Balancing protein stability and activity in cancer: a new approach for identifying driver mutations affecting CBL ubiquitin ligase activation

Minghui Li[1], **Stephen C. Kales**[2], **Ke Ma**[2], **Benjamin A. Shoemaker**[1], **Juan Crespo-Barreto**[2], **Andrew L. Cangelosi**[2], **Stanley Lipkowitz**[2,*], and **Anna R. Panchenko**[1,*]

[1]National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894

[2]Women's Malignancies Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Abstract

Oncogenic mutations in the monomeric Casitas B-lineage lymphoma (Cbl) gene have been found in many tumors, but their significance remains largely unknown. Several human c-Cbl (CBL) structures have recently been solved depicting the protein at different stages of its activation cycle and thus provide mechanistic insight underlying how stability-activity tradeoffs in cancer-related proteins may influence disease onset and progression. In this study, we computationally modeled the effects of missense cancer mutations on structures representing four stages of the CBL activation cycle to identify driver mutations that affect CBL stability, binding, and activity. We found that recurrent, homozygous, and leukemia-specific mutations had greater destabilizing effects on CBL states than did random non-cancer mutations. We further tested the ability of these computational models assessing the changes in CBL stability and its binding to ubiquitin conjugating enzyme E2, by performing blind CBL-mediated EGFR ubiquitination assays in cells. Experimental CBL ubiquitin ligase activity was in agreement with the predicted changes in CBL stability and, to a lesser extent, with CBL-E2 binding affinity. Two-thirds of all experimentally tested mutations affected the ubiquitin ligase activity by either destabilizing CBL or disrupting CBL-E2 binding, whereas about one-third of tested mutations were found to be neutral. Collectively, our findings demonstrate that computational methods incorporating multiple protein conformations and stability and binding affinity evaluations can successfully predict the functional consequences of cancer mutations on protein activity, and provide a proof of concept for mutations in CBL.

Keywords

CBL; driver mutations; protein interactions

---

*corresponding authors: Stanley Lipkowitz, M.D., Ph.D., Center for Cancer Research, National Cancer Institute, National Institutes of Health, Building 10, Room 4B54, Bethesda, MD 20892-1578, Phone: (301) 402-4276, lipkowis@mail.nih.gov, Anna R. Panchenko, Ph.D., National Center for Biotechnology Information, National Institutes of Health, 8600 Rockville Pike, Bldg 38A, Bethesda, MD, 20894, Tel: (301) 435-5891, panch@ncbi.nlm.nih.gov.

Introduction

Whole exome sequencing of cancer patients has produced unprecedented amounts of data to analyze and interpret; these studies report a very large fraction of missense mutations which can potentially be implicated in tumorigenensis (1). Although some missense mutations can provide selective growth advantage to tumor cells (driver mutations), the large majority of them are considered to be neutral (passenger mutations). The mechanisms by which the driver variants may affect protein stability, interactions, and function remain largely unknown. Various computational methods have been developed to estimate the impacts of disease mutations on proteins but most of them exclusively use sequence features and do not explicitly utilize the protein three-dimensional structures, their physico-chemical properties and dynamics (2, 3). Many cancers are characterized by (de)activation of certain proteins which may be a result of missense mutations (4). The interconversion between active and inactive states is highly regulated in proteins and it is not well understood how these regulatory mechanisms are disrupted in cancer. The development of *in silico* approaches to estimate the effects of disease mutations on protein activity, stability and binding will help to define which are likely to be driver or passenger mutations. Moreover, understanding the mechanisms of their actions would allow for prioritization of potential driver candidates for better targeted therapies to design drugs which might in turn compensate for the reduced/ enhanced protein stability or activity.

The monomeric Casitas B-lineage lymphoma (Cbl) RING finger ubiquitin ligase (E3) represents an exceptionally difficult yet important system to study the mechanisms of cancer mutations (5, 6). Strikingly, proteins from this family play both positive and negative regulatory roles in tyrosine kinase signaling which is aberrantly activated in many cancers (5). Oncogenic mutations in the c-Cbl gene (referred to as CBL thereafter) were found in human myeloid neoplasms and other tumors (5) but the significance of these mutations and their impacts on CBL function were studied only for very few mutants (7). The mechanistic aspects of CBL cancer mutations can now be adequately addressed as several CBL structures have become available which represent the snapshots of different stages of the CBL activation cycle (Fig. 1). All CBL proteins share a highly conserved N-terminus which includes a tyrosine kinase–binding domain (TKBD), a linker helix region (LHR) and a RING finger domain, while the C-terminus comprises a proline-rich region (8). The RING domain of CBL has E3 activity and ubiquitinates activated receptor tyrosine kinases which subsequently targets them for degradation (8). At the same time, since CBL proteins can bind to activated receptor tyrosine kinases via the TKBD domain, they can serve as adaptors by recruiting downstream signal transduction components such as SHP2 and P13K (9, 10).

Another aspect of CBL function which should be accounted for in modeling the effects of cancer mutations is that it can bind to the ubiquitin-conjugating enzyme (E2) in complex with ubiquitin (Ub) and a substrate protein, thereby facilitating the transfer of Ub from E2 to a lysine residue of the substrate (11). The crystal structure of the inactive complex of CBL with E2 (UbcH7) was solved more than ten years ago (12) while the active phosphorylated CBL-E2 (UbcH5B) complex was resolved fairly recently (13). According to the latter study, substrate binding and Tyr371 phosphorylation activates CBL by producing a large conformational change in order to place the RING domain and E2 in close proximity to the

substrate. It was further confirmed that phosphorylation-induced conformational change is required for positioning of ubiquitin for effective catalysis (14).

Here we present a new approach, which aims to assess the effects of cancer mutations on stability, binding and activity of cancer related proteins. We apply computational models to four different stages of the CBL activation cycle (Fig. 1) and perform blind *in vivo* experiments of CBL-mediated EGFR ubiquitination. We show a rather remarkable agreement of experimental EGFR ubiquitination by CBL mutants with the computed changes in CBL thermodynamic stability and to a lesser extent with CBL-E2 binding affinity. The computational models not only quantitatively predict the magnitude of the effects of mutations but also shed light on the mechanisms of their action. Namely, we find that cancer mutations have greater destabilizing effects on four CBL states than random non-cancer mutations for recurrent, homozygous and leukemia mutations. Most damaging cancer mutations happen in the sites involved in Zn-coordination and in the formation of salt bridges and hydrogen bonds within CBL or between CBL and E2. Overall cancer driver mutations affect different or multiple stages of the CBL activation cycle either completely abolishing its E3 activity or partially attenuating it. The computational models based on stability and binding affinity calculations can discriminate experimentally validated driver from passenger mutations (with the exception of two mutations) and outperform several state-of-the-art bioinformatics methods aiming to predict phenotypic impacts of mutations.

Materials and Methods

Computational modeling and analysis

Mapping of CBL mutations—The COSMIC database (15) stores data on somatic cancer mutations and integrates the experimental data from full-genome sequencing studies. We extracted 103 missense mutations for the CBL gene from the COSMIC database (15) that could be mapped to four available CBL structures in its activation cycle (Fig. 1). Cancer mutations were classified into different classes according to the frequency of observed samples (single and recurrent mutations), zygosity (homo- and heterozygous mutations), types of cancer (leukemia and sarcoma) and the involvement of mutations in Zn-coordination (Table S1). In addition, all possible single nucleotide substitutions resulting in amino-acid changes in the CBL gene were performed to obtain a "Random" missense mutation reference set. After excluding mutations observed in COSMIC and mutations occurring on residues without known coordinates in crystal structures, we obtained 2,102 random missense mutations (Table S1). We also searched the dbSNP database (16) but found very few benign variations in the CBL gene. Detailed description is provided in Supplementary Materials and Methods.

Model preparation—We investigated the effects of mutations on four states of the CBL activation cycle: (1) closed CBL state (nCBL), (2) partially opened CBL state bound to substrate (CBL-S), (3) unphosphorylated autoinhibitory CBL bound to substrate and conjugating enzyme UbcH5B (CBL-E2-S) and (4) phosphorylated CBL bound to substrate and UbcH5B (pCBL-E2-S) (Fig. 1). It was previously noted that both UbcH5B and UbcH7 can bind specifically to CBL but only UbcH5B can facilitate ubiquitination (17). The crystal

structures of nCBL (PDB id: 2Y1M) (13), CBL-S (PDB id: 2Y1N) (13) and pCBL-E2-S (PDB id: 4A4C) (13) were obtained directly from the Protein Data Bank (PDB) (18). Only one crystal structure of the unphosphorylated inactive state of CBL bound to E2 (UbcH7) and Zap-70 peptide (PDB id: 1FBV) (12) was available in PDB. Although sequence identity between UbcH7 and UbcH5B proteins is 38%, the structural similarity is very high with the root mean square deviation (RMSD) of 1.04 Å. The unphosphorylated autoinhibitory structure of CBL-UbcH5B-S was therefore modeled based on CBL-UbcH7-S (PDB id: 1FBV) using Chimera (19). (Detailed description is provided in Supplementary Materials and Methods.)

Minimization procedure—We applied our recently developed optimization protocol for minimizing wild-type and mutant structures (20). Heavy side chain atoms without known coordinates and hydrogen atoms were added to the crystal structures using the VMD (version 1.9.1) program (21) with models immersed into rectangular boxes of water molecules extending up to 10 Å from the protein in each direction. Wild-type protein complexes were minimized for 40,000 steps using explicit TIP3P water model. The final minimized models of wild-type protein complexes were used to produce all mutant structures and then an additional 300-step minimization for all mutant structures was performed. The energy minimization was carried out with the NAMD program (version 2.9) (22) using the CHARMM27 force field (23). For unfolding free energy calculations we applied the optimization procedure implemented in the RepairPDB module of the FoldX program (24) which optimizes the side chain configurations to provide a repaired structure. (See Supplementary Materials and Methods for more information.)

Binding and unfolding free energy calculations—Binding free energy and effects of mutations on binding affinity were calculated according to the approach introduced by us earlier (20). Energy calculations were based on the modified MM-PBSA method that combined the molecular mechanics terms with the Poisson–Boltzmann continuum representation of the solvent (25) and statistical scoring energy functions with parameters optimized on experimental sets of several thousand mutations (equation 3). Binding free energy calculations were performed on minimized wild-type and mutant structures of inactive CBL-E2-S and active pCBL-E2-S states. The binding energy is defined as a difference between the free energies of the complex and unbound proteins.

$$\Delta G_{bind} = G^{CBL-E2-S} - G^{CBL-S} - G^{E2} \quad (1)$$

The change of binding energy due to a mutation can be calculated as:

$$\Delta\Delta G_{bind} = \Delta G_{bind}^{mut} - \Delta G_{bind}^{WT} \quad (2)$$

The effect of mutations on binding affinity was calculated in this work using the following energy function:

$$\Delta\Delta G_{bind} = \alpha\Delta\Delta E_{vdw} + \beta\Delta\Delta G_{solv} + \gamma\Delta SA_{mut} + \varepsilon\Delta\Delta G_{BM} + \lambda\Delta\Delta G_{FD} + \delta \quad (3)$$

Here E_{vdw} is the change of the van der Waals interaction energy and G_{solv} is the change of the polar solvation energy of solute in water. SA_{mut} represents a term proportional to the interface area of the mutant complex. G_{BM} and G_{FD} are the binding energy changes from BeAtMuSiC and FoldX, respectively. BeAtMuSiC (26) estimates the effect of mutations on binding based on statistical potentials.

The FoldX software program (24) was used to estimate the unfolding free energy and to model the unfolding state. FoldX calculates the effects of mutations on protein stability using an empirical force field. The BuildModel module was used to introduce a mutation, optimize the configurations of the neighboring side chains and calculate the difference in stability (unfolding free energy) between mutant (ΔG_{fold}^{mut}) and repaired native structure (ΔG_{fold}^{WT}).

$$\Delta\Delta G_{fold} = \Delta G_{fold}^{mut} - \Delta G_{fold}^{WT} \quad (4)$$

For comparison we used three alternative methods to predict the impacts of mutations on unfolding free energy: the Eris server (27), the Rosetta program (28) and the PoPMuSiC server (29). We also applied four additional webtools to assess the impacts of amino acid substitutions on CBL function: PROVEAN (30), PolyPhen-2 (31), MutationAssessor (32) and InCa (33). (See Supplementary Methods for more information.)

Experimental procedures

Expression constructs

Human CBL cDNA was originally obtained from Wallace Langdon and subsequently cloned into the pCEFL expression plasmid (34). Point mutation constructs, described above, were created from wild-type CBL using the QuikChange II Site-directed Mutagenesis Kit according to the manufacturer's instructions (Stratagene, La Jolla, CA). All constructs were confirmed by DNA sequencing.

Cell culture and transfections

The human embryonic kidney cell line HEK293T, the human nonsmall cell lung cancer cell line A549 and the human cervical cancer cell line HeLa used in this study were originally obtained from ATCC and maintained in culture using DMEM (Gibco, Grand Island, NY) supplemented with 10% FBS, 100 U/ml penicillin and 100 μg/ml streptomycin sulfate. Cell lines were authenticated by Short Tandem Repeat (STR) analysis by using either Promega Powerplex 16 (Promega, Madison, WI) or the AmpFISTR Identifier Kit (Life Technologies, Carlsbad, CA) and compared to the ATCC or DSMZ databases. HeLa cells were received in the lab in 2013 last authenticated on 7-30-15, A549 were received in the lab in 2015 and last authenticated on 8-12-15, and HEK293T were received in the lab in 1995 and last authenticated on 10-8-15.HEK293T cells were transfected using calcium phosphate according to the instructions accompanying the reagent (Profection; Promega Corp., Madison, WI), incubated 18 h prior to media change and grown for a total of 48 h prior to harvesting. A549 and Hela cells were transfected with Lipofectamine 2000 (Invitrogen,

Carlsbad, CA). Transfections were allowed to incubate 6 h prior to media change, and cells were grown an additional 48 h before being harvested. Cells were starved for 4 hours and treated with EGF (100 ng/ul). Each cell-based experiment was repeated at least two times.

Immunoblotting and immunoprecipitation

To harvest proteins, cells were washed twice in ice-cold Dulbecco's PBS-containing 200 μM sodium orthovanadate (Fisher Chemicals, Fairlawn, NJ) then lysed in ice-cold lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 10% glycerol, 100 mM iodoacetamide [Sigma-Aldrich Corp., St. Louis, MO], 2 mM sodium orthovanadate, and protease inhibitors [Complete tabs®, Roche Diagnostics Corp., Indianapolis, IN]). All whole cell lysates were cleared of cellular debris by centrifugation at $16,000 \times g$ for 15 m at 4 °C. Supernatant protein concentrations were determined using the BioRad protein assay (BioRad, Hercules, CA). For immunoblotting, 20μg of whole cell lysates were boiled in a 1:1 dilution of 2X loading buffer (62.5 mM Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 1 mg/ml bromophenol blue, 0.3573 M β-mercaptoethanol) for 5 m. For immunoprecipitations, 150 μg of each of the whole cell lysates were incubated with rabbit anti-EGFR (Ab-3; Millipore) and with Protein A/G+ agarose beads (sc-2003; Santa Cruz Biotechnology, Santa Cruz, CA). All immunoprecipitations were incubated overnight at 4°C with tumbling. Immune complexes were washed five times in 1 mL cold lysis buffer, then resuspended in 2X loading buffer, boiled for 5 m, then resolved by SDS-PAGE and transferred to nitrocellulose membranes (Protran BA85; Whatman, Sanford, MA). For immunoblot detection of proteins, the following antibodies were used: rabbit anti-EGFR (2232L; Cell Signaling), rat monoclonal high affinity anti-HA-peroxidase, (clone 3F10; Roche), rabbit anti-Cbl (sc-C-15; Santa Cruz) and mouse anti-Hsc70 (sc-7298; Santa Cruz). Horseradish peroxidase linked donkey anti-rabbit IgG (NA934V; GE Healthcare, Piscataway, NJ), or donkey anti-mouse IgG (NA931: GE Healthcare, Piscataway, NJ) immunoglobulin was used with SuperSignal (Pierce Biotechnology Inc., Rockford, IL) to visualize protein detection.

Densitometric Analysis

Immunoblots were developed on HyBlot CL Autoradiography Film (Denville Scientific Inc., South Plainfield, NJ) with an X-OMAT automated processor (Eastman Kodak, Rochester, NY). Protein expression levels were then recorded using an Epson Perfection V750 PRO scanner (Epson Inc., Long Beach, CA) and densitometric analysis was performed using Adobe Photoshop software version 7.0 (Adobe Systems Inc., U.S.A). EGFR ubiquitination signal intensity was determined by optical density in a set area normalized against the EGFR band intensity in parallel immunoblots and expressed as a densitometric ratio of ubiquitination/EGFR levels. The mean densitometric ratio of EGFR ubiquitination, in the presence of each CBL mutation, was then assessed relative to wild-type CBL where the ratio was set at 1.0.

Results

Cancer mutations impact CBL stability

We estimated the thermodynamic stability (unfolding free energy) changes, G_{fold}, upon mutation for closed (nCBL) and partially opened (CBL-S) CBL states (Fig. 1). These states did not involve binding to E2. The examination of the G_{fold} distribution for random non-cancer mutations (Fig. 2A) showed that it was very similar to experimental distributions produced by random mutagenesis on a set of different proteins (35, 36). Namely, the distribution for random non-cancer mutations was asymmetrically centered at positive energy values and there were about 10% of random mutations with highly damaging effects on CBL structures by more than 5 kcal mol^{-1} (Fig. 2A). Despite the presence of highly damaging random mutations, recurrent cancer mutations overall produced significantly larger destabilizing effects compared to random (p-value $\ll$ 0.01) for both nCBL and CBL-S states (Fig. 2A and S1, Table S2). This was not the case for single cancer mutations observed in only one patient (Table S2). Overall, the form of the G_{fold} distribution for recurrent cancer mutations was different from the G_{fold} distribution of random mutations (Kolmogorov-Smirnov test p-value $\ll$ 0.01) whereas the distribution for single cancer mutations was indistinguishable from a random mutation distribution. However, it does not mean that all single cancer mutants can be considered passenger, as we show later, this is not the case for some of them.

The top 25% cancer mutations with the largest damaging effects are presented in Table S3. Many of them occurred in Zn-coordinating sites while many others did not involve Zn-coordination sites. Some of these latter mutations (G415V, G413R and G415S) introduced large Van der Waals clashes with neighboring residues which could not be accommodated by side chain rearrangements, whereas other mutations (S376F, Y371D and L405P) affected disulfide or hydrogen bonds. For example, L405P located in the middle of an α-helix affected the CBL structure by introducing an energetically unfavorable kink in the helix due to its inability to donate an amide hydrogen bond.

Effects of cancer mutations on CBL-E2 binding

The outcome of mutations can be assessed by the extent of structural changes they induce. We calculated the local root mean squared deviation (RMSD) between the minimized wild-type and mutant structures around the mutated site. We found that the protein backbone in the vicinity of a mutation of the CBL-E2 complex underwent larger local conformational changes upon recurrent cancer mutations, especially for mutations occurring in Zn-coordinating site clusters, compared to random mutations (p-value $\ll$ 0.01, Fig. S2 and S3). In the previous section we analyzed the original closed (nCBL) state and partially opened (CBL-S) state induced by substrate binding. In this section we study the effects of cancer mutations on two other CBL states: an autoinhibitory CBL state bound to substrate and conjugating enzyme UbcH5B (CBL-E2-S) and phosphorylated active CBL state bound to substrate and UbcH5B (pCBL-E2-S) (Fig. 1) (13). Effects of mutations on binding can be in general linked with their structural locations, therefore we examined the locations of mutations and found that about one third of all cancer mutations were located on the CBL-

E2 interface (Fig. S4). This preference was found to be statistically significant (p-value ≪ 0.01) (Fig. S4).

It was previously experimentally shown that binding between CBL and E2 was rather weak with a micromolar dissociation constant (13). Despite the fact that the interface between CBL and E2 in the inactive state is several residues larger than in the active state, consistent with experiments, we found somewhat stronger binding between CBL and E2 for the active state of CBL (Table S4). Similarly to the impacts of cancer mutations on stability, recurrent (but not single) mutations destabilized CBL-E2 binding significantly (Fig. 2B, S1 and Table S2). Importantly, cancer mutations reduced the CBL-E2 binding in the active state considerably more than in the inactive state (p-value = 0.002) (Fig. 2B) and the effects of cancer mutations on CBL-E2 binding of the active state were noticeably larger compared to random mutations even if we excluded mutations in Zn-coordinating sites (Table S2). This latter observation does not hold true for the autoinhibitory CBL-E2-S conformation.

Next, all cancer mutations were ranked with respect to their effects on CBL-E2 binding. About half of all highly damaging mutations (top 25% of most damaging mutations, Table S3) impacted both stability and CBL-E2 binding for all four CBL states. Among them, several mutations occurred in Zn-coordinating clusters and M400 and L405 sites. Another class of mutations (involving W408 and R420 sites) mostly influenced CBL-E2 binding. For example, the W408S mutation caused the largest perturbation in CBL-E2 binding of the active state by decreasing binding affinity up to 3.5 kcal mol^{-1} but having a moderate effect on stability of the nCBL and CBL-S states. It was previously suggested that amino acids in positions W408 and I383 constituted specificity of CBL-E2 binding and corresponded to binding hot spots (10, 12). Mutations in the I383 site are not recorded in the COSMIC database although they have a profound destabilizing effect on CBL-E2 binding for both states (see data for random mutations on the ftp site ftp://ftp.ncbi.nih.gov/pub/panch/CBL). Another site affecting binding of CBL (R420) is a highly conserved site in the CBL family, it strongly interacts with the Q92, W93 and S94 residues of E2. It was previously experimentally verified that mutations in this site disrupted the CBL activity and could be associated with cytokine-independent growth (37). There are 24 cancer patient samples where mutations of this site (R420Q, R420L and R420P) are found and they all produce strong destabilizing effects on the active pCBL-E2-S state.

According to the zygosity annotation, there are 49 heterozygous ("Hetero") and 27 homozygous ("Homo") mutations (Table S1) and for the rest of cancer mutations their zygocity status is undetermined. Overall, we found that destabilizing mutations were enriched among homozygous compared to heterozygous mutations for all stages of CBL activation cycle (p-value = 0.021 – 0.035) and were more prevalent in leukemia compared to sarcoma patients (p-value = 0.000 – 0.003) (Fig. 2B, Table S2). This observation could not be attributed to the prevalence of homozygous mutations in Zn-coordinating clusters since these clusters had almost equal numbers of heterozygous and homozygous mutations. It was reported earlier that many homozygous mutations could be connected to uniparental disomy when germline heterozygosity would lead to neoplasia upon reduction to homozygosity (6, 38, 39). If some patients with myeloproliferative neoplasms had germline heterozygous mutations in CBL, then lost the wild-type CBL allele and duplicated mutant allele, it would

be possible that the damaging mutant allele would be duplicated in a cancer cell with a higher probability. Indeed, human leukemia samples show the loss of the normal CBL allele.

CBL-mediated EGFR ubiquitination: comparing experiments with computational models

Next we randomly picked fifteen cancer mutations based on their predicted damaging status, trying to equally sample highly damaging and benign mutations irrespective to their frequencies in cancer samples, and experimentally tested the E3 activity of wild-type and mutant CBL proteins (see Methods, Table 1). As can be seen in Fig. 3A, in the presence of wild-type CBL, activation of the EGFR induced more than a 10-fold increase in the ubiquitination of the EGFR compared to empty vector (endogenous CBL levels) in HEK293T cells (compare lane 4 to lane 2 in top panel of Fig. 3A). The E3 activity of wild-type CBL was mirrored by a decrease in the levels of immunoprecipitated EGFR consistent with the targeting of ubiquitinated EGFR for degradation (compare lanes 2 and 4 in second panel of Fig. 3A).

We classified all mutations into three groups (damaging, attenuating and benign) according to their experimental relative densitometry data and compared experimental data with the estimates produced by the computational models. The first group of damaging mutants included C396R, H398Q, Y371H, K382E and C381A which completely abolished CBL activity (relative densitometry was less than 10%) (Fig. 3B) and the levels of total EGFR were not decreased by these mutants (Fig. 3A). Only two of these mutants belonged to the Recurrent class of mutations. Table 1 shows predicted and experimentally verified effects of mutations on all four CBL states. All five mutants were predicted to be damaging by stability and binding affinity calculations but the mechanisms of their action were different. C396R, H398Q and C381A mutations disrupted Zn-coordinating clusters and had very damaging consequences according to the stability model for both nCBL and CBL-S states and damaging effects on CBL-E2 binding even though none of these mutations were located on the CBL-E2 interface. On the other hand, the K382E mutation did not affect Zn-coordination but destabilized the nCBL state. This, in turn, could be explained by the charge substitution that led to the disruption of the K382-E373 salt bridge within the closed state of nCBL (Figure 4 and Figure S5A). Moreover, K382E mutation had a significant impact on CBL-E2 binding due to the disruption of a salt bridge between pY371 and K382 affecting the stability of CBL in the active state (Figure 4 and Figure S5D). No significant changes were observed for CBL-S and CBL-E2-S states (Figure S5B,C). K382E mutation was previously observed in Noonan syndrome and was speculated to affect CBL stability or binding (40). Finally, the Y371H mutation not only abolished phosphorylation at the Y371 site but also had a profound destabilizing effect on all four states as evident from Table 1.

The second group of mutants (M374V, V430M, P428L, Q249E and double mutant S80N/H94Y) maintained the CBL activity equivalent to or greater than wild-type CBL (relative densitometry of 80% or higher) (Fig. 3B). Consistent with this, the levels of the activated EGFR were also decreased in these samples (Fig. 3A). To confirm that the retained E3 activity of these mutants was not cell type specific, we transfected these mutants into the non-small cell lung cancer cell line A549 and the cervical cancer cell line Hela (Fig. S6). The S80N/H94Y and Q249E mutations were previously identified in human non-small cell

lung cancers making the A549 cell line a relevant cell type to investigate the function of these CBL mutants while other cancer mutations were found in other cancer types. As in the HEK293T cells, transfection of wild-type CBL into A549 and HeLa cells resulted in a significant increase in EGF stimulated ubiquitination of EGFR compared to the cells transfected with empty vector (Fig. S6). The increase in transfected CBL protein compared to endogenous CBL protein was less in HeLa than in either the A549 or HEK293T cells. Consistent with this, the fold increase in EGF stimulated ubiquitination of the EGFR by wild-type CBL was smaller in Hela cells (about three fold) compared to either A549 or HEK293T cells (more than 10 fold in each). As in HEK293T cells, the Y371H mutant did not stimulate ubiquitination compared to the vector transfected control for A549 or HeLa cells (Fig. S6A and C). The CBL mutants that were fully active in HEK293T cells (Fig. 3A and B) maintained CBL E3 activity in A549 and HeLa cells (Fig. S6). Two of these CBL mutants resulted in more ubiquitination of EGFR compared to wild type CBL. For example, the Q249E mutant showed an increased ubiquitination of EGFR in 293T and HeLa cells while the M374V mutant resulted in higher ubiquitination levels in A549 and HeLa cells (Figure 3 and S6). Only V430M mutant was on the borderline with the densitometry ratio of 0.8. All mutations from the second group were predicted to be benign according to our computational models, whereas V430M, consistent with experimental data, had a borderline destabilizing impact on CBL-E2 binding (Table 1).

Finally, the third group constituted mutations (L399V, G375P, P395A and V391I) which attenuated the CBL E3 activity according to the relative densitometry data (Fig. 3B). Only one of these mutants was observed in two cancer samples while other three belonged to the Single mutation class. Concordant with this, there were intermediate levels of EGFR (Fig. 3A). G375P and P395A mutations were predicted to have partially damaging effects while V391I and L399V were classified as benign in our predictions. Interestingly, mutations from this third group affected only some of the CBL states. In contrast, the highly damaging mutations from the first group had damaging impacts on almost all CBL states.

Next we tested if the reduction in CBL ubiquitination activity was directly correlated with the effects of mutations on stability (Fig. 5A). The relationship between experimental densitometry data and G_{fold} was better described by an exponential dependence with correlation coefficient (R) of 0.77 and 0.78 for CBL-S and nCBL states respectively (Table S5). Indeed, the change in absorption between wild-type and mutant proteins which refers to a fraction of active mutants can be described by the Boltzmann equation relating the probability of a state with the energy of this state. A mutation may lead to damaging effects and a loss of function if it impacts at least one of the CBL functional states. Taking this into consideration we calculated the correlation between densitometry and stability changes taking a maximum of G_{fold} values for the nCBL and CBL-S states. As shown in Fig. 5A and Table S5, changes in stability can indeed explain the effects of mutations on CBL ubiquitination activity with high correlation of R = 0.83. For comparison, several alternative methods were applied to predict the effect of mutations on unfolding free energy, they all reported correlation coefficients ranging from 0.21 to 0.52, with only one method PopMusic reporting a statistically significant correlation of 0.52 (Table S5). The relationship between densitometry and binding affinity changes was on the borderline of significance with a linear

correlation coefficient of 0.48 and 0.62 (if two highly damaging mutations were excluded) (Fig. 5B and Table S5).

There are various methods which predict the phenotypic effects of mutations, (30–33, 41) and some of them use structural features (31, 41). We applied four state-of-the-art independent methods to predict the impacts of mutations on CBL function: PROVEAN (30), PolyPhen-2 (31), MutationAssessor (32) and InCa (33). Some of these methods outperformed our model in classifying cancer from random mutations since they were trained to distinguish disease from neutral variants (Table S6). However, all four methods had a very limited accuracy in classifying CBL cancer mutations into those that disrupted function and those that did not. As evident from Table 1 and Table S5, all methods except for PROVEAN over-predicted damaging effects of experimentally tested mutations while PROVEAN was the only method which produced a significant correlation between the densitometry data and PROVEAN's score (R=0.58, Table S5).

Discussion

Stability-activity balance in cancer related proteins

Evolutionary selection to maintain structural, foldable and functional proteins eliminates many mutations in protein sequences. On the other hand, thermodynamic stability can be compromised in evolution to ensure certain arrangements of catalytic and binding sites, which might not be energetically optimal (36, 42). In tumorigenesis, protein stability or binding may be reduced (or in some cases increased) due to cancer mutations. As a consequence it can lead to decreased fitness at the protein level, but may confer a fitness advantage for the population of tumor cells (43, 44). However, the extent of the stability-activity tradeoff in oncogenes and tumor suppressors remains largely unknown. Using the example of the CBL protein, here we tried to elucidate the stability-activity balance and to understand whether the loss or gain of activity in cancer-related proteins can be accompanied by compromised stability or binding.

In contrast to many other computational studies, which attempt to link stability with activity by mostly focusing on one protein conformation, we performed an analysis on four different stages of the CBL activation cycle. Our predictions of changes in stability and binding were further elucidated by the experimental CBL-mediated EGFR ubiquitination assays. We found a strong relationship between the effects of mutations on CBL stability and experimentally obtained densitometry data (quantifying activity), while a relatively weaker correlation was observed between changes in densitometry and CBL-E2 binding affinity. It could be explained either by a less significant impact of experimentally tested mutations on CBL-E2 binding or by a limited coupling (compared to CBL stability) between CBL-E2 binding and E3 activity.

Drivers or passengers?

According to our study, about two thirds of all experimentally tested mutations either completely abolished or attenuated E3 activity, while one third of them were neutral. Trying to assess the limitations of our models, we applied state-of- the-art methods commonly used

to predict functional effects of mutations. However, these methods failed to distinguish inactivating from neutral mutations as accurately as the computational models which were solely based on accounting for the effects on stability and binding for different CBL states. Indeed, many methods are not specifically designed to discriminate driver from passenger mutations within the pool of cancer mutations, but rather they perform a task of distinguishing cancer from neutral mutations.

Certainly, the large majority of all mutations detected in cancer genomics studies are likely to be neutral although the collective burden of passenger mutations may also alter the course of tumorigenesis (45). When we analyzed the distribution of all single cancer missense mutations observed in the CBL gene, their average effects on stability or binding were not found to be significantly different from the pool of random mutations. These mutations either mostly constitute true passenger mutations or their oncogenic mechanisms are not directly connected to protein destabilization. The story is quite the opposite for the recurrent mutations as they have on average significantly higher destabilizing effects than random mutations and the majority of them should be drivers. An interesting group of mutations includes those which are highly damaging but found only in one cancer sample (10% of all single cancer mutations). These mutations may represent either rare driver or latent driver mutations (46). Several of these mutations were experimentally tested and are listed in Table 1.

CBL-E2 binding in deciphering the mechanisms of cancer

Our analysis showed that cancer mutations reduce CBL-E2 binding in the active state considerably more than in the inactive state. It is in contrast to the stability-activity tradeoff reported for EGFR and other cancer related proteins (47), where cancer mutations may disrupt autoinhibitory interactions and activate the kinase (48, 49). Importantly, all mutations with experimentally tested high inactivating effects have impacts on both CBL stability and CBL-E2 binding (Table 1). In fact, the correlation between G_{fold} and G_{bind} for CBL cancer mutations from COSMIC is positive (R = 0.20 – 0.48 depending on the CBL state, p-value < 0.05). It might seem counterintuitive as some residues maintain stability by sustaining the RING-TKBD autoinhibitory interactions within CBL. This interface overlaps with the CBL-E2 interface and competes against E2 binding (13). One might think that disruption of RING-TKBD interactions and destabilization of CBL might facilitate binding to E2 and therefore lead to CBL activation. A slight activation was actually observed for two mutants, one of which (M374V) was directly located on the CBL-E2 interface. However, as we showed through the positive high correlation between stability and activity changes, this mechanism is rarely observed and the vast majority of mutations disrupt CBL-E3 activity by either destabilizing CBL and/or CBL-E2 binding or by directly affecting phosphorylation of the critical tyrosine Y371.

Overall, our results support the idea that decrease of E3 activity, of the ability to ubiquitinate receptor tyrosine kinases, of CBL stability and/or CBL-E2 binding can give a selective advantage for tumor cells. However, there are several factors that complicate deciphering the mechanisms of action of CBL cancer mutations. CBL mutations can be dominant-negative and mutated CBL might not only change the CBL E3 activity but also might affect

the concentration and activity of the wild-type CBL. According to a ratiometric method to identify driver genes in cancer (1), CBL can be regarded as an oncogene since it has several mutation hot spots. On the other hand, as shown in our study and in other studies, cancer driver mutations can inactivate the E3 activity of CBL, so it can also be regarded as a tumor suppressor. Although the latter fact complicates the development of CBL targeted therapies, understanding the delicate balance among different CBL-affected pathways may facilitate the indirect drug targeting of damaged CBL proteins. The current genetics-based frameworks to analyze cancer genome-wide sequence data are necessary but not sufficient for understanding the processes of carcinogenesis and developing informed, targeted therapies. Our approach, which can be applied in general to different proteins of interest, emphasizes the importance of the physics of binding and protein conformational ensembles in deducing the mechanisms of cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013; 339:1546–58. [PubMed: 23539594]

2. Gonzalez-Perez A, Mustonen V, Reva B, Ritchie GR, Creixell P, Karchin R, et al. Computational approaches to identify functional genetic variants in cancer genomes. Nature methods. 2013; 10:723–9. [PubMed: 23900255]

3. Kroncke BM, Vanoye CG, Meiler J, George AL Jr, Sanders CR. Personalized biochemistry and biophysics. Biochemistry. 2015; 54:2551–9. [PubMed: 25856502]

4. Torkamani A, Verkhivker G, Schork NJ. Cancer driver mutations in protein kinase genes. Cancer letters. 2009; 281:117–27. [PubMed: 19081671]

5. Kales SC, Ryan PE, Nau MM, Lipkowitz S. Cbl and Human Myeloid Neoplasms: The Cbl Oncogene Comes of Age. Cancer research. 2010; 70:4789–94. [PubMed: 20501843]

6. Naramura M, Nadeau S, Mohapatra B, Ahmad G, Mukhopadhyay C, Sattler M, et al. Mutant Cbl proteins as oncogenic drivers in myeloproliferative disorders. Oncotarget. 2011; 2:245–50. [PubMed: 21422499]

7. Tan Y-HC, Krishnaswamy S, Nandi S, Kanteti R, Vora S, Onel K, et al. CBL Is Frequently Altered in Lung Cancers: Its Relationship to Mutations in MET and EGFR Tyrosine Kinases. PloS one. 2010; 5:e8972. [PubMed: 20126411]

8. Nau, MM.; Lipkowitz, S. Welcome to the family: Cbl-family gene organization, overview of structure and functions of Cbl-related proteins in various taxonomical groups. In: Tsygankov, A., editor. Cbl Proteins. New York: Nova Science Publishers, Inc; 2008. p. 3-25.

9. Schmidt MHH, Dikic I. The Cbl interactome and its functions. Nature reviews Molecular cell biology. 2005; 6:907–19. [PubMed: 16227975]

10. Kar G, Keskin O, Nussinov R, Gursoy A. Human proteome-scale structural modeling of E2–E3 interactions exploiting interface motifs. J Proteome Res. 2012; 11:1196–207. [PubMed: 22149024]

11. Budhidarmo R, Nakatani Y, Day CL. RINGs hold the key to ubiquitin transfer. Trends in biochemical sciences. 2012; 37:58–65. [PubMed: 22154517]

12. Zheng N, Wang P, Jeffrey PD, Pavletich NP. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. Cell. 2000; 102:533–9. [PubMed: 10966114]

13. Dou H, Buetow L, Hock A, Sibbet GJ, Vousden KH, Huang DT. Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. Nat Struct Mol Biol. 2012; 19:184–92. [PubMed: 22266821]

14. Dou H, Buetow L, Sibbet GJ, Cameron K, Huang DT. Essentiality of a non-RING element in priming donor ubiquitin for catalysis by a monomeric E3. Nat Struct Mol Biol. 2013; 20:982–6. [PubMed: 23851457]

15. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic acids research. 2011; 39:D945–D50. [PubMed: 20952405]

16. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001; 29:308–11. [PubMed: 11125122]

17. Huang A, de Jong RN, Wienk H, Winkler GS, Timmers HT, Boelens R. E2-c-Cbl recognition is necessary but not sufficient for ubiquitination activity. Journal of molecular biology. 2009; 385:507–19. [PubMed: 18996392]

18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic acids research. 2000; 28:235–42. [PubMed: 10592235]

19. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–12. [PubMed: 15264254]

20. Li M, Petukh M, Alexov E, Panchenko AR. Predicting the Impact of Missense Mutations on Protein–Protein Binding Affinity. J Chem Theory Comput. 2014; 10:1770–80. [PubMed: 24803870]

21. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996; 14:33–8. 27–8. [PubMed: 8744570]

22. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–802. [PubMed: 16222654]

23. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. The journal of physical chemistry B. 1998; 102:3586–616. [PubMed: 24889800]

24. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. Journal of molecular biology. 2002; 320:369–87. [PubMed: 12079393]

25. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Accounts of Chemical Research. 2000; 33:889–97. [PubMed: 11123888]

26. Dehouck Y, Kwasigroch JM, Rooman M, Gilis D. BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations. Nucleic acids research. 2013; 41:W333–W9. [PubMed: 23723246]

27. Yin S, Ding F, Dokholyan NV. Eris: an automated estimator of protein stability. Nature methods. 2007; 4:466–7. [PubMed: 17538626]

28. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. Meth Enzymol. 2004; 383:66–93. [PubMed: 15063647]

29. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. Bioinformatics. 2009; 25:2537–43. [PubMed: 19654118]

30. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PloS one. 2012; 7:e46688. [PubMed: 23056405]

31. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010; 7:248–9. [PubMed: 20354512]

32. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic acids research. 2011; 39:e118. [PubMed: 21727090]

33. Espinosa O, Mitsopoulos K, Hakas J, Pearl F, Zvelebil M. Deriving a Mutation Index of Carcinogenicity Using Protein Structure and Protein Interfaces. PloS one. 2014; 9:e84598. [PubMed: 24454733]

34. Ettenberg SA, Keane MM, Nau MM, Frankel M, Wang LM, Pierce JH, et al. cbl-b inhibits epidermal growth factor receptor signaling. Oncogene. 1999; 18:1855–66. [PubMed: 10086340]

35. Zeldovich KB, Chen P, Shakhnovich EI. Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc Natl Acad Sci USA. 2007; 104:16152–7. [PubMed: 17913881]

36. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Current opinion in structural biology. 2009; 19:596–604. [PubMed: 19765975]

37. Sargin B, Choudhary C, Crosetto N, Schmidt MH, Grundler R, Rensinghoff M, et al. Flt3-dependent transformation by inactivating c-Cbl mutations in AML. Blood. 2007; 110:1004–12. [PubMed: 17446348]

38. Grand FH, Hidalgo-Curtis CE, Ernst T, Zoi K, Zoi C, McGuire C, et al. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. Blood. 2009; 113:6182–92. [PubMed: 19387008]

39. Chen PC, Yin J, Yu HW, Yuan T, Fernandez M, Yung CK, et al. Next-generation sequencing identifies are variants associated with Noonan syndrome. Proc Natl Acad Sci USA. 2014; 111:11473–8. [PubMed: 25049390]

40. Tsai C-J, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Protein Science. 1999; 8:1181–90. [PubMed: 10386868]

41. Hurst JM, McMillan LEM, Porter CT, Allen J, Fakorede A, Martin ACR. The SAAPdb web resource: A large-scale structural analysis of mutant proteins. Human Mutation. 2009; 30:616–24. [PubMed: 19191322]

42. Studer RA, Christin P-A, Williams MA, Orengo CA. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. Proc Natl Acad Sci USA. 2014; 111:2223–8. [PubMed: 24469821]

43. Wilcken R, Wang G, Boeckler FM, Fersht AR. Kinetic mechanism of p53 oncogenic mutant aggregation and its inhibition. Proc Natl Acad Sci USA. 2012; 109:13584–9. [PubMed: 22869713]

44. Zhao S, Lin Y, Xu W, Jiang W, Zha Z, Wang P, et al. Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. Science. 2009; 324:261–5. [PubMed: 19359588]

45. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. Proc Natl Acad Sci USA. 2013; 110:2910–5. [PubMed: 23388632]

46. Nussinov R, Jang H, Tsai CJ. The structural basis for cancer treatment decisions. Oncotarget. 2014; 5:7285–302. [PubMed: 25277176]

47. Nishi H, Tyagi M, Teng S, Shoemaker BA, Hashimoto K, Alexov E, et al. Cancer missense mutations alter binding properties of proteins and their interaction networks. PloS one. 2013; 8:e66273. [PubMed: 23799087]

48. Hashimoto K, Rogozin IB, Panchenko AR. Oncogenic potential is related to activating effect of cancer single and double somatic mutations in receptor tyrosine kinases. Hum Mutat. 2012; 33:1566–75. [PubMed: 22753356]

49. Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, et al. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer cell. 2007; 11:217–27. [PubMed: 17349580]
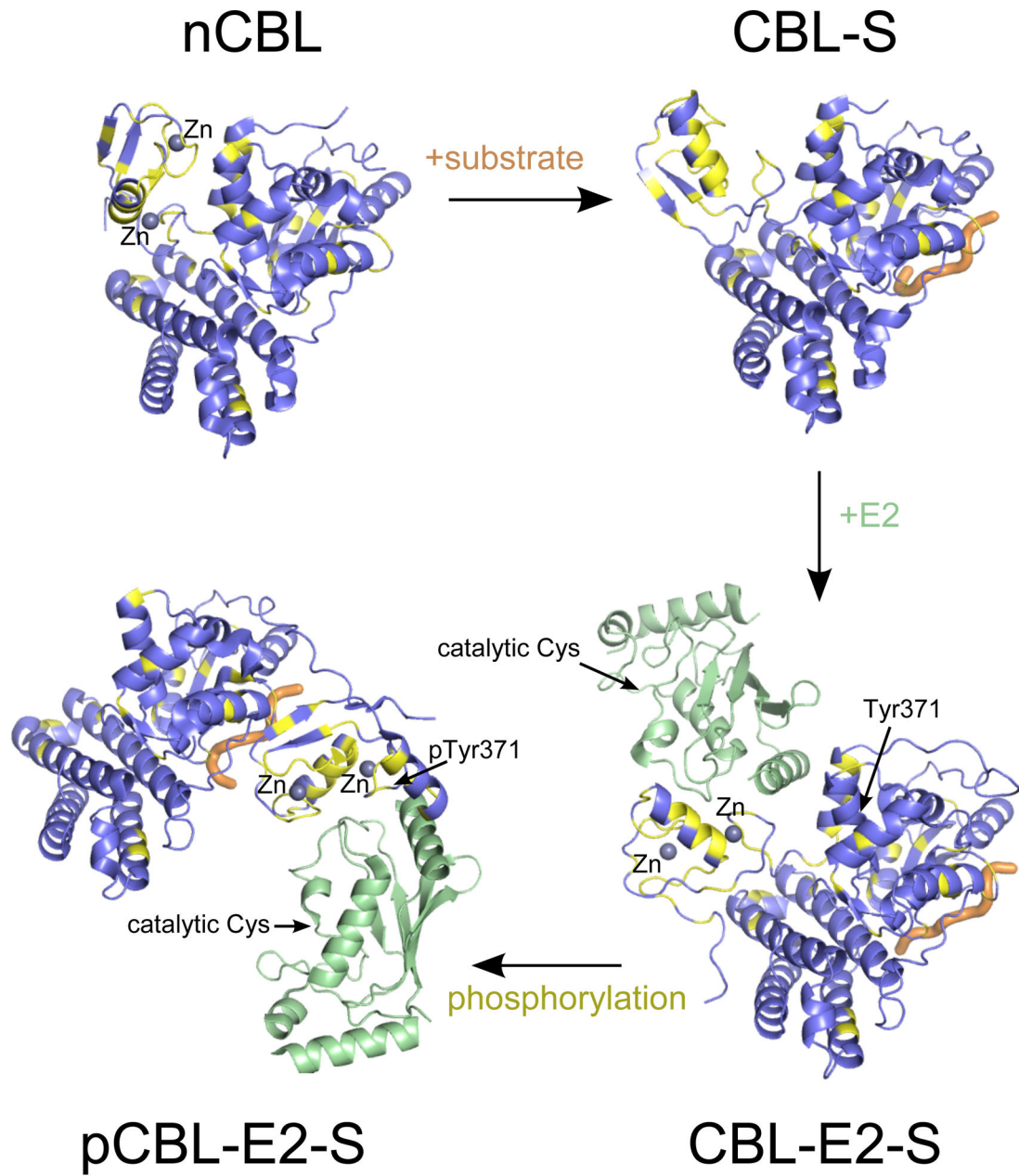
Figure 1.
CBL activation cycle. The structures representing the activation cycle of CBL are shown. In the inactive closed state (nCBL) the protein exists in the cytosol. Upon activation of the RTK, CBL can bind to the phosphorylated RTK (orange peptide) via TKBD domain. Upon phosphorylation (pCBL-E2-S) CBL undergoes a large conformational change resulting in positioning the E2 active site close to the RTK. CBL and E2 are shown in blue and green colors and all cancer mutation sites are shown in yellow.
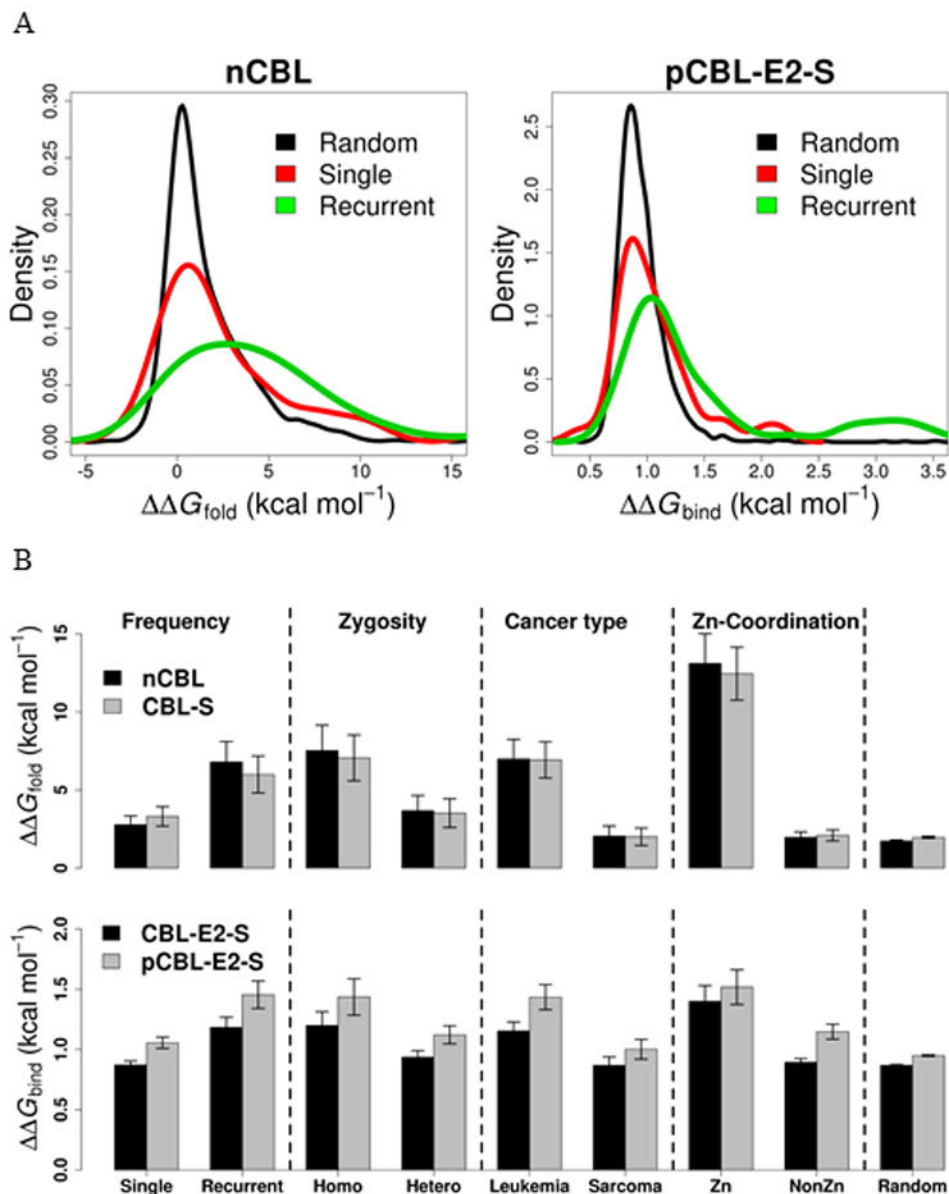
Figure 2.
(A) The distribution of unfolding (ΔG_{fold}) and binding (ΔG_{bind}) free energy changes for Single, Recurrent and Random class mutations on nCBL and pCBL-E2-S states (see SI for other states). Positive and negative values correspond to destabilizing and stabilizing effects. Recurrent cancer mutations (green line) produce significantly larger destabilizing effects for both nCBL and pCBL-E2-S (p-value $\ll 0.01$) states compared to random mutations. The distributions are smoothed by the Gaussian kernel density estimation. (B) Mean values and standard errors of ΔG_{fold} and ΔG_{bind} for different classes of mutations for nCBL, CBL-S, CBL-E2-S and pCBL-E2-S states.
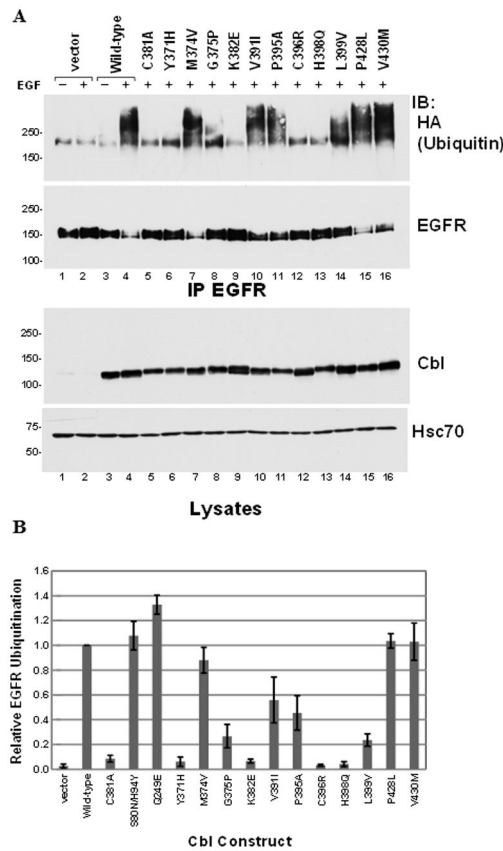
Figure 3.
(A) EGFR ubiquitination. HEK293T cells were transfected with plasmids encoding wild-type or mutant CBL proteins along with the EGFR and hemagglutinin (HA)-epitope tagged ubiquitin. The vector control and wild-type, CBL transfectant cells were incubated with or without EGF and EGFR was immunoprecipitated (IP EGFR) from each lysate. Ubiquitination of the EGFR was measured by blotting the immunoprecipitated EGFR for HA. EGFR levels in the immunoprecipitate were determined in parallel blots. The bottom two panels show CBL protein expression and HSC70 for loading in the cell lysates. The MW in kDa is shown to the left of the blots. (B) Quantification of the blots. Densitometry was performed for ubiquitination and EGFR levels; the data represent the ubiquitination/EGFR levels normalized to the ubiqutination/EGFR levels of wild-type CBL in the presence of EGF; the mean value +/− SE for at least three experiments for each CBL protein. Results for A549 and Hela cell lines are presented in Fig S5. The densitometry data for S80/H94Y and Q249E for HEK293T cells were quantified based on our previous EGFR ubiquitination experiments (7).
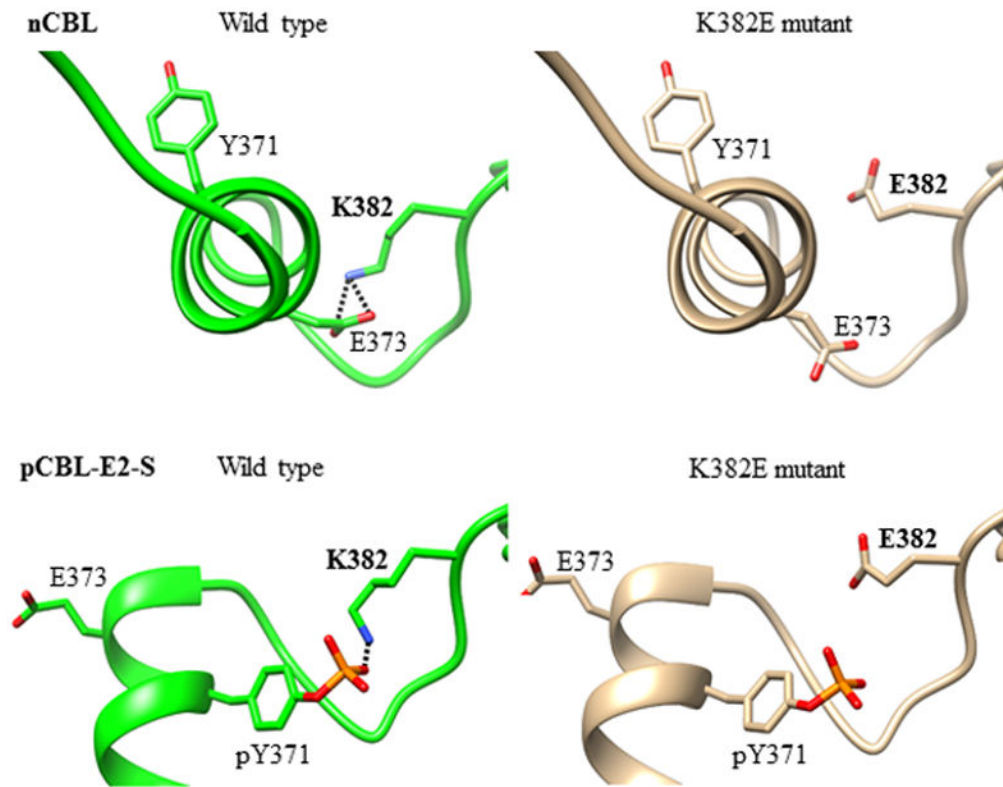
Figure 4.
Structures of wild type (green) and K382E mutant (tan) are shown for nCBL and pCBL-E2-S states. Salt bridges are shown as dashed lines and in both cases mutation affects the stability by disrupting the salt bridge.
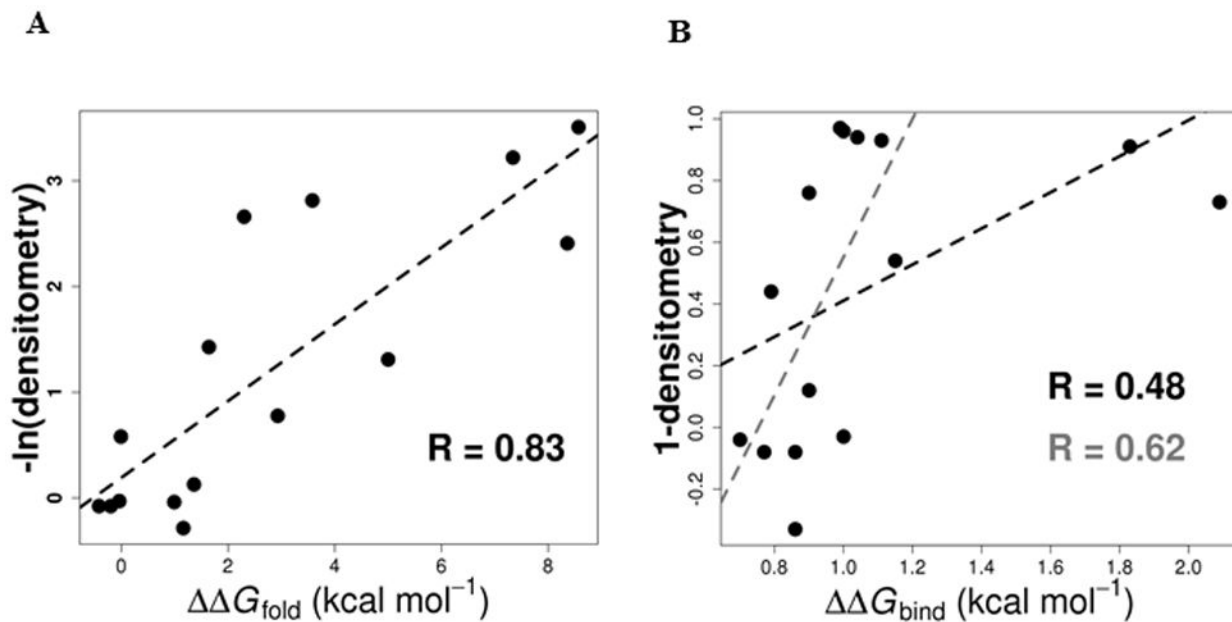
Figure 5.
(A) Comparison between experimental densitometry values (exponential model) and maximal predicted changes in stability. (B) Comparison between experimental densitometry values (linear model) and predicted changes in binding affinity of pCBL-E2-S state. Black line corresponds to 15 mutations and the grey line corresponds to the data with the exclusion of highly damaging C381A and G375P.

Table 1

Experimental relative densitometry values for HEK293T cells, predicted changes in stability (ΔG_{fold} kcal mol^{-1}), binding affinity (ΔG_{bind} kcal mol^{-1}) and comparison with other methods for 15 experimentally tested mutations.

Mutations	Densitometry	Stability		Binding affinity		PROVEAN	PolyPhen-2	Mutation Assessor	InCa
		nCBL	CBL-S	CBL-E2-S	pCBL-E2-S				
C396R	0.03±0.007	4.65	8.57	0.98	0.99	-11.39	0.99	4.69	0.88
H398Q	0.04±0.019	7.28	7.34	0.79	1.00	-7.57	1	4.34	0.73
Y371H	0.06±0.038	3.58	3.43	0.98	1.04	-4.70	1	2.35	0.92
K382E[+]	0.07±0.016	2.30	0.56	0.82	1.11	-3.74	1	2.56	0.77
C381A	0.09±0.028	8.36	7.55	2.29	1.83	-8.48	1	4.72	0.59
L399V	0.24±0.051	1.64	0.22	0.80	0.90	-2.83	1	1.70	0.34
G375P[+]	0.27±0.094	-0.19	5.00	0.67	2.09	-7.56	1	2.09	0.94
P395A	0.46±0.138	1.97	2.93	0.66	1.15	-7.54	1	2.54	0.34
V391I	0.56±0.185	-0.14	-0.01	0.81	0.79	-0.37	0.13	0.45	0.78
M374V[+]	0.88±0.104	1.36	0.87	0.81	0.90	-3.56	0.83	2.28	0.35
V430M	1.03±0.150	-0.04	-1.15	0.87	1.00	-2.19	1	2.16	0.35
P428L	1.04±0.059	0.80	0.99	0.85	0.70	-4.29	0.98	2.13	0.71
S80N	1.08±0.115	-0.42	-0.5	0.71	0.86	-2.69	1	2.56	0.84
H94Y	1.08±0.115	-1.00	-0.2	0.80	0.77	-4.26	1	2.22	0.78
Q249E	1.33±0.077	0.88	1.16	0.79	0.86	-2.81	1	2.78	0.84
Cutoff		1.80	2.04	0.87	0.95	-4.75	0.87	2.07	0.49

Cutoff is derived from the score distribution of CBL random mutations (calculated separately for each method) and is equal to the mean value plus standard error. Mutation names are in bold, italic or regular font if mutants abolish, attenuate or does not affect ligase activity respectively.

[+] indicates that mutation sites is located on CBL-E2 interface of the active state. Experimental data for S80N and H94Y mutations are derived from double mutation of S80N/H94Y and data for A549 and HeLa cell lines can be found in Fig S5. Classification of mutations for different methods based on their default cutoffs is presented in Table S7.