-unbound form of cereblon, in which the free protein population is likely tipped further toward the degrader binding-incompetent state). This hypothesis is consistent with the reported weak  $-5.2$  kcal/mol  $\Delta H$ <sup>44</sup> for lenalidomide binding to the DDB1-unbound state.

A



B

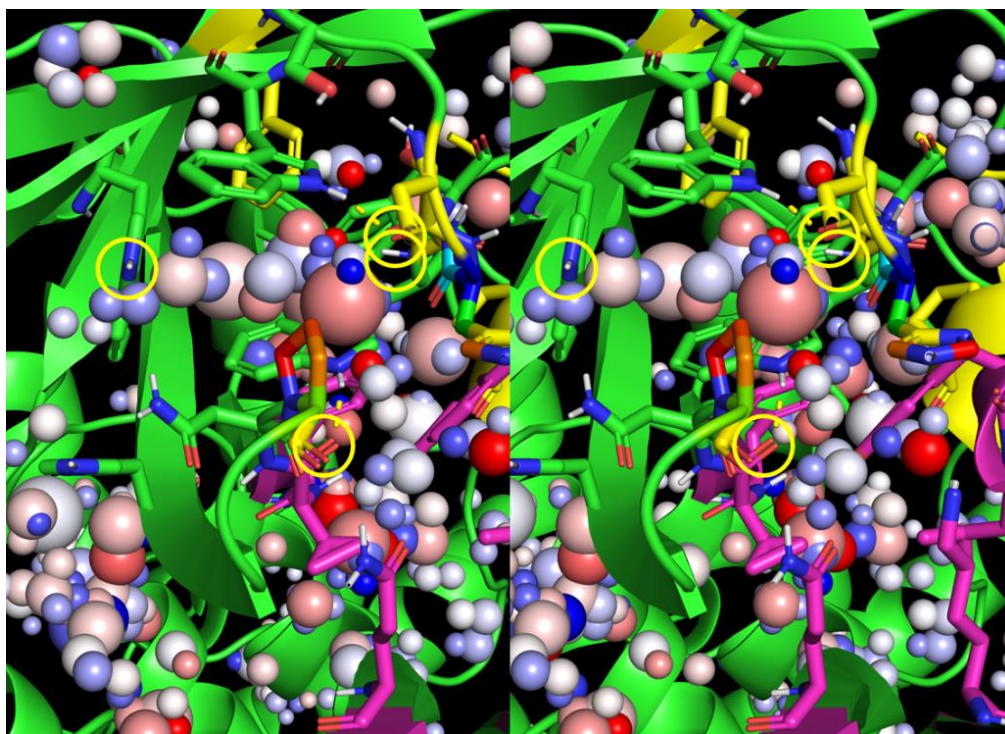


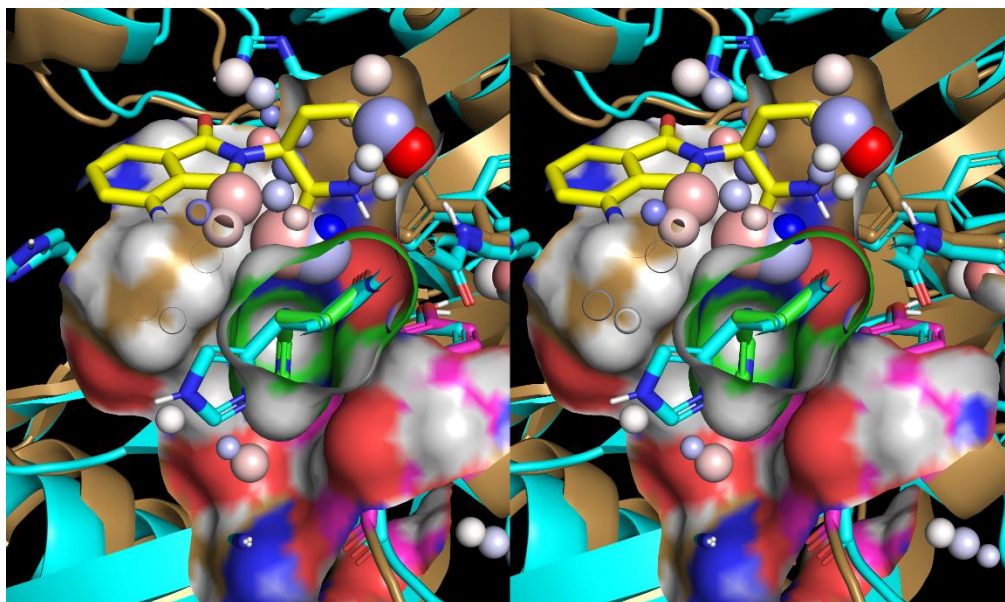
Figure 38. Stereo views of the voxels within the partially formed state of pocket L and fully formed state of pocket T in structure A\_closed. Lenalidomide (yellow) from structure C is overlaid for reference. We postulate that the pocket is destabilized by impeded solvation caused by the narrowed opening, manifesting as non-H-bond enriched HOVs. (A) Zoomed-out view of the voxels, highlighting the restricted flow of water to/from the pocket. (B) The solvating water within pocket L putatively resides at the upper end of a free energy Goldilocks zone, in which the aromatic/non-polar pocket lining is partially counter-balanced by protein H-bond groups (circled in yellow).

Pocket T is most prominent in structure A\_closed (left arrow in Figure 23C), and is also present in structures A\_closed', C, and F (but is fully absent in structures A\_closed'', C', and C''). This pocket resides at the three-way N-terminal domain-N-terminal peptide-CULT domain interface and is occupied by His332 in the degrader-unbound state. We postulate that His332 serves as a linchpin between the aforementioned substructures, which is expelled from pocket T during degrader association via desolvation of an H-bond enriched HOV (Figure 39A). Solvation of pocket T (exchanged for His332) weakens the three-way interface, resulting in dissociation of the N-terminal peptide and rotation of the CULT domain (Figure 39B).

Lastly, we used WATMD to study the solvation properties of the CULT domain within the peri-N-terminal peptide binding interface (the putative “latch”) (Figures 28 and 40A). The N-terminal peptide is stabilized in the CULT-bound state by a large number of ULOVs corresponding to H-bond depleted solvation (small white spheres). We postulate that the peptide dissociates in response to degrader binding via displacement of His332 from pocket T and transient solvation thereof, which weakens the binding interface (Figure 40B). Dissociation of the N-terminal and

CULT domains is putatively promoted by dissociation of this peptide, which we postulate facilitates processive substrate poly-ubiquitylation.

A



B

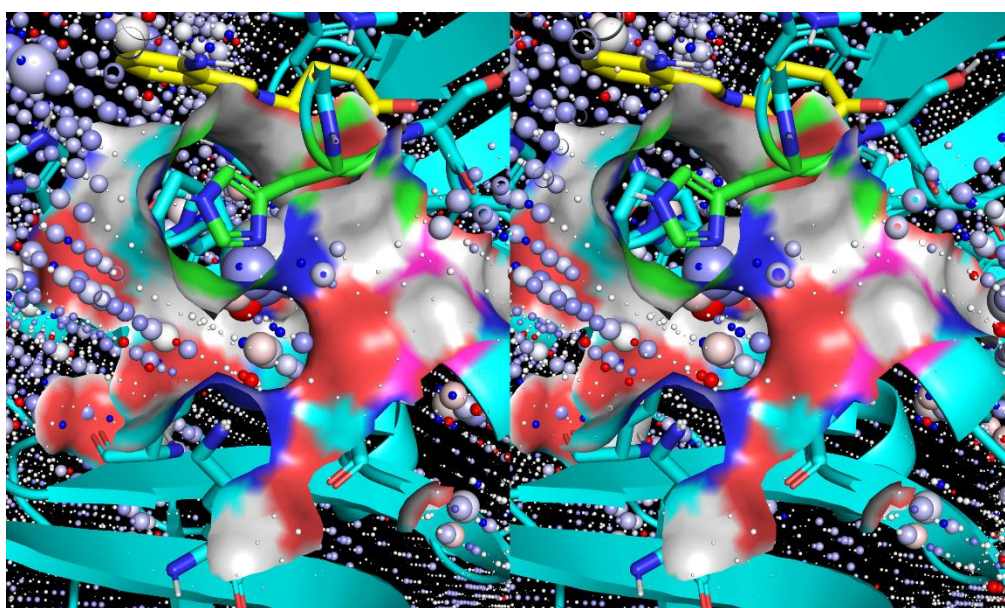
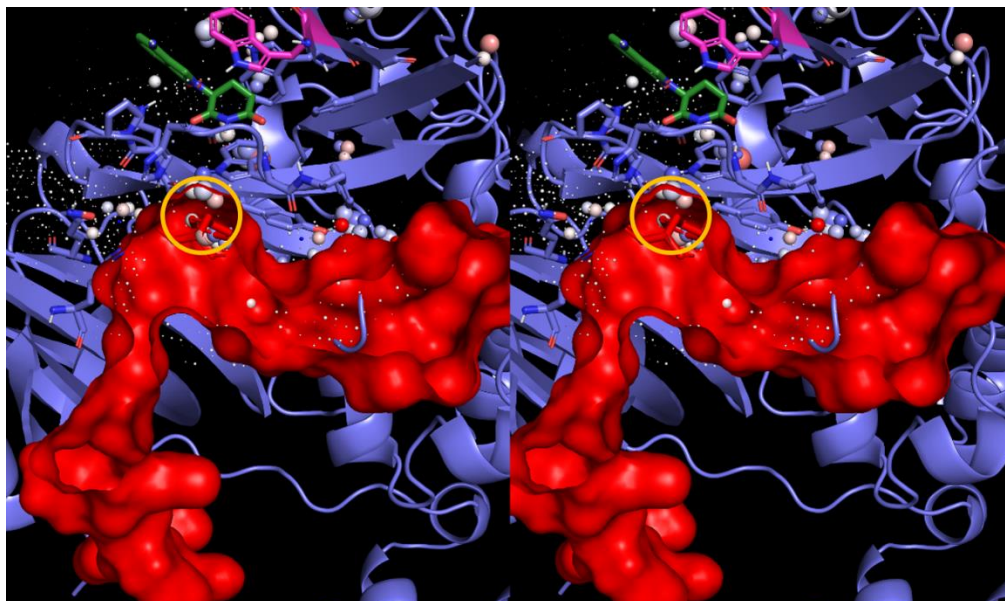


Figure 39. (A) Same as Figure 29C, except showing the waters within the empty pocket L, including a cluster of HOVs corresponding to H-bond enriched solvation that bridges between pockets L and T, and overlaps the southern O of the glutarimide moiety of lenalidomide (overlaid for reference, and shown as yellow sticks). Optimal H-bond replacement of this water by the glutarimide O putatively results in low cost expulsion thereof via pocket T, accompanied by displacement of His332 from the pocket. (B) The degrader-bound state of structure F, in which His332 (green sticks) has been displaced from pocket T by lenalidomide (yellow sticks). Weakening of the three-way interface between the N-terminal domain, N-terminal peptide (magenta surface), and CULT domain by His332 displacement putatively results in dissociation of these substructures, followed by rotation of the CULT domain to the open state captured in 6H0F and structure A\_open (not shown). We postulate that the closed state is adapted for degrader binding (and possibly substrate binding as well), whereas the open state is adapted for substrate ubiquitylation. The closed state present in the starting crystal structure (5FQD) and time-averaged degrader-bound structures is due to crystal packing.

A



B

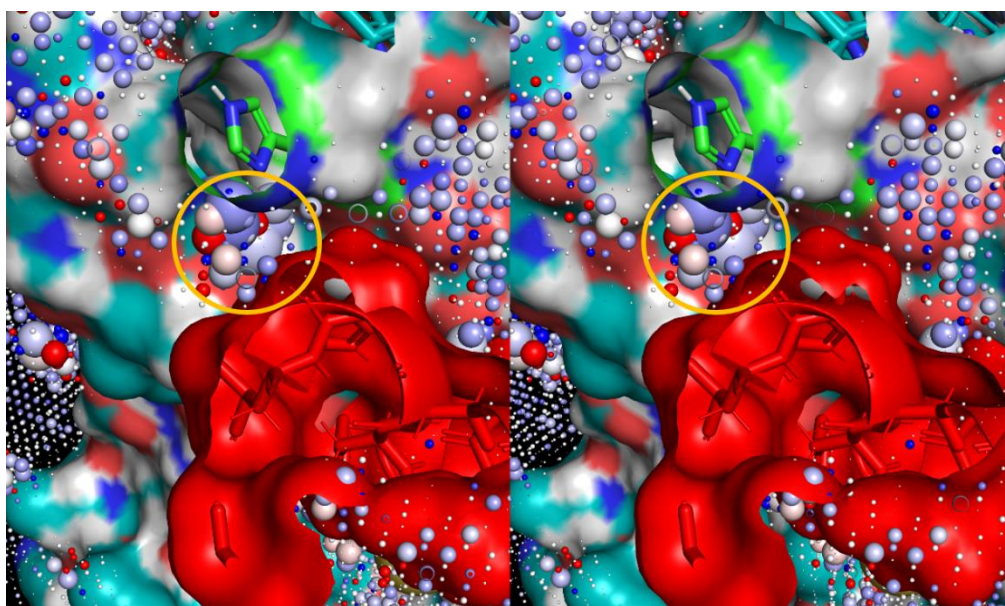


Figure 40. (A) Stereo view of the isolated N-terminal peptide extracted from C (red surface) overlaid on the ULOVs and HOVs of C" (blue cartoon) within the footprint of the N-terminal peptide. The entrance to pocket T is visible at the upper left horizontal section of the peptide (circled in yellow). (B) Close-up stereo view of the entrance to pocket T (circled in yellow) prior to degrader-induced peptide (red surface) dissociation. His332 (green sticks), which is displaced from the pocket, is shown for reference.

## Discussion

### **The non-covalent rearrangements of aqueous solutes are powered by solvation free energy**

We proposed more than a decade ago that non-covalent intra- and intermolecular rearrangements among the sterically-accessible states of aqueous biomolecules and multi-molecular systems under native cellular conditions are powered principally by solvation free energy.<sup>23</sup> More recently, we expanded this concept into a general first principles multi-scale theory on the means by which cellular functions are conveyed by networked molecular fluxes (“physical equations”) operating in the NLD regime.<sup>1,3,4,46</sup> According to our theory, molecular fluxes are powered in a source-sink fashion by covalent free energy and undergo interdependent solvation-powered non-covalent and enzyme-catalyzed state transitions manifesting as cellular states and sub-states (the instantaneous solutions of the multiplexed flux “equations”). In this work, we further extend and exemplify the solvation dynamics concept, in which enthalpic and entropic potential energy is dynamically stored and released in the form of:

- 1) Impeded or trapped water within dynamically rearrangeable internal channels and cavities, in which exchanges between such water and bulk solvent (the reference state of liquid water) are slowed to varying degrees. Trapped/impeded high energy water manifests as



HOVs in our WATMD calculations (not to be confused with HOVs corresponding to low energy solvation at external surfaces).

- 2) Solvating water that is “repelled” from H-bond depleted solute surfaces to bulk solvent. Such water manifests as ULOVs in our calculations.

High potential energy solvation (relative to that of bulk solvent) resides within a Goldilocks zone, where:

- 1) Structural states above the Goldilocks zone of free energy form slowly or not at all.
- 2) Structural states below the Goldilocks zone of free energy are inert (i.e., are stuck in such states).

**Potential energy is released via relaxation of high energy H-bond fields, in which H-bond depleted solvation is expelled (at zero cost) and replaced by non-polar/weakly polar intra- and/or inter-solute groups that likewise exhibit low or no desolvation cost.** Rearrangements necessarily begin and end with the respective storage and release of positive solvation potential energy, including exchanges between:

- 1) Unfavorable states  $i$  and  $j$ , in which H-bond depleted, impeded, or trapped solvation expelled from one or more positions in state  $i$  appears elsewhere in state  $j$  (noting that restrained SD simulations result in artificially exaggerated impedance).
- 2) Transitions to/from states containing counter-balanced favorable-unfavorable solvation configurations to/from states in which the unfavorable contribution is desolvated at the

expense of the favorable contribution (e.g., Trp340 gating in cereblon depicted in Figure 32B).

Molecular populations are distributed across their sterically accessible states under equilibrium conditions according to the free energy of each state relative to the global free energy minimum, or between any states of interest ( $\Delta G_k$ ). Whereas  $G_j^{abs} - G_i^{abs}$  (denoted as scenario 1 (Figure 9A) and  $\Delta G_{i \rightarrow j}^\ddagger - \Delta G_{j \rightarrow i}^\ddagger$  (denoted as scenario 2) (Figure 9B)) are both valid representations of  $\Delta G$  under equilibrium conditions, the interpretation of measured  $\Delta G$  or prediction of  $\Delta G$  based implicitly or explicitly on  $G_j^{abs} - G_i^{abs}$  is, in practice, precluded for the following reasons:

- 1)  $G^{abs}$  resides largely in solvation dynamics ( $G^{abs} \approx G^{bulk} - (G^{enriched} + G^{depleted}) = \delta G$ ), rather than solute structure per se. However,  $G^{abs}$  is implicitly assumed in force-field (FF)- and quantum mechanics (QM)-based computational approaches to reside principally in solute interatomic contacts. As such,  $\Delta G$  (or  $\Delta \Delta G$  among closely related molecules) cannot be predicted reliably using such methods in the absence of data fitting (e.g., free energy perturbation).
- 2) Water-water, water-solute and solute-solute H-bond contributions to free energy (principally internal energy) are weighted similarly in FF-based methods, and solute-solute contacts are overweighted relative to water-water and water-solute H-bond free energy contributions (noting that *ab initio* QM methods are impractical for explicitly solvated systems). We have provided examples of inter- and intramolecular cereblon rearrangements that are ostensibly driven by overweighted H-bond force-field energies, including rotation of Trp340 into pocket L and retention of the interfacial H-bonds across the hypothetical binary cereblon-CK1 $\alpha$  interface (time-averaged structure B). Whereas

Trp340 rotation makes sense in the context of pocket L desolvation (to which the AMBER force-field is largely ambivalent), the resemblance of the time-averaged structure B to the starting ternary structure D does not. As such, structural rearrangements predicted by MD simulations are more likely to be right for the wrong reasons (or entirely wrong), unlike SD simulations that are aimed at predicting the properties of H-bond-dominated solvation fields.

- 3) All of the aforementioned contributions to  $G^{abs}$  depend on dynamic water exchanges between bulk solvent and solvation that are neglected in both FF- and QM-based computational methods, which are limited to static solute structures and solvation configurations. Our WATMD results suggest that the H-bonds of solvating water switch continually among co-localized solute acceptors and donors, either via rotational/translational rearrangements (“smearing”) or during separate exchanges with bulk solvent, manifesting as HOV clusters that, in many cases, exhibit a range of water H and O preferences. H-bond switching reduces entropic losses incurred when solvating water is constrained rigidly in concurrent multi-H-bond interactions (noting that water participates in 3.5 H-bonds/molecule in bulk solvent via rapid partner-partner switching, in which entropic losses are avoided<sup>3</sup>).

Scenario 2 applies to both the biochemically relevant equilibrium and native cell relevant non-equilibrium regimes, whereas scenario 1 is limited strictly to the former regime. Furthermore, the qualitative interpretation of measured  $\Delta G$  or the prediction of  $\Delta G$  based on  $\Delta G_{i \rightarrow j}^{\ddagger} - \Delta G_{j \rightarrow i}^{\ddagger}$  within the Biodynamics framework is possible for the following reasons:

- 1) An explicit relationship is assumed between  $\Delta G_{i \rightarrow j}^\ddagger$  and desolvation and  $\Delta G_{j \rightarrow i}^\ddagger$  and resolution during the  $i \rightarrow j$  and  $j \rightarrow i$  state transitions, respectively.
- 2) SD simulations exploit the fact that water-water H-bonds dominate the free energy landscape in aqueous systems, manifesting as high sensitivity of solvating water to H-bond gains and losses relative to bulk solvent (capable of resolving a single carbon atom difference among solutes<sup>3</sup>).
- 3)  $\delta G^{enriched} = (G^{bulk} - G^{enriched})$  and  $\delta G^{depleted} = (G^{bulk} - G^{depleted})$  are assumed to be reflected qualitatively in the time-averaged voxel occupancies calculated by WATMD.

The overall non-linear dynamic behavior of solutes and their solvation fields is treated implicitly in SD simulations and WATMD, but not in free energy calculations in which static solute-solvation configurations are assumed. Solvation contributions to predicted structural state preferences are therefore further underestimated relative to the intra- and inter-solute contributions by these methods due to their heavy bias toward maximization of van der Waals contacts, H-bonds, electrostatic contacts, ring-stacking, etc.

In our previous work, we used Biodynamics and WATMD to study the putative intramolecular and observed intermolecular rearrangements of COVID M<sup>pro</sup> in the non-equilibrium context.<sup>4</sup> Here, we have used these methods to probe the structure and function of the binary and ternary complexes of the cereblon subunit of the CUL4A/RBX1/DDB1/CRBN E3 ligase, LMW degraders, and the neo substrate CK1 $\alpha$ , likewise in the non-equilibrium context. Substrate binding to both proteins, degrader binding to cereblon, and M<sup>pro</sup> dimerization are powered by external solvation, whereas the up/down state transitions of M<sup>pro</sup> and release of the N-terminal peptide of cereblon are powered by internal solvation similar to that which we postulated for M<sup>pro</sup>.<sup>4</sup> We have

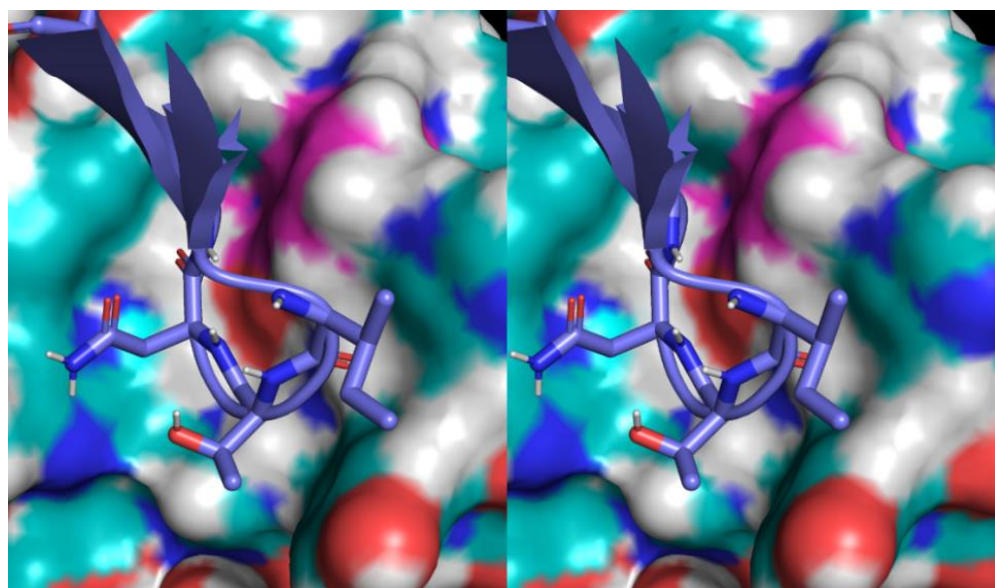
further postulated that both  $M^{pro}$  and cereblon are activated by substrate-induced dimerization and degrader binding, respectively. We propose that HMW-HMW (e.g., protein-protein) and HMW-LMW binding are, in general, powered predominantly by desolvation and resolution costs underlying  $k_{on}$  (or  $k_1$ ) and  $k_{off}$  (or  $k_{-1}$ ), respectively, which may be regulated by additional inputs (e.g., allosteric activators/inhibitors) governing interfacial surface shape and H-bond group complementarity among the partners. We further propose that the solvation fields across the full complement of aqueous biomolecules constitute a new omics class that we refer to as the “solvome” (form follows function follows free energy).

### **Cereblon is switched by degraders from binding incompetent to competent forms**

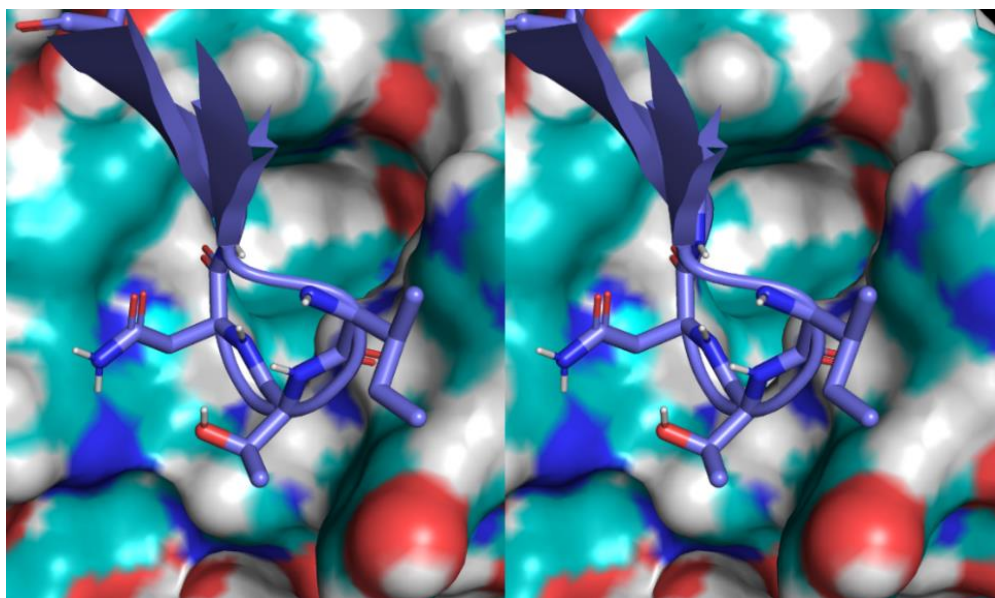
Our findings challenge the widely held view that cereblon is “glued” to its substrates via LMW degraders (i.e., acting as a switch between fast  $k_{-1}$  in their absence and slow  $k_{-1}$  in their presence). The relatively small fraction of ULOVs contributed by lenalidomide in the cereblon-lenalidomide complex is apparent in Figure 30B (the yellow region of the surface above lenalidomide), consistent with incremental, rather than radical  $k_{-1}$  slowing in ternary complex formation. **Furthermore, impeded substrate turnover in the binary complex due to fast  $k_{-1}$  relative to  $k_{cat}$ , rather than slowed  $k_1$  (where  $k_1$  remains similar to that of the ternary complex), would result in substrate inhibition of cereblon.** Instead, Type I degraders far more plausibly speed  $k_1$  (i.e., by lowering the association barrier) versus slow  $k_{-1}$  (i.e., by raising the dissociation barrier), noting that slowing  $k_{-1}$  from the ternary complex is well-explained by the association-induced release of H-bond depleted water present in the overall cereblon-degrader-CK1 $\alpha$  interface. **Therefore, we propose that Type I degraders activate cereblon by aligning the surface shape and key exposed H-bond groups with those of its cognate substrates** (Table 3). The substrate

binding surface of binary cereblon-Type I degrader complexes consists of both cereblon and the exposed phthalimide ring of bound synthetic degraders (Figures 41A and B) or the ribose ring of uridine (Figure 41 C), which putatively speed  $k_1$  (and to a lesser extent, slow  $k_{-1}$ ), and further contribute to substrate selectivity/complementarity. The smaller footprint of BRD4 on the corresponding cereblon-degrader complex (Figure 15C) relative to that of CK1 $\alpha$  (Figure 15A) is consistent with insufficient desolvation of H-bond depleted solvation to achieve a physiologically meaningful  $k_{-1}$ . As such, Type II degraders are expected to both speed  $k_1$  and slow  $k_{-1}$  via expansion of the interface deeper within the target substrate (acting as both “glues” and “activators”).

A



B



C

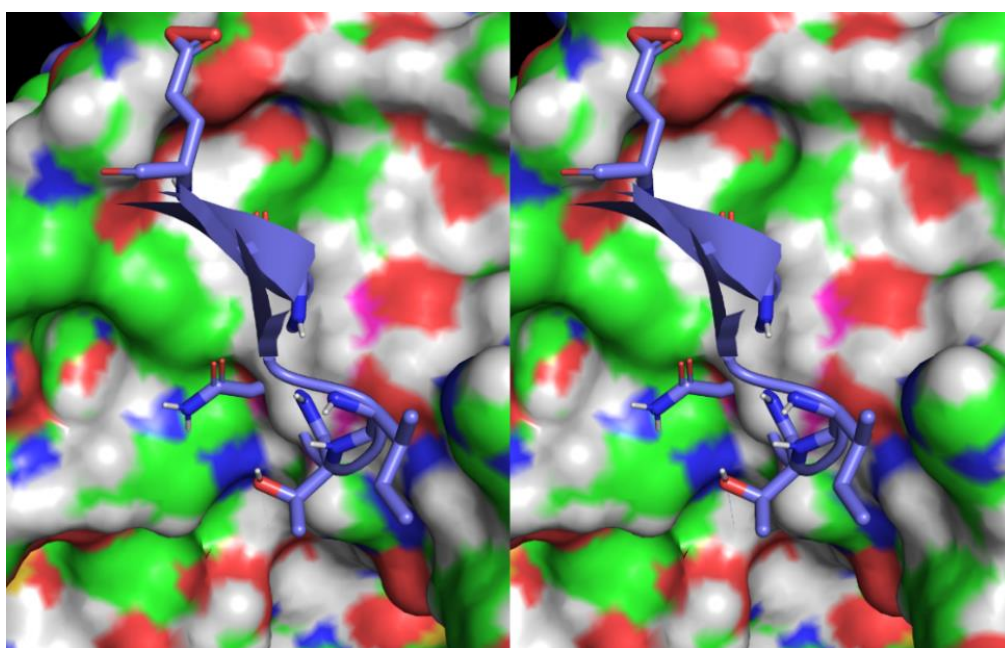


Figure 41. (A) Stereo view of the blended surfaces of cereblon plus lenalidomide (blue and magenta, respectively), together with CK1 $\alpha$  (blue cartoon) (structure F). High complementarity between the binary surface and the substrate is apparent. (B) Same as A, except showing the cereblon surface in the absence of the degrader contribution. (C) Same as A, except for docked uridine (structure K) with CK1 $\alpha$  from structure F overlaid for reference. The CK1 $\alpha$  pocket is completely absent in this structure, consistent with the lack of CK1 $\alpha$  binding.

Table 3. Implications of “glue” versus “accelerator/activator” mechanisms for substrate-cereblon binding kinetics in the absence ( $k'_1, k'_{-1}$ ) and presence ( $k_1, k_{-1}$ ) of Type I and II degraders (noting that, based on our rationale, Type I and II degraders are designated as “activator” and “glues”, respectively).

		<b>Ratio of substrate association and dissociation rate constants in the absence (primed) and presence (unprimed) of degraders</b>	
<b>Type</b>	<b>Scenario</b>	$k'_1/k_1$	$k'_{-1}/k_{-1}$
II	Glue (slows $k_{-1}$ )	$\sim 1$	$\gg 1$
I	Activator (speeds $k_1$ )	$\ll 1$	$\gtrsim 1$



## Conclusion

We have extended our previous work on molecular fluxes,<sup>1,4,5</sup> focusing on the general means by which:

- 1) Cellular behaviors are conveyed by multiplexed multi-flux systems operating in the non-linear dynamics regime via a molecular form of analog computing. Fluxes undergo switchable covalent and non-covalent intra- and intermolecular state transitions governed by the overall system behavior.
- 2) Switching is powered by desolvation of H-bond depleted water from internal/buried and external surfaces.
- 3) Unfavorable and favorable free energy is stored in the H-bond network of water solvating both HMW and LMW species (referred to as the “solvation field”), equating to entry and exit barriers underlying the intrinsic rates) to/from non-covalent states.
- 4) Solvation fields are governed by solute composition and structure (i.e., structure-solvation field relationships).

We exemplified the aforementioned principles via cereblon, the substrate-recruiting subunit of the CUL4A/RBX1/DDB1/CRBN E3 ligase complex, which serves as the sink for both native and neo substrates. In particular, we explored the:

- 1) Relationships between the extrinsic and intrinsic rates of cereblon binding to LMW Type I and II degraders and substrates (assuming that the partners build and decay over time) using the symbolic mathematical model provided in the Supplementary Information, **focusing on the optimal rates for achieving the maximum dynamic degrader**

**occupancy at the lowest possible degrader concentration (based on the “kinetic tuning” concept that we advanced previously<sup>1</sup>).** We postulate that for Type I degraders, the substrate degradation rate ( $k_{cat(4)}$ ) is optimally  $\gg$  the rate of downstream substrate-mediated effector activation ( $k_{16}$ ), and the rate of binary cereblon-degrader formation ( $k_9$ ) is optimally  $\approx$  the rate of cereblon buildup ( $k_7$ ). We postulate that for Type II degraders the rate of degrader-substrate binding ( $k_{20}$ ) is ideally  $>$  the rate of degrader-cereblon binding ( $k_9$ ), which simultaneously promotes substrate inhibition and degradation, while limiting cereblon inhibition.

- 2) Qualitative structure-solvation field relationships of cereblon and its binding partners to deduce the solvation-powered dynamic binding mechanism (analogous to our previous work on COVID M<sup>Pro</sup>).<sup>4</sup> We postulate that occupancy of the degrader pocket of cereblon shifts the population toward the substrate binding-competent state, which is otherwise populated predominantly in degrader and substrate binding-incompetent states. We further postulate that degrader binding is powered by the release of high potential energy stored as H-bond depleted solvation within the empty degrader pocket (counter-balanced by co-localized H-bond enriched solvation), which is transduced into alignment between the interfacial H-bond groups of cereblon with those of cognate substrates, followed by release and rotation of the N-terminal peptide from the CULT domain. We proposed that Trp340 and His332 play key roles in “gating” the empty degrader pocket L and the adjacent water-filled pocket T. Last but not least, we showed that the solvation fields of uridine, lenalidomide, and thalidomide are highly complementary to that of the degrader pocket, translating to low (but specific) desolvation costs of the associating degrader species.

## References

1. Selvaggio, G. & Pearlstein, R. A. Biodynamics: A novel quasi-first principles theory on the fundamental mechanisms of cellular function/dysfunction and the pharmacological modulation thereof. *PLoS One* **13**, e0202376 (2018).
2. Pearlstein, R. A., Wan, H. & Aravamuthan, V. Toward in vivo relevant drug design. *Drug Discov. Today* **26**, 637–650 (2021).
3. Pearlstein, R. A. *et al.* Building New Bridges between In Vitro and In Vivo in Early Drug Discovery: Where Molecular Modeling Meets Systems Biology. *Curr. Top. Med. Chem.* **17**, 1–1 (2017).
4. Wan, H., Aravamuthan, V. & Pearlstein, R. A. Probing the Dynamic Structure-Function and Structure-Free Energy Relationships of the Coronavirus Main Protease with Biodynamics Theory. *ACS Pharmacol. Transl. Sci.* **3**, 1111–1143 (2020).
5. Wan, H., Spiru, K., Williams, S. & Pearlstein, R. A. General structure-free energy relationships of hERG blocker binding under native cellular conditions. *bioRxiv* (2021) doi:10.1101/2021.10.07.463585.
6. Wan, H., Selvaggio, G. & Pearlstein, R. Toward in vivo -relevant hERG safety assessment and mitigation strategies based on relationships between non-equilibrium blocker binding, three-dimensional channel-blocker interactions, dynamic occupancy, dynamic exposure, and cellular arrhythmia. *PLoS One* **15**, e0234946 (2020).
7. Pearlstein, R. A. *et al.* Implications of Dynamic Occupancy, Binding Kinetics, and Channel Gating Kinetics for hERG Blocker Safety Assessment and Mitigation. *Curr. Top. Med. Chem.* **16**, 1792–1818 (2016).
8. Eshraghi, M. *et al.* Mutant Huntingtin stalls ribosomes and represses protein synthesis in a

- cellular model of Huntington disease. *Nat. Commun.* **12**, 1461 (2021).
9. Ou, S., Hu, Y., Patel, S. & Wan, H. Spherical monovalent ions at aqueous liquid-vapor interfaces: Interfacial stability and induced interface fluctuations. *J. Phys. Chem. B* **117**, 11732–11742 (2013).
  10. Wan, H., Zhou, G. & Voelz, V. A. A Maximum-Caliber Approach to Predicting Perturbed Folding Kinetics Due to Mutations. *J. Chem. Theory Comput.* **12**, 5768–5776 (2016).
  11. Wan, H., Ge, Y., Razavi, A. & Voelz, V. A. Reconciling Simulated Ensembles of Apomyoglobin with Experimental Hydrogen/Deuterium Exchange Data Using Bayesian Inference and Multiensemble Markov State Models. *J. Chem. Theory Comput.* **16**, 1333–1348 (2020).
  12. Wan, H. & Voelz, V. A. Adaptive Markov state model estimation using short reseeding trajectories. *J. Chem. Phys.* **152**, (2020).
  13. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* vol. 9 646–652 (2002).
  14. Levitt, M. Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.* **170**, 723–764 (1983).
  15. Voelz, V. A., Bowman, G. R., Beauchamp, K. & Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J. Am. Chem. Soc.* **132**, 1526–1528 (2010).
  16. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science (80-. )*. **334**, 517–520 (2011).
  17. Shan, Y. *et al.* How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **133**, 9181–9183 (2011).

18. LaConte, L. E. W., Voelz, V., Nelson, W., Enz, M. & Thomas, D. D. Molecular dynamics simulation of site-directed spin labeling: Experimental validation in muscle fibers. *Biophys. J.* **83**, 1854–1866 (2002).
19. Grant, B. J. *et al.* Novel allosteric sites on ras for lead generation. *PLoS One* **6**, e25711 (2011).
20. Perez, A., MacCallum, J. L. & Dill, K. A. Accelerating molecular simulations of proteins using Bayesian inference on weak information. *Proc. Natl. Acad. Sci.* **112**, 11846–11851 (2015).
21. Sheinerman, F. B. & Brooks, C. L. Molecular picture of folding of a small  $\alpha/\beta$  protein. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1562–1567 (1998).
22. Zhao, J. *et al.* EGCG binds intrinsically disordered N-terminal domain of p53 and disrupts p53-MDM2 interaction. *Nat. Commun.* **2021 121** **12**, 1–11 (2021).
23. Pearlstein, R. A. *et al.* New hypotheses about the structure-function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor-like repeat A docking site using watermap. *Proteins Struct. Funct. Bioinforma.* **78**, 2571–2586 (2010).
24. Pearlstein, R. A., Sherman, W. & Abel, R. Contributions of water transfer energy to protein-ligand association and dissociation barriers: Watermap analysis of a series of p38 $\alpha$  MAP kinase inhibitors. *Proteins Struct. Funct. Bioinforma.* **81**, (2013).
25. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 808–813 (2007).
26. Tran, Q.-T. *et al.* Structure-kinetic relationship of carbapenem antibacterials permeating through E. coli OmpC porin. *Proteins Struct. Funct. Bioinforma.* **82**, (2014).

27. Tran, Q. T., Williams, S., Farid, R., Erdemli, G. & Pearlstein, R. The translocation kinetics of antibiotics through porin OmpC: Insights from structure-based solvation mapping using WaterMap. *Proteins Struct. Funct. Bioinforma.* **81**, 291–299 (2013).
28. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
29. Fouad, S., Wells, O. S., Hill, M. A. & D'Angiolella, V. Cullin Ring Ubiquitin Ligases (CRLs) in Cancer: Responses to Ionizing Radiation (IR) Treatment. *Frontiers in Physiology* vol. 10 (2019).
30. Hartmann, M. D. *et al.* Thalidomide mimics uridine binding to an aromatic cage in cereblon. *J. Struct. Biol.* **188**, 225–232 (2014).
31. Boichenko, I. *et al.* Chemical Ligand Space of Cereblon. *ACS Omega* **3**, 11163–11171 (2018).
32. Boichenko, I., Deiss, S., Bär, K., Hartmann, M. D. & Hernandez Alvarez, B. A FRET-Based Assay for the Identification and Characterization of Cereblon Ligands. *J. Med. Chem.* **59**, 770–774 (2016).
33. Sievers, Q. L. *et al.* Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science (80-. ).* **362**, (2018).
34. Matyskiela, M. E. *et al.* A novel cereblon modulator recruits GSPT1 to the CRL4 CRBN ubiquitin ligase. *Nature* **535**, 252–257 (2016).
35. Chamberlain, P. P. *et al.* Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nat. Struct. Mol. Biol.* **21**, 803–809 (2014).
36. Petzgold G, Fischer ES & Thoma NH. Structural basis of lenalidomide-induced CK1a

- degradation by the CRL4 (CRBN) ubiquitin ligase. *Nature* **532**, 127–130 (2016).
37. Schreiber, S. L. The Rise of Molecular Glues. *Cell* **184**, 3–9 (2021).
  38. Chamberlain PP, H. L. Development of targeted protein degradation therapeutics. *Nat Chem Biol* **15**, 937–944 (2019).
  39. Han, B. A suite of mathematical solutions to describe ternary complex formation and their application to targeted protein degradation by heterobifunctional ligands. *J. Biol. Chem.* **295**, 15280–15291 (2020).
  40. Petzold, G., Fischer, E. S. & Thomä, N. H. Structural basis of lenalidomide-induced CK1 $\alpha$  degradation by the CRL4 CRBN ubiquitin ligase. *Nature* **532**, 127–130 (2016).
  41. Nowak, R. P. *et al.* Plasticity in binding confers selectivity in ligand-induced protein degradation article. *Nat. Chem. Biol.* **14**, 706–714 (2018).
  42. Fischer, E. S. *et al.* Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49 (2014).
  43. Krönke, J. *et al.* Lenalidomide induces ubiquitination and degradation of CK1 $\alpha$  in del(5q) MDS. *Nature* **523**, 183–188 (2015).
  44. Akuffo, A. A. *et al.* Ligand-mediated protein degradation reveals functional conservation among sequence variants of the CUL4-type E3 ligase substrate receptor cereblon. *J. Biol. Chem.* **293**, 6187–6200 (2018).
  45. Velez-Vega, C. *et al.* Estimation of solvation entropy and enthalpy via analysis of water oxygen-hydrogen correlations. *J. Chem. Theory Comput.* **11**, (2015).