Published in final edited form as:

J Mol Biol. 2008 February 1; 375(5): 1416–1433.

Coarse-grained models for simulations of multi-protein complexes:

Application to ubiquitin binding

Young C. Kim and Gerhard Hummer*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, U.S.A.

Abstract

We develop coarse-grained models and effective energy functions for simulating thermodynamic and structural properties of multi-protein complexes with relatively low binding affinity ($K_d > 1 \mu\text{M}$) and apply them to binding of Vps27 to membrane-tethered ubiquitin. Folded protein domains are represented as rigid bodies. The interactions between the domains are treated at the residue level with amino-acid dependent pair potentials and Debye-Hückel-type electrostatic interactions. Flexible linker peptides connecting rigid protein domains are represented as amino-acid beads on a polymer with appropriate stretching, bending, and torsion-angle potentials. In simulations of membrane-attached protein complexes, interactions between amino acids and the membrane are described by residue-dependent short-range potentials and long-range electrostatics. We parameterize the energy functions by fitting the osmotic second virial coefficient of lysozyme and the binding affinity of the ubiquitin-CUE complex. For validation, extensive replica-exchange Monte Carlo simulations are performed of various protein complexes. Binding affinities for these complexes are in good agreement with the experimental data. The simulated structures are clustered on the basis of distance matrices between two proteins and ranked according to cluster population. In ~70% of the complexes, the distance root mean square (DRMS) is less than 5 Å from the experimental structures. In ~90% of the complexes, the binding interfaces on both proteins are predicted correctly, and in all other cases at least one interface is correct. Transient and non-specifically bound structures are also observed. With the validated model, we simulate the interaction between the Vps27 multiprotein complex and a membrane-tethered ubiquitin. Ubiquitin is found to bind preferentially to the two UIM domains of Vps27, but transient interactions between ubiquitin and the VHS and FYVE domains are observed as well. These specific and non-specific interactions are found to be positively cooperative, resulting in a substantial enhancement of the overall binding affinity beyond the ~300 μM of the specific domains. We also find that the interactions between ubiquitin and Vps27 are highly dynamic, with conformational rearrangements enabling binding of Vps27 to diverse targets as part of the multivesicular-body protein-sorting pathway.

Keywords

protein-protein interactions; protein-protein docking; replica exchange Monte Carlo; energy landscape; energy function; contact potentials; multivesicular body; Vps27

*Corresponding Author: Bldg. 5, Rm. 132, NIH, Bethesda, MD 20892-0520, email: hummer@helix.nih.gov, Phone: 301-402-6290, Fax: 301-496-0825

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Protein-protein interactions play an essential role in the biological function of many proteins.^{1,2} Nevertheless, the number of protein complexes with structures determined at atomic resolution is relatively small, despite the enormous advances in X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryoelectron microscopy. In fact, most of the structures deposited in the Protein Data Bank (PDB) are individual isolated monomeric proteins. A major difficulty in resolving protein complex structures arises from the fact that they are often too transient to be subject to structural characterization. Computational methods have thus become useful and important tools to provide valuable complementary information in the characterization of protein-protein interactions.³

A broad range of docking algorithms has been developed to predict the structures of protein complexes.⁴⁻¹⁹ The main focus has been on specific binding with high affinity (dissociation constant in the nM to pM range) for possible application to drug design. Most protein docking studies have concentrated on binary protein complexes. However, interactions between proteins inside a cell often involve many constituents, such as multiple proteins, membranes, DNAs, RNAs, etc.²⁰⁻²⁵ In addition, many proteins have several folded domains connected by unstructured peptides. For example, Vps27, responsible for membrane trafficking, contains VHS, FYVE, UIM1 and UIM2 domains that are connected by flexible linkers.²⁴ These folded domains are thought to remain almost rigid during the interactions with other proteins, but the flexible linkers connecting the domains permit large conformational changes of the overall system.

Studies of Vps27 and similar multi-domain protein complexes with conventional docking face major challenges. Understanding such multi-protein interactions requires not only binary docking but also cooperativity between the various components, including proteins and membrane. Simulating such large complexes, however, cannot be achieved easily by available docking algorithms that are suited mainly for binary interactions. Furthermore, thermodynamic properties, such as the binding affinity, or transient intermediate structures not detectable easily by experiment, are not considered by many algorithms.

Here, our goal is to develop a simple model for simulating thermodynamic and structural properties of multi-protein complexes. Routine atomic-level simulations of multi-protein complexes to predict their structure, stability, and dynamics are currently not practical. As possible alternatives, several coarse-grained representations for proteins have been introduced and adapted in various docking algorithms.³ In this work, proteins or folded domains are treated as rigid bodies. Flexible linkers between the domains are treated as polymers at the amino-acid level. The interactions between proteins are represented at the residue-level. The pair potentials between two residues are adapted from statistical contact potentials, which have been developed originally for investigations of protein folding and protein-protein interactions.²⁶⁻³⁴ The long-range electrostatic interactions between charged residues are treated at the Debye-Hückel level with explicit salt-concentration dependence. For the interactions between proteins and a membrane, residue-dependent membrane potentials are used, which have been derived from distributions of amino acids in the trans-membrane domain.³⁵ While the focus here is on calculations of the structure and binding affinity of weakly-interacting protein complexes, the potential energy function is also well-suited for Brownian dynamics simulations to explore the association and dissociation kinetics of protein-protein complexes.³⁶⁻³⁹

The outline of the paper is as follows. We first determine the free parameters in the model energy function by using experimental affinity data, without using complex-structure information. In particular, we use data for (1) the second virial coefficient of lysozyme and (2)

the binding affinity of the ubiquitin-CUE complex. Among a group of models that differ in the way surface exposure is accounted for, we select an optimal model based on the binding affinity for a ubiquitin-UIIM complex. In a second stage, we validate that model first by comparing the calculated second virial coefficient of staphylococcal nuclease to experiment. We also use the model to study the binding affinities and complex structures of cytochrome *c* and cytochrome *c* peroxidase, of seven additional ubiquitin complexes not used in parametrization and model selection, and of various other protein complexes (K_d of 0.01 to $\sim 300 \mu\text{M}$). We show that for all ubiquitin complexes and about half of the other complexes the model predicts accurate complex structures without having used complex-structure information during parametrization. In particular, the model identifies the proper binding interface of at least one protein in all cases studied, and in most cases has the top-ranked complex configurations within less than 5 Å DRMS (Distance Root Mean Square) from the respective X-ray crystal and NMR solution structures of the complexes. Finally, we apply the model to study the interaction of a membrane-tethered ubiquitin and the Vps27 complex, which is involved in sorting cargo proteins from early endosomes to the ESCRT-I complex. Here we focus on binding of ubiquitin to various Vps27 domains in the presence of a membrane. We address the biologically important question how Vps27 can be recruited to form a sufficiently long-lived complex with ubiquitin-tagged membrane proteins when the measured interactions of the Vps27 UIM domains and ubiquitin are weak, with binding affinities of $\sim 300 \mu\text{M}$? Furthermore, we explore the structures and motions of ubiquitin-bound Vps27 to gain insights into the relation between conformational flexibility and the ability to bind diverse targets. We conclude with a discussion of the Vps27 results as well as the quality, limitations, and possible applications of the model.

Results

Parametrization of the energy functions

The potential energy model contains two free parameters, λ and e_0 (see Models and Methods), that account for the strength of non-electrostatic interactions, and the affinity between amino acids relative to that of the solvent, respectively. We first determine λ as a function of e_0 by fitting the second virial coefficient of lysozyme. The optimal (λ, e_0) parameter combination is then obtained by fitting the binding affinity of the ubiquitin-CUE complex. This procedure is repeated for six models A-F that account differently for the relative surface exposure of amino acids.

λ and e_0 are determined by fitting the second virial coefficient of hen egg lysozyme (PDB code: 1W6Z), which has been studied extensively by experiments.⁴⁰ Using the molecular weight $M_w \simeq 14400$ of lysozyme, we calculate the second virial coefficient over a wide range of the two-dimensional parameter space (e_0, λ) . Figure 1(a) presents B_{22} as a function of the scale factor λ for different offsets e_0 , calculated for model A. As expected, smaller offsets e_0 yield more negative values for B_{22} (i.e., more attractive interactions), since more pairs of residues become attractive. For a given e_0 , B_{22} decreases with λ , indicating that stronger attractive interactions are induced between the two proteins. Similar behavior for B_{22} is observed for the other models B-F. Figure 1(b) shows combinations of the scale factor λ and offset e_0 that yield the experimentally measured $B_{22} \simeq -3 \times 10^{-4} \text{ mol}\cdot\text{ml}/\text{g}^2$ at pH = 7 and 100 mM salt concentration.⁴⁰

To determine an optimal combination of the two parameters, a second set of independent experimental data is used. Here, we compute the binding affinity of the complex between ubiquitin and the CUE domain of the Cue2 protein (PDB code: 1OTR), which has been measured experimentally as $K_d \simeq 155 \mu\text{M}$.⁴¹ Replica exchange Monte Carlo (REMC) simulations are performed on the complex for a system within a sphere of radius $R = 140 \text{ Å}$, which corresponds to $145 \mu\text{M}$ protein concentration. Figure 1(c) presents the binding probability calculated for the (λ, e_0) parameter combinations in Figure 1(b) for the different

models A-F. Interestingly, the binding probability is a sigmoidal function of the scale parameter λ (with similar dependences on the offset e_0). For larger λ values, the proteins experience stronger attractive interactions resulting in larger probabilities of being bound. Using the experimental $K_d \approx 155 \mu\text{M}$ for the complex, we obtain the optimal parameter combinations for each of the models A-F, as listed in Table 1. For model A, in which all residues are weighted equally irrespective of their distance from the protein surface, the fitted offset, $e_0 = -2.27 k_B T$, is close to the value $-2.55 k_B T$ obtained originally by Miyazawa and Jernigan.²⁷ As the inner residues are weighted less in models B-F, the offset increases so that more residue pairs become attractive. This can be understood by the fact that without the contributions from inner residues the pair-wise interaction ϕ_{ij} is required to be more attractive in order to yield the same binding affinity.

As an initial validation of the parameter values, the second virial coefficient of staphylococcal nuclease (PDB code: 1SNO) is calculated. For models A-F, the estimated second virial coefficients lie between 0.5×10^{-4} and $1.2 \times 10^{-4} \text{ mol} \cdot \text{ml/g}^2$, which are all in good agreement with experimental data, $B_{22} \approx 0.0$ to $1.0 \times 10^{-4} \text{ mol} \cdot \text{ml/g}^2$, extrapolated to pH = 7 and 100 mM salt concentration.⁴²

To choose among the different surface-exposure models A-F, we calculate the binding affinity for the complex of ubiquitin and the UIM1 domain of Vps27 (UIM/Ub)⁴³ (PDB code: 1Q0W). Ubiquitin is a small protein (76 residues) that binds weakly ($K_d \sim 5$ to $500 \mu\text{M}$) to various ubiquitin-binding domains. It serves as a protein tag in numerous biological processes. UIM1 is a ubiquitin binding, helical domain of the yeast class E protein Vps27 and its mammalian homolog, hepatocyte receptor tyrosine kinase substrate (Hrs). Vps27 and Hrs are required for protein sorting in the multivesicular body (MVB) sorting pathway.²⁴

REMC simulations are performed on the UIM/Ub complex for varying system sizes, $R = 60$ to $160 \text{ \AA}$. The system is well equilibrated, switching regularly between low energy states and high energy states. The energy of the system exhibits a bimodal distribution, which indicates the existence of two dominant states, i.e., a bound state with lower energy and an unbound state with higher energy. As the temperature decreases, the population of the bound state increases.

To estimate the binding affinity of the UIM/Ub complex, the fraction of bound states is calculated as a function of the protein concentration. Note that the protein concentration is equal to the inverse of the system volume. The bound fraction is then fitted to a titration curve, equation (18), as shown in Figure 2(a). The binding affinities K_d obtained from the fits are listed in Table 2 for models A-F.

The calculated binding affinities of the UIM/Ub complex are within a factor of four of the experimentally measured value $K_d \approx 280 \mu\text{M}$,⁴³ except for model D in which inner residues are least weighted. As seen in Table 2, the binding affinity of the UIM/Ub complex increases with decreasing weight of the inner residues. Model C yields a K_d in excellent agreement with experiment and will be used exclusively in the following calculations, except if noted otherwise.

To illustrate the “calorimetric” estimate of the binding affinity of the UIM/Ub complex, Figure 2(b) presents the excess heat capacity $C = k_B T^2 \langle (U_{\text{tot}} - \langle U_{\text{tot}} \rangle)^2 \rangle$ calculated as an ensemble average of the fluctuations in the protein-protein interaction energy U_{tot} . Plotted as a function of the protein concentration, the heat capacity exhibits a maximum at a concentration equal to K_d within statistical uncertainties. This result can be understood by noticing that at K_d the two modes of the binding energy distribution have about equal population, such that the variance

is largest. The maximum in the heat capacity thus provides an alternative, “titration calorimetry” estimate of K_d .

Validation of potential models: binding affinity and structure

In the following, we will validate our potential models by calculating the binding affinities and structures of a number of protein complexes. In particular, we will study the complex of yeast cytochrome *c* and cytochrome *c* peroxidase (Cc/CcP; PDB code: 2PCC), which serves as a paradigm of a low-affinity protein complex.⁴⁴ Cytochrome *c* peroxidase is a heme protein found in mitochondria. It is a biological redox partner of cytochrome *c* and catalyzes the reduction of alkyl hydroperoxides. As shown in Table 2, the estimated binding affinities for models B (620 μM) and C (850 μM) at 150 mM salt concentration are in agreement with experimental values, $K_d \approx 100 \mu\text{M}$ and $K_d \approx 300 \mu\text{M}$ to 1000 μM at 100 mM and 200 mM salt, respectively.^{45,46} For the other models, the Cc/CcP complex exhibits weaker affinities between 1.5 and 2.0 mM (Figure 2). These results provide further evidence for model C being a well-balanced potential function.

Structural information on protein complexes can provide key insights into their biological function. To assess the quality of the potential models in predicting complex structures, we calculate DRMS values (see Models and Methods) for each configuration as a measure of distance from the experimental complex structure. We note that complex-structure information had not been used in the above parametrization.

Figure 3 shows scatter plots of the DRMS versus potential energy for the UIM/Ub and Cc/CcP complexes simulated for model C. Similar results are obtained for the other models. For both complexes, the low-energy structures exhibit small DRMS values, $\sim 2 \text{ \AA}$ for the UIM/Ub complex and $\sim 4 \text{ \AA}$ for the Cc/CcP complex. The higher DRMS values for the Cc/CcP complex arise from the fact that for large proteins a slight orientational change results in a comparably large DRMS increase.

The bound structures for the complexes are clustered (see Models and Methods) by using a cutoff distance of 3 $\text{\AA}$ for the UIM/Ub complex and 10 $\text{\AA}$ for the Cc/CcP complex. The higher cutoff value for the Cc/CcP complex was used to account for the relatively larger DRMS values. In particular, structures with the same binding pocket on cytochrome *c* peroxidase but small differences in the orientation of the cytochrome *c* cause large differences in the distance matrix.

The results for the five top-ranked clusters are presented in Figure 4. For each complex, the largest cluster captures the correct binding interface on both complex partners, with populations $> 40\%$ and low DRMS values. All native interfacial residues of the ubiquitin and UIM1 are correctly identified within the lowest-energy structure, as well as $\sim 100\%$ of the interfacial contacts. For the UIM/Ub complex, the second-largest cluster has a population of $\sim 20\%$ and a DRMS of $\sim 5 \text{ \AA}$. This cluster contains structures that have the UIM docked into the same binding pocket of the ubiquitin, but with the helix orientation inverted (Figure 5(b)). Such “inverted UIM” (IUIM) domains have indeed been observed experimentally.⁴⁷

Figure 5(c) compares the structures predicted for the Cc/CcP complex with the crystallographic structure. Three structures representative of the top three clusters are shown in green, with cytochrome *c* peroxidase shown in blue. For comparison, the experimental structure of cytochrome *c* is shown in red. The cluster with the highest population and a DRMS close to 4 $\text{\AA}$ yields the correct binding interface on both proteins (top right of Figures 5(c) and 5(d)). The fraction of native interfacial contacts, $Q_{\text{Cc/CcP}}$, present in the structure is about 0.43, while the fractions, Q_{CcP} and Q_{Cc} , of native interfacial residues present in the interface are 0.97 for the cytochrome *c* peroxidase and 0.78 for the cytochrome *c*, respectively. The fact that the structure has $Q_{\text{Cc}} \approx Q_{\text{CcP}} > Q_{\text{Cc/CcP}}$ is caused by a slight change in the the docked orientation of

cytochrome *c* relative to that in the experimental structure. We note, however, that the extent of the structural agreement is remarkably good if one considers that the Cc/CcP interface is extensively hydrated, with more than 20 water sites in the crystal interface.⁴⁴

Clusters 2 and 3 (bottom right and left of Figure 5(c), respectively), in contrast, have DRMS values ~ 12 Å and ~ 17 Å, respectively. However, these structures have higher energies (about $2 k_B T$ higher than the lowest-energy structure) and are considerably less populated than those with the correct binding site. The possibility of alternative binding modes is already suggested by the original Cc/CcP co-crystal,⁴⁴ in which an additional Cc/CcP interface is present. Moreover, upon mutations of surface residues, Kang and Crane identified additional binding modes in Cc/CcP co-crystals.⁴⁸ While those binding sites do not agree closely with those found here, these experimental results suggest that in complexes with relatively weak binding affinities the specific structure may compete with other binding modes. For other proteins, paramagnetic NMR has recently been used to demonstrate, at equilibrium, the presence of transient protein encounter complexes⁴⁹ and to determine their structures. Such non-specific complexes may be relevant for function.⁴⁹⁻⁵²

Comparison of different contact potentials

For the potential energy functions of our model we have adapted the contact potentials derived from Miyazawa and Jernigan (MJ).²⁷ To assess the sensitivity of our models to different contact potentials, we also use those obtained by Skolnick et al. (SK),³⁰ Betancourt and Thirumalai (BT),³¹ and a more recent re-parametrization by Miyazawa and Jernigan (MJ1999).³² For each set of contact potentials, the two free parameters, λ and e_0 , are fitted against the second virial coefficient of lysozyme and the binding affinity of the ubiquitin and CUE complex. The values for models A and C are presented in Table 3. REMC simulations on the UIM/Ub complex yield comparable binding affinities K_d for all contact potentials as listed in Table 3. Furthermore, more than half of the simulated structures show DRMS values less than 5 Å. For the Cc/CcP complex, the binding affinities, calculated using model A, are similar for the different contact potentials. However, the simulated structures from SK contact potentials exhibit less native-like structures. For model C, MJ1999 contact potentials yield K_d and DRMS similar to that of MJ contact potentials. In contrast, the binding affinities and DRMS values obtained from BT and SK contact potentials show larger deviations from that of MJ contact potentials. Thus for the protein complexes tested, the MJ contact potentials give consistently good results, and hence will be used exclusively in the following simulations, except if stated otherwise.

Unbound docking

To determine the sensitivity of our models to the configuration of the complex partners, we have simulated the UIM/Ub and Cc/CcP complexes by using coordinates from independently determined X-ray and NMR structures of unbound proteins. As shown in Table 4, the resulting binding affinities and structures are essentially the same as those obtained for bound structures. The reason for the slightly larger DRMS value (2.1 vs 1.1 Å) and smaller Q value ($Q_{AB} = 0.8$ vs. 0.97) for unbound docking simulations of the UIM/Ub complex (1UBQ/1Q0V) is that the helical UIM1 domain is rather flexible, with a C_α root mean square deviation (RMSD) of 4.6 Å between bound and unbound states. These results indicate that the predictions of the model are relatively insensitive to small differences in back-bone conformations, reflecting the coarse-grained character of the structure representation.

Complexes of ubiquitin/ubiquitin-binding domains and other proteins

In the following, we apply our model to complexes between ubiquitin [or ubiquitin-like domain (Ubl)] and a series of ubiquitin-binding domains (UBD).⁵³⁻⁶¹ Other weakly interacting protein complexes with $k_d > 0.01 \mu\text{M}$ are also studied.⁶²⁻⁶⁹ Characteristics of these complexes

are listed in Table 4, including experimental and calculated binding affinities. Figure 6 presents a comparison between the experimental and simulated binding affinities obtained by using MJ and MJ1999 contact potentials. For reference, we also include the experimental results for single-site mutants of several complexes between ubiquitin and ubiquitin-binding domains. The experimental K_d values cover about four orders of magnitude, ranging from 0.014 μM to 280 μM . We find that for most complexes with $K_d^{\text{exp}} > 1\mu\text{M}$ the estimated affinities are in good agreement with experimental values, deviating by less than a factor of four (corresponding to an error in the absolute protein interaction energy of less than $k_B T \ln 4 \approx 1.4 k_B T \approx 0.84$ kcal/mol). Among Ub(UbL)/UBD complexes, the calculated and measured affinities of the complex between ubiquitin and UBA of Dsk2 (PDB code: 1WR1) show the largest discrepancy with the calculated binding affinity of 430 μM being about 30-fold lower than the experimental result, 14.8 μM . In mutagenesis experiments on this complex, a single mutation of a Dsk2-UBA residue in the ubiquitin-binding region (Gly-343, Phe-344, and Leu-369) lowered the affinity more than 25-fold, bringing it in line with our estimate as shown in Figure 6.⁵⁴ This result suggests that specific atomic-level interfacial contacts are essential for the tighter binding of Dsk2-UBA to ubiquitin. Such specific interactions are beyond the resolution limit of the coarse-grained models used here. For complexes of ubiquitin and S5a UIM (PDB codes: 1UEL and 1P9D), similar results are obtained for MJ contact potentials, as shown in Table 4.

For the non-ubiquitin protein complexes, the calculated and measured binding affinities agree within a factor of ~ 10 (or about 1.5 kcal/mol in free energy) in $\sim 50\%$ of the cases. As expected, we find the largest deviations for proteins with high binding affinities (i.e., $K_d^{\text{exp}} < 1\mu\text{M}$). For such strongly interacting complexes atomic-level interactions play an important role in binding, which is beyond the scope of the coarse-grained model and potential energy functions developed in this paper.

To illustrate that the coarse-grained model can also predict correct bound configurations of the complexes, we compute the DRMS values for the simulated structures. For most complexes, more than 50 % of the simulated structures yield DRMS values less than 5 Å with the exception of Ub-Dsk2 UBA (1WR1: 38 %, 2BWE: 24 %) and Ub-GAT (1YD8: 25 %).

To further assess the quality of simulated structures, Table 4 lists the DRMS values and the fractions of native contacts and native interfacial residues present in the lowest-energy structures. Also listed are the corresponding means, obtained by averaging over the ensemble of bound structures. The lowest-energy structures yield lower DRMS and higher Q values than the averages. In most cases DRMS values lie below 5 Å, although the simulated structures for several complexes (3HHR, 1MEL, 1MLC, 1WEJ, 1MDA, 1GLA, 1AVZ) yield large DRMS values ($\gtrsim 9$ Å). For some complexes (1Q0W, 2G3Q, 1P3Q, 2D3G, 2FID, 1UEL, 1P9D, 1EFU, 1FIN) more than 50 % of native contacts are correctly identified in the lowest-energy structures. Among the complexes between the ubiquitin and ubiquitin-binding domains, the binding interfaces of the two domains are accurately predicted. In all cases, even for the complexes with large DRMS values, the binding interface of at least one protein is correctly identified. We conclude from these results that further improvements of the quality of bound structures will require higher-resolution models and energy functions.

Vps27/Ub complex

To illustrate applications to flexible multi-protein complexes, we study the interactions of Vps27 and a membrane-anchored ubiquitin.²⁴ The protein Vps27, which is a subsystem of the Vps27/Hse1 complex, is required for protein sorting in the MVB pathway. This MVB system plays a key role in a number of biological processes, such as receptor down-regulation, cell growth, antigen presentation, and the budding of certain viruses from host cells. The Vps27 protein consists of several folded domains, including VHS, FYVE and two UIMs (UIM1 and

UIM2) as shown in Figure 7(a). The FYVE domain is known to bind to the membrane through strong interactions with the endosomally enriched lipid phosphatidylinositol 3-phosphate (PI(3)P). The UIMs bind to membrane-tethered ubiquitins. The membrane-bound Vps27 then recruits the first protein complex of the endosomal sorting complex for transport (ESCRT) machinery (ESCRT-I), and thereby initiates the MVB sorting reaction.^{24,70} Although the interactions between individual UIMs and ubiquitin are well studied experimentally, less is known how the various domains of the whole protein associate with the membrane-tethered ubiquitins and other domains.⁷¹

As a demonstration of our coarse-grained models in studies of large multi-protein complexes, we simulate the Vps27-ubiquitin complex in the presence of a membrane. To mimic the interaction between the FYVE domain and PI(3)P, we add a harmonic potential between the PI(3)P binding pocket of the FYVE domain and the planar membrane. This interaction restricts the movement of the FYVE domain away from the membrane and allows movement only parallel to the membrane with flexible rotational motion. Ubiquitin is tethered to the membrane via a 12-residue-long flexible linker.

To measure the binding affinities of membrane-tethered ubiquitin to various domains of Vps27, we calculate the probabilities of the ubiquitin being bound to the respective domains. For comparison, we also simulate the complex of membrane-tethered ubiquitin and a fragment of Vps27 that contains the FYVE and UIM1 domains connected by a flexible linker. Note that the simulations of an isolated UIM/Ub complex in solution have yielded a binding affinity of 270 μM close to the experimental value (Table 2). As shown in Figure 7(b) we find that the binding affinity of Ub-UIM1 in the presence of the full Vps27 is slightly *higher* than that of the fragmented Vps27, despite the competition by other domains (UIM2 and VHS) for ubiquitin binding. In contrast, the binding affinities of Ub-UIM2 and Ub-VHS are slightly lower than that of Ub-UIM1.

In combination, the overall Vps27-ubiquitin binding affinity is greatly enhanced over those of the individual domains, as shown in Figure 7(b). We find that the probability of ubiquitin being bound to one of the Vps27 domains is about five times higher than that of ubiquitin-UIM1 binding alone at low ubiquitin concentrations. This result suggests positive cooperativity, in which non-specific ubiquitin binding to the non-UIM1 domains enhances the overall binding affinity of the complex without interfering with specific ubiquitin-UIM1 binding.

Interestingly, we also observe frequent interactions between ubiquitin and the VHS domain of Vps27. Figure 7(c) presents the distances between the centers of mass of the three ubiquitin-interacting domains, VHS, UIM1 and UIM2, and the membrane-tethered ubiquitin. The binding of a VHS domain to ubiquitin has been observed in several proteins.^{72,73} In contrast, the interaction between the VHS domain of Vps27 and ubiquitin has so far not been established experimentally. To explore their interaction further, we performed simulations also of isolated Vps27 VHS and ubiquitin complexes. The results of those simulations, $K_d \approx 50 \mu\text{M}$, indicate that their binding may be stronger than that between UIM1 and ubiquitin. It would be interesting to see whether this interaction can be detected experimentally.

Figure 8 show representative structures of the Vps27-ubiquitin complex. We find that the structures of the isolated ubiquitin-UIM complexes are retained in the context of the Vps27-ubiquitin complex. Specifically, the pairwise interactions of ubiquitin with UIM1 and UIM2 produce structures in excellent agreement with the respective experimental structures, with DRMS values $< 2 \text{ \AA}$. Furthermore, the simulations suggest a highly dynamic conformational equilibrium within the Vps27-ubiquitin complex, as illustrated in Figure 8. We find that the UIM1 domain is the major binding partner for ubiquitin. Binding to UIM1 alone accounts for $\sim 32\%$ of the simulated structures (Figure 8(a)). In addition, we find that in $\sim 11\%$ of the

bound structures, ubiquitin interacts simultaneously with both UIM1 and UIM2 (Figure 8(b)). A significant fraction of the bound structures also involves non-specific interactions. In particular, we find interactions between ubiquitin and the VHS domain alone in ~26% of the bound structures (Figure 8(c)), and with both VHS and FYVE domains (~12%).

The observed flexibility in the interactions not only enhances the overall binding affinity to ubiquitin, but also permits binding to a diverse set of targets. Both aspects appear to be central to the proper function of Vps27 in the MVB protein-sorting pathway, which is recruited by a broad range of ubiquitin-tagged proteins.^{24,70,71}

Discussion

The observed agreement of the simulated binding affinities for a series of weakly interacting protein complexes with experiment, the correct identification of the binding interfaces, and the accurate prediction of bound structures (DRMS of 2-5 Å) in many cases indicate that the potential energy functions are transferable at least between weakly bound protein complexes. Even for the complex of ubiquitin and Dsk2 UBA (1WR1), with the largest observed discrepancy between the calculated and measured binding affinities, the 30-fold difference in K_d amounts to an error of only $k_B T \ln 30 \approx 3.6 k_B T \approx 2.2$ kcal/mol in the binding free energy. Such small differences in the free energy cannot be resolved in general within coarse-grained models that ignore atomic details. Indeed, a single mutation brings the experimental affinity down to values close to the calculated one. Note that the formation of a single hydrogen bond is expected to contribute $\gtrsim 1 k_B T$ to the energy.

The ability of the coarse-grained model to provide semi-quantitative binding affinities and to predict correct bound structures for the protein complexes may rest on the fact that the proteins tested here interact relatively weakly with their binding partners. For such weakly bound protein complexes, atomic-level interactions may not be as important as for strongly bound complexes. Conversely, for strongly bound protein complexes ($K_d < 1 \mu\text{M}$) the coarse-grained models and the potential energy functions developed here are not expected to be as successful in estimating binding affinities as demonstrated above. However, even for such complexes the models can be expected to predict correct binding sites, in part because strong binders may have evolved from weak binders while maintaining the original binding interface.

The comparison between different contact potentials shows that the coarse-grained models are not particularly sensitive to the details of the potentials. As a further simplification for the potential energy functions, one may consider a uniform attractive interaction between all pairs of residues. However, such potentials give size-dependent protein-protein interactions, yielding stronger attraction for larger protein complexes. Furthermore, simulations of several protein complexes show bound structures deviating strongly from native ones. This result suggests that although details of residue-specific interactions between amino acid residues are not crucial in protein-protein interactions, the character of these interactions, such as repulsive or attractive interactions between pairs of polar-polar, polar-apolar and apolar-apolar residues, is important to obtain accurate protein-protein interactions.

A natural question is why the relatively simple models studied in this paper work so well in predicting the structures and affinities of weakly-bound protein complexes. A possible answer lies in a recent observation that the landscapes for functional protein binding are relatively smooth, guiding the binding partners toward the specific complex.⁷⁴ The relative insensitivity of the results to the details of the interaction energies used would then be a reflection of a “funneled” landscape for protein-protein interactions. Indeed, one may expect that such a global characteristic of the landscape should be most apparent for the weakly bound complexes studied here, whose structure is determined primarily by many spatially diffuse and weak

interactions, not the strong and specific ones of tight complexes. But clearly, exploring the character of the binding landscape deserves further attention, as it is important both practically to find correctly bound structures,³⁹ and fundamentally to understand the connection between amino-acid sequence and function in proteomics studies.^{75,76}

To improve the coarse-grained models significantly will require atomic-level information of protein-protein interactions. Such improvements can be achieved, for instance, by adopting a multi-scale approach. A major challenge for high resolution all-atom models are the prohibitively large configurational space. The coarse-grained models can greatly reduce the search space by providing good candidates for all-atom Molecular Dynamics (MD) simulations. Such all-atom MD simulations can provide further information on atomic-level protein-protein interactions, which then permit refinements of the coarse-grained models and the effective energy functions. Protein flexibility, as a major factor in high-affinity binding,⁷⁷ is then considered at a higher level of resolution.

The main objective here is to develop tools for the study of large multi-protein complexes in which protein domains can be connected by peptide linkers and may be tethered to a membrane. We have shown that we can efficiently sample the large configuration space of the Vps27 complex involved in MVB protein sorting. In an earlier work,⁷¹ we used similar models to study the interactions between the full Vps27/Hse1 complex and three membrane-tethered ubiquitins (~ 1080 residues). The simulations showed that the complex is open, with a radius of gyration $R_g \simeq 50$ Å, and highly dynamic. The overall binding affinities of different UIMs to the membrane-tethered ubiquitins were shown to remain close to those of isolated ones, resulting in additive interactions that can explain how multiple weak Ub/UIM interactions can recruit the Vps27 complex to the membrane.

The models developed here can easily incorporate additional experimental information (e.g., from fluorescence, scattering, NMR, electron microscopy). In simulations of the Vps27/Ub complex some experimental information has already been incorporated by adding a harmonic constraint between the FYVE domain and the membrane. Further experimental information regarding interactions between various domains can be added in the simulations via additional constraints, which will enhance the ability of our coarse-grained models to capture established aspects of protein-protein interactions.

Conclusions

Coarse-grained models for simulations of weakly bound multi-protein complexes were developed, validated, and applied. To parametrize a coarse-grained energy function, we used experimental data for lysozyme second virial coefficients and ubiquitin-CUE binding affinities. The resulting models were applied to a series of complexes between ubiquitin and ubiquitin-binding domains, the complex of cytochrome *c* and cytochrome *c* peroxidase, other protein complexes, and finally the membrane-attached Vps27 domain of the MVB pathway. Using REMC simulations, we found that the binding affinities estimated from the ensemble of simulated structures agree quantitatively with experiment (in most cases within a factor of four, corresponding to an error in the overall free energy of less than $\sim 1.5 k_B T \simeq 1$ kcal/mol). In $\sim 90\%$ of the cases both binding interfaces are correctly identified; in the other cases, at least one of the interfaces is correct. Moreover, the predicted complex structures agree with the experimental structures to within 2 to 5 Å DRMS for all ubiquitin complexes studied, and about half the other complexes. In particular, the dominant cluster of complexes contains native-like structures and has a population exceeding $\sim 40\%$. For the UIM/Ub complex, we found in addition a cluster with about 20 % population, which contains structures that occupy the same binding pocket in the ubiquitin but with an inverted orientation of the UIM1. Such inverted UIMs have indeed been seen experimentally.⁴⁷ These results combined suggest that the coarse-

grained model and the energy function should be transferable to other protein complexes with intermediate to low binding affinities.

In an illustrative application to a multi-protein complex, we used the coarse-grained model to study the interaction of Vps27 with membrane-tethered ubiquitin. We found a dynamic complex structure in which ubiquitin interacts frequently with the UIM1, UIM2 and VHS domains of Vps27. It was shown that the competition between the various domains for ubiquitin does not lower the overall binding probability to the UIM1 domain. Instead, binding to UIM1 is even slightly enhanced in the context of the full Vps27 relative to isolated UIM1. These results suggest that the binding of Vps27 to ubiquitin derives part of its strength from positive cooperativity between non-specific ubiquitin interactions, in addition to the specific interactions of UIM1 and UIM2 with ubiquitin. The observed flexibility at the structural level likely plays a role not only in enhancing the binding affinity, but also in facilitating binding to the diverse set of targets in the MVB protein-sorting pathway.^{24,70,71}

Models and Methods

Models

A main objective here is to build a model for simulating large multi-protein complexes and to identify and distinguish various protein-protein interactions. Several requirements have to be met to achieve such a goal: (i) computational cost-effectiveness and the ability (ii) to predict correct bound structures and (iii) to distinguish between different protein-protein interactions by their selective binding affinities. Here proteins or folded domains are treated as rigid bodies, which greatly enhances the sampling efficiency in configuration space and thus reduces the overall simulation time. All-atom-level potential energy functions, although desirable, increase the cost of simulations, thus becoming prohibitive in simulations of large protein complexes. Therefore the interactions between two proteins are specified here at the residue-level. Each amino acid residue of a protein is represented by a spherical bead centered at the corresponding C_α atom. The interaction potential, $\phi_{ij}(r)$, between two residues, i and j , then consists of short-range Lennard-Jones (LJ) type potentials and long-range electrostatic interactions:

$$\phi_{ij}(r) = u_{ij}(r) + u_{ij}^{\text{el}}(r), \quad (1)$$

where r is the distance between the C_α atoms of the two residues.

Depending on the amino acid types of the pair ij , the interactions, beyond a short-range steric repulsion, are either attractive or long-range repulsive. Repulsive interactions are used for amino acid pairs that interact less favorably with each other than with the solvent. If a pair of residues experiences an attractive interaction with a strength $\epsilon_{ij} < 0$, we use the usual LJ potential form given by

$$u_{ij}(r) = 4|\epsilon_{ij}| \left[\left(\sigma_{ij}/r \right)^{12} - \left(\sigma_{ij}/r \right)^6 \right], \quad (\epsilon_{ij} < 0), \quad (2)$$

where σ_{ij} are the residue-dependent interaction radii. For pairs of residues repelling each other, i.e., $\epsilon_{ij} > 0$, the potential energy function is given by

$$u_{ij}(r) = \begin{cases} 4\epsilon_{ij} \left[\left(\sigma_{ij}/r \right)^{12} - \left(\sigma_{ij}/r \right)^6 \right] + 2\epsilon_{ij}, & \text{if } r < r_{ij}^0, \\ -4\epsilon_{ij} \left[\left(\sigma_{ij}/r \right)^{12} - \left(\sigma_{ij}/r \right)^6 \right], & \text{if } r \geq r_{ij}^0, \end{cases} \quad (3)$$

where $r_{ij}^0 = 2^{1/6}\sigma_{ij}$. At contact, this form of the energy function yields equal contributions from attractive ($\epsilon_{ij} < 0$) and repulsive ($\epsilon_{ij} > 0$) pairs, but with opposite sign, $u_{ij}(r_{ij}^0) = \epsilon_{ij}$, as illustrated in Figure 9(a). Explicit values for ϵ_{ij} will be discussed below.

The long-range electrostatic interaction is given by a simple Debye-Hückel-type potential:

$$u_{ij}^{\text{el}}(r) = q_i q_j \exp(-r/\xi) / 4\pi D r, \quad (4)$$

where q_i is the charge of residue i , located at the center of the corresponding bead, ξ is the Debye screening length, and D is the dielectric constant of the solvent medium. In this study we set $D = 80$, the dielectric constant of water at nearambient conditions. The Debye screening length ξ depends on the salt concentration. For simplicity, we focus on physiological salt concentrations of ~ 100 mM, which lead to $\xi \approx 10$ Å. Residue charges correspond to pH = 7 such that $q_i = +e$ for Lys and Arg, $-e$ for Asp and Glu, and $+0.5e$ for His, where e is the elementary charge.

The LJ interaction strengths ϵ_{ij} are adapted from the knowledge-based statistical contact potentials obtained by Miyazawa and Jernigan (MJ).²⁷ Here, we need to scale the contact potentials relative to the known electrostatic interactions. Specifically, we set

$$\epsilon_{ij} = \lambda (e_{ij} - e_0), \quad (5)$$

where $e_{ij} (< 0)$ is the MJ contact potential between residues i and j ²⁷ while e_0 is an offset parameter that balances the preference of residue-residue interactions relative to residue-solvent interactions. The parameter λ scales the strength of the LJ interactions compared to the physical electrostatic interactions. The two free parameters λ and e_0 are determined by fitting against experimental data. For comparison, contact potentials developed by Skolnick et al. (SK),³⁰ Betancourt and Thirumalai (BT),³¹ and Miyazawa and Jernigan 1999 (MJ1999)³² are used in fitting the potential energy functions and simulating protein complexes. Other contact potentials^{26,29,34,78} may also be adapted in the potential energy functions.

The interaction radii, σ_{ij} , are given by

$$\sigma_{ij} = (\sigma_i + \sigma_j) / 2, \quad (6)$$

where σ_i is the van der Waals diameter of a residue i , which is calculated from the van der Waals volume⁷⁹ by assuming a spherical shape for the residue (see Table 5).

To account for the solvation effects on the potential energy, the pairwise interaction $\phi_{ij}(r)$ between two residues i and j is weighted according to the degree of the solvent accessible surface area (SASA) for residue pairs. The total potential energy, U_{tot} , is then given by

$$U_{\text{tot}} = \sum_{i,j} f_i f_j \phi_{ij}(r_{ij}), \quad (7)$$

where r_{ij} is the distance between two residues i and j , while $0 \leq f_i \leq 1$ is a weight factor for a residue i . The simplest case will be $f_i = 1$ for any i so that all residues are equally weighted, independent of their location. However, we also consider models in which inner residues with low SASAs have smaller effects on the overall protein-protein interaction than the outer ones with high SASAs. In these models we set $f_i = f(s_i)$ where $0 \leq s_i \leq 1$ is the relative SASA for a residue i compared to the value for a random-coil configuration.⁸⁰ Here we consider six models for $f(s)$:

$$\begin{aligned} (A) f(s) &= 1 \\ (B) f(s) &= \tanh(10 - \tan(\pi s/2)), \\ (C) f(s) &= \tanh(5 - \tan(\pi s/2)), \\ (D) f(s) &= \tanh(2 - \tan(\pi s/2)), \\ (E) f(s) &= (1 + \tanh(2 - \tan(\pi s/2))) / 2, \\ (F) f(s) &= (1 + \tanh(\tan(\pi s/2))) / 2 \end{aligned} \quad (8)$$

The shapes of these weighting functions, $f(s)$, are shown in Figure 9(b). Note that the hyperbolic functional form for $f(s)$ makes the weighting factor insensitive to outer residues with $s \gtrsim 0.5$.

Flexible linkers

Flexible linkers are modeled as bead polymers. Besides the LJ-type potentials and the electrostatic potentials between residue pairs separated by more than three residues, $|i - j| \geq 4$, the interactions for linker molecules contain additional contributions from bond-stretching potentials for pseudo bonds $C_\alpha-C_\alpha$, angle potentials for pseudo angles $C_\alpha-C_\alpha-C_\alpha$, and torsion-angle potentials for pseudo torsions $C_\alpha-C_\alpha-C_\alpha$. For the bond-stretching potential we use the harmonic potential,

$$E_{\text{bond}}(r) = \frac{1}{2}k(r - r_0)^2, \quad (9)$$

where r is the $C_\alpha-C_\alpha$ distance, r_0 is the reference $C_\alpha-C_\alpha$ distance, and k is the spring constant. Here we set $r_0 = 3.81 \text{ \AA}$, the average length of the $C_\alpha-C_\alpha$ pseudo-bond,⁸¹ and $k = 378 \text{ kcal/(mol} \cdot \text{\AA}^2)$ taken from Karanicolas and Brooks.⁸² The $C_\alpha-C_\alpha-C_\alpha$ pseudo-angles are described by⁸¹

$$\exp[-\gamma E_{\text{angle}}(\theta)] = \exp\left[-\gamma(k_\alpha(\theta - \theta_\alpha)^2 + \epsilon_\alpha)\right] + \exp\left[-\gamma k_\beta(\theta - \theta_\beta)^2\right], \quad (10)$$

where the constants are $\gamma = 0.1 \text{ mol/kcal}$, $\epsilon_\alpha = 4.3 \text{ kcal/mol}$, $\theta_\alpha = 1.60 \text{ rad}$, $\theta_\beta = 2.27 \text{ rad}$, $k_\alpha = 106.4 \text{ kcal/(mol} \cdot \text{rad}^2)$, and $k_\beta = 26.3 \text{ kcal/(mol} \cdot \text{rad}^2)$. This double-well potential accounts for both helical and extended $C_\alpha-C_\alpha-C_\alpha$ pseudo angles. Finally, the torsion-angle potential is expressed in the form

$$E_{\text{torsion}}(\varphi) = \sum_{n=1}^4 [1 + \cos(n\varphi - \delta_n)] V_n. \quad (11)$$

The constants, V_n and δ_n ($n = 1, 2, 3, 4$), depend only on the sequence of the middle two residues, X-A-A-X, and are taken from Karanicolas and Brooks.⁸²

Membrane interactions

Short-range interaction potentials between amino acids and planar membranes are taken from the potentials of mean force derived from integral membrane protein structures (see Figure 6 and Table 1 of Ref. 35). To prevent proteins from penetrating the membrane, a sharp repulsive potential, $0.05(z_0/z)^{12}$, is added, where z is the distance from the membrane center to the amino acid and $z_0 = 20 \text{ \AA}$.

Electrostatic interactions between a negatively charged membrane and charged residues are included by using the Guoy-Chapman formula. The membrane electrostatic potential at a distance z from the membrane surface is given by⁸³

$$\psi(z) = \frac{2k_B T}{e} \ln \frac{[1 + \alpha \exp(-\kappa z)]}{[1 - \alpha \exp(-\kappa z)]}, \quad (12)$$

where $\alpha = (\exp[e\psi(0)/2k_B T] - 1)/(\exp[e\psi(0)/2k_B T] + 1)$ and κ is the inverse Debye length. For the surface potential $\psi(0)$, we assume $e\psi(0) = -30 \text{ mV}$.⁸³ The total electrostatic interaction between a membrane and a protein is then given by

$$E_{\text{mem}}^{\text{el}} = \sum_i q_i \psi(z_i), \quad (13)$$

where q_i is the charge of a residue i and z_i the distance between the membrane and the residue.

Simulation methods

Structures of protein complexes are taken from the Protein Data Bank, using both bound and unbound configurations. The proteins are confined into a sphere of radius R with an impenetrable hard wall to prevent proteins from escaping. Each protein is allowed to move both translationally and rotationally. One may also consider a conventional cubic box with periodic boundary conditions but the results remain essentially unchanged. If a membrane is present, we consider only a hemisphere with the membrane located at the center plane. By varying the radius R , one effectively changes the concentration of molecules, which is inversely proportional to the volume of the sphere, $4\pi R^3/3$.

To obtain equilibrium properties for protein-protein interactions we perform Monte Carlo (MC) simulations for different sizes of the sphere using U_{tot} of equation 7 as potential energy. For systems that contain flexible linkers or membranes, additional interactions are included as described above. To enhance the equilibrium sampling, the REMC method is adopted⁸⁴ with replicas at twenty different temperatures ranging from 40 percent above room temperature, $T = 300$ K, to 20 percent below. Specifically, we use $\beta^* \equiv T/T_i = 0.6, 0.64, 0.68, \dots, 1.15, 1.175, 1.2$. Observed acceptance ratios for the replica exchange lie between 40 to 70 %. Basic MC steps are rigid body translational and rotational moves on each protein or folded domain. For flexible linker molecules, in addition to local MC moves on each residue, crank shaft moves are employed to enhance sampling.⁸⁵ For MC translational and rotational movements, we use step sizes of 0.5 Å and 0.2 rad, respectively, which yield acceptance ratios of 40 to 60 % at room temperature. After initial equilibration amounting to $\sim 10^4$ to 10^5 MC steps, on the order of 10^8 to 10^9 MC steps are performed for data acquisition. We consider 10^4 to 10^5 configurations chosen every 10^3 to 10^4 MC steps for analysis.

Osmotic virial coefficients

To parameterize the energy functions, we use osmotic virial coefficient data. This procedure allows us to find energy-function parameters relevant for the regime of weak protein-protein interactions. We calculate the osmotic second virial coefficient, $B_{22} \equiv B'_{22}/M_w^2$, defined by

$$B'_{22} = -\frac{1}{16\pi^2} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^\infty \left[e^{-U_{\text{tot}}/k_B T} - 1 \right] r^2 dr \sin \theta d\theta d\varphi d\alpha \sin \beta d\beta d\gamma, \quad (14)$$

where M_w is the protein molecular weight, while $U_{\text{tot}}(r, \theta, \varphi, \alpha, \beta, \gamma)$ is the total potential energy between two protein molecules given by equation (7). Here r is the distance between the centers of the two proteins, θ and φ are the orientation angles for the first molecule, while α, β and γ are the Euler angles that specify the orientation of the second molecule relative to the first one. MC integration was used to compute the five-dimensional orientation integral in equation (14). For a given set of the angles $(\theta, \varphi, \alpha, \beta, \gamma)$, the radial integration can be written as

$$I_{in} \equiv \int_0^\infty \left[e^{-U_{\text{tot}}/k_B T} - 1 \right] r^2 dr = -\frac{r_c^3}{3} + \int_{r_c}^\infty \left[e^{-U_{\text{tot}}/k_B T} - 1 \right] r^2 dr, \quad (15)$$

where r_c is a cutoff distance chosen such that $U_{\text{tot}} \geq 40k_B T$ for $r < r_c$. The remaining integration was performed by using Simpson's rule with an interval $dr = 0.05$ Å up to $10r_0$ where r_0 is the diameter of the protein. Note that B_{22} measures the overall strength for protein pairwise interactions (repulsive for positive B_{22} and attractive for negative B_{22}) and has been related to protein crystallization conditions.^{40,86,87}

Binding affinity

To determine the binding affinity measured by the dissociation constant, K_d , for a protein complex, we consider two-state kinetics for the formation of a complex AB from two components A and B at equilibrium, which may be expressed as

where k_a is the second-order rate constant for the bimolecular association reaction while k_d is the first-order rate constant for the monomolecular dissociation. The ratio between the two rate constants yields the equilibrium association constant K_a given by

$$K_a = \frac{k_a}{k_d} = \frac{[AB]}{[A][B]}, \quad (17)$$

where $[A]$ denotes the concentration of the molecule A. Note that K_a has units of inverse concentration. It is often intuitively easier to consider the dissociation constant, $K_d = 1/K_a$, which has units of concentration. Here $[A] = [B]$, such that the fraction of complex AB formed, y , is given by

$$y = \frac{[AB]}{[A] + [AB]} = \frac{[A]}{[A] + K_d}. \quad (18)$$

Note that when a protein concentration is equal to K_d , half the protein molecules form complexes.

From the simulations, conducted as described above with one A and one B molecule in the sphere, we calculate K_d using geometric, energetic, and calorimetric approaches: by counting a configuration as bound (1) if at least one pair of residues from the two proteins is within a distance of 8 Å, or (2) if the protein pair interaction energy is less than $-2 k_B T$; and (3) by using a “calorimetric” definition in which K_d is identified as the protein concentration at which the specific heat has a maximum. For the binary complexes studied here, we find that all three definitions give essentially the same binding affinities.

Scoring and ranking

Complex structures obtained from MC simulations are clustered according to the distance matrix $[d_{ij}]$ where d_{ij} is the distance between two residues i and j from different proteins. Two structures are assigned to the same cluster if for all pairs (i, j) the difference between distances is less than a cutoff value, $|d_{ij} - d'_{ij}| < d_{\text{cutoff}}$. A simple clustering algorithm is adapted to group the simulated bound structures of protein complexes. Clusters are then ranked according to their population and compared to the native structures by using the DRMS values defined by

$$\text{DRMS} = \frac{1}{N} \sum_{i,j} |d_{ij}^{\text{sim}} - d_{ij}^{\text{exp}}|, \quad (19)$$

where N is the total number of pairs (i, j) while d_{ij}^{exp} and d_{ij}^{sim} are the distance matrices from the experimental and simulated structures, respectively. For large protein complexes, the DRMS is a better measure for structure comparison than the RMSD, since small orientational changes of a protein relative to the other can cause large RMSD changes, but only small DRMS variations. For each cluster, the structure with the largest number of neighbors is chosen as a representative

The quality of simulated structures can be further assessed by calculating the fraction of native contacts between proteins A and B, $0 \leq Q_{AB} \leq 1$, present in the simulated structures. A pair of residues, i and j , is considered a native contact if d_{ij}^{exp} is less than 12 Å in the experimental structure. For simulated structures, Q_{AB} is defined as the fraction of native contacts which satisfy $d_{ij}^{\text{sim}} < 1.25 d_{ij}^{\text{exp}}$. We further define Q_A and Q_B as the fractions of native interfacial

residues of proteins A and B, respectively. A residue i of protein A is considered to be in the binding interface if it lies within 12 Å from any residue of protein B.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. James H. Hurley for many stimulating discussions. This work was supported by the Intramural Research Program of the NIH, NIDDK.

References

1. Kleanthous, C. Protein-Protein Recognition. Oxford University Press; Oxford: 2000.
2. Smith GR, Sternberg MJE. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol* 2002;12:28–35. [PubMed: 11839486]
3. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins: Struct. Funct. Genet* 2002;47:409–443. [PubMed: 12001221]
4. Wodak SJ, Janin J. Computer analysis of protein-protein interaction. *J. Mol. Biol* 1978;124:323–342. [PubMed: 712840]
5. Warwicker J. Investigating protein-protein interaction surfaces using a reduced stereochemical and electrostatic model. *J. Mol. Biol* 1989;206:381–395. [PubMed: 2541255]
6. Totrov M, Abagyan R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. *Nature Struct. Biol* 1994;1:259–263. [PubMed: 7656055]
7. Moont G, Gabb HA, Sternberg MJE. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins: Struct. Funct. Genet* 1999;35:364–373. [PubMed: 10328272]
8. Ritchie D, Kemp GJL. Protein docking using spherical polar Fourier correlations. *Proteins: Struct. Funct. Genet* 2000;39:178–194. [PubMed: 10737939]
9. Norel R, Sheinerman F, Petrey D, Honig B. Electrostatic contributions to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci* 2001;10:2147–2161. [PubMed: 11604522]
10. Camacho CJ, Vajda S. Protein docking along smooth association pathways. *Proc. Natl. Acad. Sci. USA* 2001;98:10636–10641. [PubMed: 11517309]
11. Zacharias M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci* 2003;12:1271–1282. [PubMed: 12761398]
12. Chen R, Mintseris J, Janin J, Weng Z. A protein-protein docking benchmark. *Proteins: Struct. Funct. Genet* 2003;52:88–91. [PubMed: 12784372]
13. Schueler-Furman O, Wang C, Baker D. Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins: Struct. Funct. Bioinfo* 2005;60:187–194.
14. Wiehe K, Pierce B, Mintseris J, Tong WW, Anderson R, Chen R, Weng Z. ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. *Proteins: Struct. Funct. Bioinfo* 2005;60:207–213.
15. Zacharias M. ATTRACT: Protein-protein docking in CAPRI using a reduced protein model. *Proteins: Struct. Funct. Bioinfo* 2005;60:252–256.
16. Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. *Proteins: Struct. Funct. Bioinfo* 2005;60:296–301.
17. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ. Prediction of multimolecular assemblies by multiple docking. *J. Mol. Biol* 2005;349:435–447. [PubMed: 15890207]
18. Tobi D, Bahar I. Optimal design of protein docking potentials: efficiency and limitations. *Proteins: Struct. Funct. Bioinfo* 2006;62:970–981.

19. Janin J, Henrick K, Moult J, Eyck L, Sternberg MJE, Vajda S, Vakser I, Wodak SJ. CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins: Struct. Funct. Genet* 2003;52:2–9. [PubMed: 12784359]
20. Misra S, Hurley JH. Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27. *Cell* 1999;97:657–666. [PubMed: 10367894]
21. Hurley JH, Misra S. Signaling and subcellular targeting by membrane-binding domains. *Annu. Rev. Biophys. Biomol. Struct* 2000;29:49–79. [PubMed: 10940243]
22. Katzmann DJ, Babst M, Emr SD. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* 2001;106:145–155. [PubMed: 11511343]
23. Gavin A-C, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;415:141–147. [PubMed: 11805826]
24. Katzmann DJ, Stefan CJ, Babst M, Emr SD. Vps27 recruits ESCRT machinery to endosomes during MVB sorting. *J. Cell Biol* 2003;160:413–423. [PubMed: 12900393]
25. Teo H, Perisic O, González B, Williams RL. ESCRT-II, an endosome-associated complex required for protein sorting: crystal structure and interactions with ESCRT-III and membranes. *Develop. Cell* 2004;7:559–569.
26. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 1985;18:534–552.
27. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term for simulation and threading. *J. Mol. Biol* 1996;256:623–644. [PubMed: 8604144]
28. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol* 1996;257:457–469. [PubMed: 8609636]
29. Thomas PD, Dill KA. An iterative method for extracting energy-like quantities from protein structures. *Proc. Natl. Acad. Sci. USA* 1996;93:11628–11633. [PubMed: 8876187]
30. Skolnick J, Jaroszewski L, Kolinski A, Godzik A. Derivation and testing of pair potentials for protein folding. When is the quasichemical approximation correct? *Protein Sci* 1997;6:676–688. [PubMed: 9070450]
31. Betancourt MR, Thirumalai D. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci* 1999;8:361–369. [PubMed: 10048329]
32. Miyazawa S, Jernigan RL. Self-consistent estimation of interresidue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins: Struct. Funct. Genet* 1999;34:49–68. [PubMed: 10336383]
33. Zhang C, Kim SH. Environment-dependent residue contact energies for proteins. *Proc. Natl. Acad. Sci. USA* 2000;97:2550–2555. [PubMed: 10706611]
34. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins: Struct. Funct. Genet* 2001;43:89–102. [PubMed: 11276079]
35. Ulmschneider MB, Sansom MSP, Di Nola AD. Properties of integral membrane protein structures: Derivation of an implicit membrane potential. *Proteins: Struct. Funct. Bioinfo* 2005;59:252–265.
36. Elcock AH, Sept D, McCammon JA. Computer simulation of protein-protein interactions. *J. Phys. Chem. B* 2001;105:1504–1518.
37. Spaar A, Dammer C, Gabdoulline RR, Wade RC, Helms V. Diffusional encounter of barnase and barstar. *Biophys. J* 2006;90:1913–1924. [PubMed: 16361332]
38. Alsallaq R, Zhou HX. Prediction of protein-protein association rates from a transition-state theory. *Structure* 2007;15:215–224. [PubMed: 17292839]
39. Alsallaq R, Zhou HX. Energy landscape and transition state of protein-protein association. *Biophys. J* 2007;92:1486–1502. [PubMed: 17142273]
40. Neal BL, Asthagiri D, Velez OD, Lenhoff AM, Kaler EW. Why is the osmotic second virial coefficient related to protein crystallization? *J. Cryst. Growth* 1999;196:377–387.

41. Kang RS, Daniels CM, Francis SA, Shih SC, Salerno WJ, Hicke L, Radhakrishnan I. Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. *Cell* 2003;113:621–630. [PubMed: 12787503]
42. Paliwal A, Asthagiri D, Abras D, Lenhoff AM, Paulaitis ME. Light-scattering studies of protein solutions: role of hydration in weak protein-protein interactions. *Biophys. J* 2005;89:1564–1573. [PubMed: 15980182]
43. Swanson KA, Kang RS, Stamenova SD, Hicke L, Radhakrishnan I. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *EMBO J* 2003;22:4597–4606. [PubMed: 12970172]
44. Pelletier H, Kraut J. Crystal structure of a complex between electron transfer partners, cytochrome *c* peroxidase and cytochrome *c*. *Science* 1992;258:1748–1755. [PubMed: 1334573]
45. Erman JE, Vitello LB. The binding of cytochrome *c* peroxidase and ferricytochrome *c*. *J. Biol. Chem* 1980;255:6224–6227. [PubMed: 6248515]
46. Kresheck GC, Vitello LB, Erman JE. Calorimetric studies on the interaction of horse ferricytochrome *c* and yeast cytochrome *c* peroxidase. *Biochemistry* 1995;34:8398–8405. [PubMed: 7599130]
47. Lee S, Tsai YC, Mattera R, Smith WJ, Kostelansky MS, Weissman AM, Bonifacino JS, Hurley JH. Structural basis for ubiquitin recognition and auto-ubiquitination by Rabex-5. *Nature Struct. Biol* 2006;13:264–271.
48. Kang SA, Crane BR. Effects of interface mutations on association modes and electron-transfer rates between proteins. *Proc. Natl. Acad. Sci. USA* 2005;102:15465–15470. [PubMed: 16227441]
49. Tang C, Iwahara J, Clore GM. Visualization of transient encounter complexes in protein-protein association. *Nature* 2006;444:383–386. [PubMed: 17051159]
50. Crowley PB, Ubbink M. Close encounters of the transient kind: protein interactions in the photosynthetic redox chain investigated by NMR spectroscopy. *Acc. Chem. Res* 2003;36:723–730. [PubMed: 14567705]
51. Nuresanu L, Pristovsek P, Löhr F, Maneg O, Mukrasch MD, Rüterjans H, Ludwig B, Lücke C. The electron transfer complex between cytochrome *c*₅₅₂ and the Cu_A domain of the *thermus thermophilus* *ba*₃ oxidase: A combined NMR and computational approach. *J. Biol. Chem* 2006;281:14503–14513. [PubMed: 16554303]
52. Blundell TL, Fernández-Recio J. Brief encounters bolster contacts. *Nature* 2006;444:279–280. [PubMed: 17051147]
53. Hicke L, Schubert HL, Hill CP. Ubiquitin-binding domains. *Nature Rev. Mol. Cell Biol* 2005;6:610–621. [PubMed: 16064137]
54. Ohno A, Jee J-G, Fujiwara K, Tenno T, Goda N, Tochio H, Kobayashi H, Hiroaki H, Shirakawa M. Structure of the UBA domain of Dsk2p in complex with ubiquitin: Molecular determinants for ubiquitin recognition. *Structure* 2005;13:521–532. [PubMed: 15837191]
55. Lowe ED, Hasan N, Trempe J-F, Fonso L, Noble MEM, Endicott JA, Johnson LN, Brown NR. Structures of the Dsk2 UBL and UBA domains and their complex. *Acta Crystallogr. D Biol. Crystallogr* 2006;62:177–188. [PubMed: 16421449]
56. Swanson KA, Hicke L, Radhakrishnan I. Structural basis for monoubiquitin recognition by the Ede1 UBA domain. *J. Mol. Biol* 2006;358:713–724. [PubMed: 16563434]
57. Prag G, Misra S, Jones EA, Ghirlando R, Davies BA, Horazdovsky BF, Hurley JH. Mechanism of ubiquitin recognition by the CUE domain of Vps9p. *Cell* 2003;113:609–620. [PubMed: 12787502]
58. Prag G, Lee S, Mattera R, Arighi CN, Beach BM, Bonifacino JS, Hurley JH. Structural mechanism for ubiquitinated-cargo recognition by the golgi-localized, γ -ear-containing, ADP-ribosylation-factor-binding proteins. *Proc. Natl. Acad. Sci. USA* 2005;102:2334–2339. [PubMed: 15701688]
59. Hirano S, Kawasaki M, Ura H, Kato R, Railborg C, Stenmark H, Wakatsuki S. Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nature Struct. Biol* 2006;13:272–277.
60. Fujiwara K, Tenno T, Sugasawa K, Jee J-G, Ohki I, Kojima C, Tochio H, Hiroaki H, Hanaoka F, Shirakawa M. Structure of the ubiquitin-interacting motif S5a bound to the ubiquitin-like domain of HR23B. *J. Biol. Chem* 2004;279:4760–4767. [PubMed: 14585839]
61. Mueller TD, Feigon J. Structural determinants for the binding of ubiquitin-like domains to the proteasome. *EMBO J* 2003;22:4634–4645. [PubMed: 12970176]

62. Heitz F, Morris MC, Fesquet D, Cavadore JC, Doree M, Divita G. Interactions of cyclins with cyclin-dependent kinases: a common interactive mechanism. *Biochemistry* 1997;36:4995–5003. [PubMed: 9125522]
63. Romero G, Chau V, Biltonen RL. Kinetics and thermodynamics of the interaction of elongation factor Tu with elongation factor Ts, guanine nucleotides, and aminoacyl-tRNA. *J. Biol. Chem* 1985;260:6167–6174. [PubMed: 3846595]
64. Huang M, Syed R, Stura EA, Stone MJ, Stefanko RS, Ruf W, et al. The mechanism of an inhibitory antibody on TF-initiated blood coagulation revealed by the crystal structures of human tissue factor, Fab 5G9 and TF.G9 complex. *J. Mol. Biol* 1998;275:873–894. [PubMed: 9480775]
65. Akerstrom B, Bjorck L. A physicochemical study of protein G, a molecule with unique immunoglobulin G-binding properties. *J. Biol. Chem* 1986;261:10240–10247. [PubMed: 3733709]
66. Ascenzi P, Amiconi G, Menegatti E, Guarneri M, Bolognesi M, Schnebli H. Binding of the recombinant proteinase inhibitor eglin c from leech *Hirudo medicinalis* to human leukocyte elastase, bovine alpha-chymotrypsin and subtilisin Carlsberg: thermodynamic study. *J. Enzyme Inhib* 1988;2:167–172. [PubMed: 3071573]
67. Lee CH, Leung B, Lemmon MA, Zheng J, Cowburn D, Kuriyan J, Saksela K. A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. *EMBO J* 1995;14:5006–5015. [PubMed: 7588629]
68. Mannherz HG, Goody RS, Konrad M, Nowak E. The interaction of bovine pancreatic deoxyribonuclease I and skeletal muscle actin. *Eur. J. Biochem* 1980;104:367–379. [PubMed: 6244947]
69. Ward ES, Gussow D, Griffiths AD, Jones PT, Winter G. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* 1989;341:544–546. [PubMed: 2677748]
70. Hurley JH, Emr SD. The ESCRT complexes: structure and mechanism of a membrane-trafficking network. *Annu. Rev. Biophys. Biomol. Struct* 2006;35:277–298. [PubMed: 16689637]
71. Prag G, Watson H, Kim YC, Beach BM, Ghirlando R, Hummer G, Bonifacino JS, Hurley JH. The Vps27/Hse1 complex is a GAT domain-based scaffold for ubiquitin-dependent sorting. *Develop. Cell* 2007;12:973–986.
72. Mizuno E, Kawahata K, Kato M, Kitamura N, Komada M. STAM proteins bind ubiquitinated proteins on the early endosome via the VHS domain and ubiquitin-interacting motif. *Mol. Bio. Cell* 2003;14:3675–3689. [PubMed: 12972556]
73. Yamakami M, Yoshimori T, Yokosawa H. Tom1, a VHS domain-containing protein, interacts with Tollip, ubiquitin and clathrin. *J. Biol. Chem* 2003;278:52865–52872. [PubMed: 14563850]
74. Levy Y, Cho SS, Onuchic JN, Wolynes PG. A survey of flexible protein binding mechanisms and their transition on states using native topology based energy landscapes. *J. Mol. Biol* 2005;346:1121–1145. [PubMed: 15701522]
75. Zhou HX, Shan YB. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins: Struct. Funct. Genet* 2001;44:336–343. [PubMed: 11455607]
76. Salwinski L, Eisenberg D. Computational methods of analysis of protein-protein interactions. *Curr. Opin. Struct. Biol* 2003;13:377–382. [PubMed: 12831890]
77. Ehrlich LP, Nilges M, Wade RC. The impact of protein flexibility on protein-protein docking. *Proteins: Struct. Funct. Bioinfo* 2005;58:126–133.
78. Tobi D, Shafran G, Linial N, Elber R. On the design and analysis of protein folding potentials. *Proteins: Struct. Funct. Bioinfo* 2000;40:71–85.
79. Creighton, TE. *Proteins: structures and molecular properties*. 2nd Ed.. W.H. Freeman and Company; New York: 1993.
80. Fraczkiwicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem* 1998;19:319–333.
81. Best RB, Chen Y-G, Hummer G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of Arc repressor. *Structure* 2005;13:1755–1763. [PubMed: 16338404]
82. Karanicolas J, Brooks CL. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* 2002;11:2351–2361. [PubMed: 12237457]

83. McLaughlin S. The electrostatic properties of membranes. *Annu. Rev. Biophys. Biophys. Chem* 1989;18:113–136. [PubMed: 2660821]
84. Swendsen RH, Wang J-S. Replica Monte Carlo simulation of spin-glasses. *Phys. Rev. Lett* 1986;57:2607–2609. [PubMed: 10033814]
85. Kumar SK, Vacatello M, Yoon DY. Off-lattice Monte Carlo simulations of polymer melts confined between two plates. *J. Chem. Phys* 1988;89:5206–5215.
86. Bonneté F, Finet S, Tardieu A. Second virial coefficient: variations with lysozyme crystallization conditions. *J. Cryst. Growth* 1999;196:403–414.
87. Ruppert S, Sandler SI, Lenhoff AM. Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnol. Prog* 2001;17:182–187. [PubMed: 11170497]

Figure 1.

Model parametrization. (a) Second virial coefficient of lysozyme computed as a function of λ and e_0 (-1.5 to -2.5 $k_B T$) for model A. The horizontal line is the experimental result⁴⁰ interpolated to pH = 7 and 100 mM salt concentration. (b) Combinations of parameter values yielding the experimental second virial coefficient for models A-F. (c) Binding probability for ubiquitin-CUE complex calculated using the (λ, e_0) combinations of (b) for models A-F. The horizontal line is the binding probability expected from the experimental affinity $K_d = 155$ μM ⁴¹ for the simulation protein concentration of 145 μM .

Figure 2.

(a) Fraction of bound states as a function of protein concentration for UIM/Ub complex. The solid curves are fits to equation (18), with K_d listed in Table 2. (b) Excess heat capacity versus protein concentration for the UIM/Ub complex at room temperature calculated for model C. The vertical dashed lines indicate the binding affinities estimated by using a structural criterion. (c) Fraction of bound states as a function of protein concentration for Cc/CcP complex and (d) excess heat capacity for Cc/CcP complex calculated for model C.

Figure 3. DRMS versus energy for the bound (a) UIM/Ub and (b) Cc/CcP complexes for model C at a system size of $R = 100$ Å. The right panels show the cumulative distributions.

Figure 4. Population of top-five clusters for the (a) UIM/Ub and (b) Cc/CcP complexes (top panels), and the corresponding DRMS values of their representative structures (bottom panels).

Figure 5.

Structures of the top ranked complexes. (a) Largest and (b) second-largest cluster for the ubiquitin (blue) and UIM1 domain (experiment: red; simulation: green). (c) Three top-ranked complex structures for Cc/CcP, with cytochrome *c* peroxidase shown in blue, cytochrome *c* from experiments⁴⁴ in red, and predicted cytochrome *c* in green. The top ranked structure is shown on the top right and with a close-up in (d). On the bottom right and left, the second and third ranked clusters are shown.

Figure 6.

Comparison between calculated and measured binding affinities for complexes of ubiquitin and several ubiquitin-binding domains (circles and triangles), and for other protein complexes (open rectangles) as listed in Table 4. Results are shown for MJ (solid symbols) and MJ1999 (open symbols) contact potentials. The solid and dashed lines represent $K_d^{\text{sim}} = K_d^{\text{exp}}$ and $K_d^{\text{sim}} = 3K_d^{\text{exp}}$, respectively.

Figure 7. Vps27-ubiquitin complex. (a) Structure. (b) Probability of membrane-tethered ubiquitin being bound to Vps27 domains as a function of the ubiquitin area concentration in the membrane. Results are shown for a fragmented Vps27 consisting of FYVE and UIM1 alone, and for UIM1, UIM2, and VHS domains of a complete Vps27 protein. Results for the full Vps27 complex can be found in Ref. 40. (c) Distances between ubiquitin and the three Vps27 domains UIM1, UIM2 and VHS along the simulation trajectories, together with the corresponding distributions.

Figure 8. Simulation structures of four different ubiquitin-binding modes for Vps27-ubiquitin complex. Representative structures of ubiquitin bound to (a) UIM1 only, (b) to both UIM1 and UIM2, (c) to VHS, and (d) to VHS and FYVE domains.

Figure 9.

(a) Plot of the normalized $u_{ij}(r)$ as a function of r/σ_{ij} for $\epsilon_{ij} < 0$ (solid) and $\epsilon_{ij} > 0$ (dashed). (b) Plots of $f(s)$ as functions of s for models A-F.

Table 1

Fitted values of λ and e_0 for models A-F.

Model	A	B	C	D	E	F
λ	0.159	0.186	0.192	0.228	0.194	0.223
$e_0(k_B T)$	-2.27	-1.95	-1.85	-1.67	-2.00	-1.96

Table 2
Calculated and measured binding affinities (K_d [μ M]) for the Ub/UIIM and Cc/CcP complexes.

Model	A	B	C	D	E	F	Exp.
Ub/UIIM	1240	530	270	40	210	120	280
Cc/CcP	1570	620	850	1680	1520	2000	100-1000

Table 3

Comparison of different contact potentials.

	Model A				Cc/CcP				Model C			
	λ	$e_0(k_B T)$	$K_d(\mu M)$	UIM/Ub % ($< 5 \text{ \AA}$)	$K_d(\mu M)$	UIM/Ub % ($< 5 \text{ \AA}$)	$e_0(k_B T)$	$K_d(\mu M)$	UIM/Ub % ($< 5 \text{ \AA}$)	$K_d(\mu M)$	Cc/CcP % ($< 5 \text{ \AA}$)	
MJ1999	1.48	0.0328	3600	77	2400	34	0.0981	770	80	1400	30	
BT	1.70	0.0372	1300	57	1600	52	0.0488	190	50	<0.1	10	
SK	0.87	0.265	760	80	2200	8	0.280	63	80	6400	7	
MJ	0.159	-2.27	1240	65	1570	20	-1.85	270	67	850	20	

Table 4

Affinity and structure of various protein complexes studied with model C. The DRMS and Q values are calculated from the lowest-energy bound structures. Averages over the ensembles of bound structures are given in parentheses.

PDB ID	Complex (A/B)	Experiment K_d^{exp} (μM)	K_d^{sim} (μM)	DRMS ($\text{\AA}$)	Q_{AB}	Simulation Q_A	Q_B	Ref.
<i>Ubiquitin/Ubiquitin binding domains</i>								
1OTR	Ub/CUE	155	155	1.3(3.7)	0.89(0.45)	0.93(0.84)	1.0(0.81)	41
1Q0W	Ub/UIM1	280	270	1.1(3.8)	0.97(0.53)	1.0(0.76)	1.0(0.86)	43
1WR1	Ub/Dsk2 UBA	14.8	430	3.8(5.8)	0.40(0.19)	0.92(0.77)	0.84(0.55)	54
2BWE	Ubl/Dsk2 UBA	80	1630	3.6(7.0)	0.35(0.14)	0.95(0.61)	0.85(0.44)	55
2G3Q	Ub/Ede1 UBA	83	340	1.2(4.3)	0.99(0.38)	1.0(0.82)	0.96(0.70)	56
1P3Q	Ub/Vps9 CUE dimer	1.6	4	0.8(4.2)	1.00(0.51)	1.0(0.89)	1.0(0.80)	57
1YD8	Ub/GGA3 GAT	181	225	4.5(7.4)	0.35(0.18)	0.84(0.76)	0.84(0.51)	58
2D3G	Ub/Hrs DUIM	190	90	1.3(3.6)	0.86(0.51)	0.90(0.80)	1.0(0.93)	59
2FID	Ub/Rabex-5 IUIIM	29	55	1.5(5.6)	0.93(0.54)	0.97(0.82)	0.97(0.81)	47
1UEL	Ubl/S5a UIM	3.4	30	2.8(3.6)	0.62(0.48)	0.94(0.84)	0.95(0.94)	60
1P9D	Ubl/S5a UIM2	12.9	230	2.5(4.4)	0.56(0.35)	0.90(0.72)	1.0(0.90)	61
<i>Other complexes</i>								
2PCC	Cytochrome <i>c</i> /Cytochrome <i>c</i> peroxidase	100	850	4.2(9.0)	0.43(0.14)	0.97(0.46)	0.78(0.50)	44
1EFU	E. coli Eftu/Efts	0.014	118	2.2(12)	0.67(0.24)	0.72(0.38)	0.83(0.50)	62
3HHR	Human growth hormone/Receptor	0.016	138	16(16)	0.00(0.1)	0.94(0.40)	0.26(0.17)	63
1MEL	Vh single-domain antibody/Lysozyme	0.02	2710	10(11)	0.09(0.03)	0.83(0.40)	0.29(0.17)	64
1FIN	CDK2 cyclin-dependent kinase 2/Cyclin	0.05	0.4	2.4(4.6)	0.59(0.43)	0.75(0.63)	0.78(0.69)	65
1MLC	IgG1 D44.1 Fab fragment/Lysozyme	0.07	820	12(19)	0.05(0.01)	0.91(0.26)	0.23(0.16)	66
1WEJ	IgG1 E8 Fab fragment/Cytochrome <i>c</i>	0.1	2790	9(14)	0.00(0.02)	0.95(0.30)	0.00(0.14)	66
1A0O	Che Y/Che A	1.1	1940	4.0(8.3)	0.43(0.10)	1.0(0.33)	0.96(0.47)	66
1MDA	Methylamine dehydrogenase/Amicyanin	4.5	54	15(15)	0.00(0.0)	0.0(0.01)	0.88(0.78)	67
1GLA	Glycerol kinase/GSF III	11	160	10(11)	0.00(0.02)	0.93(0.65)	0.0(0.10)	68
1AVZ	HIV-1 NEF/FYN tyrosine kinase SH3 domain	20	60	10(11)	0.00(0.02)	0.31(0.28)	0.11(0.27)	69
<i>Unbound docking</i>								
1UBQ/1Q0V	Ub/UIM1	280	270	2.1(5.1)	0.80(0.42)	0.90(0.69)	0.87(0.74)	
1YFC/2CYP	Cytochrome <i>c</i> /Cytochrome <i>c</i> peroxidase	100	850	4.2(9.0)	0.43(0.14)	0.97(0.46)	0.78(0.50)	

Table 5

van der Waals radii (in Å) for 20 amino acids.

ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
5.0	6.6	5.7	5.6	5.5	6.0	5.9	4.5	6.1	6.2	6.2	6.4	6.2	6.4	5.6	5.2	5.6	6.8	6.5	5.9