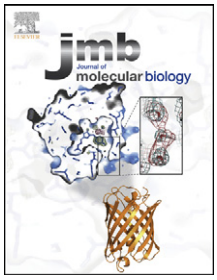


# Identifying Residues that Determine SCF Molecular-Level Interactions through a Combination of Experimental and *In silico* Analyses

**Eitan Rabinovich[1], Michael Heyne[1,2], Anna Bakhman[3], Mickey Kosloff[3], Julia M. Shifman[2] and Niv Papo[1]**

***1 - Department of Biotechnology Engineering,*** *and the National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel*
***2 - Department of Biological Chemistry,*** *The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel*
***3 - Department of Human Biology,*** *Faculty of Natural Sciences, University of Haifa, Haifa, Israel*

***Correspondence to Niv Papo:*** *Department of Biotechnology Engineering, Ben-Gurion University of the Negev, 1 Ben Gurion Avenue, Beer-Sheva 8410501, Israel. papo@bgu.ac.il*
http://dx.doi.org/10.1016/j.jmb.2016.11.018
***Edited by Sachdev Sidhu***

## Abstract

The stem cell factor (SCF)/c-Kit receptor tyrosine kinase complex—with its significant roles in hematopoiesis and angiogenesis—is an attractive target for rational drug design. There is thus a need to map, in detail, the SCF/c-Kit interaction sites and the mechanisms that modulate this interaction. While most residues in the direct SCF/c-Kit binding interface can be identified from the existing crystal structure of the complex, other residues that affect binding through protein unfolding, intermolecular interactions, allosteric or long-distance electrostatic effects cannot be directly inferred. Here, we describe an efficient method for protein-wide epitope mapping using yeast surface display. A library of single SCF mutants that span the SCF sequence was screened for decreased affinity to soluble c-Kit. Sequencing of selected clones allowed the identification of mutations that reduce SCF binding affinity to c-Kit. Moreover, the screening of these SCF clones for binding to a structural antibody helped identify mutations that result in small or large conformational changes in SCF. Computational modeling of the experimentally identified mutations showed that these mutations reduced the binding affinity through one of the three scenarios: through SCF destabilization, through elimination of favorable SCF/c-Kit intermolecular interactions, or through allosteric changes. Eight SCF variants were expressed and purified. Experimentally measured *in vitro* binding affinities of these mutants to c-Kit confirmed both the yeast surface display selection results and the computational predictions. This study has thus identified the residues crucial for c-Kit/SCF binding and has demonstrated the advantages of using a combination of computational and combinatorial methods for epitope mapping.
© 2016 Elsevier Ltd. All rights reserved.

## Introduction

Ligand–receptor interactions play a central role in many cellular functions. A detailed understanding of these interactions is thus a major goal of modern biology [1]. Precise detection—at the single-residue level—of the specific binding recognition sites on a protein ligand is a prerequisite for understanding specificity in ligand–receptor interactions [2]. The strength of these interactions (i.e., affinity) and their precision (i.e., specificity) is determined by geometrical complementarity and by the physicochemical features of the interacting residues across the binding interface. Specifically, van der Waals forces, hydrophobic interactions, salt bridges, and hydrogen bonds largely define the affinity and specificity of protein–protein interactions (PPIs) [3,4]. It is likely that the free energy of binding would be mostly affected through mutations located in the direct binding interface, especially at binding hot spots. However, mutations at more distant sites can also influence binding affinity indirectly through allosteric effects [5], global protein unfolding, or local conformational rearrangements.

Identification of the amino acid residues that participate in ligand–receptor interactions is critical both for the elucidation of mechanisms that underlie biological function and for rational drug design [6]. The currently available technique for the experimental determination of interface residues, namely, epitope mapping, is expensive, labor intensive, and time consuming [7,8]. One of the most commonly used experimental techniques for epitope mapping is alanine-scanning mutagenesis, in which the contribution of each binding interface position to binding is inferred by systematic alanine substitution and subsequent affinity measurement [1,9,10]. However, a central drawback of this technique is that alanine scanning truncates the side chain, removing its functionality. Thus, detailed information on the particular role of each residue in complex formation, such as mediating allosteric conformational changes, electrostatic/hydrophilic/hydrophobic contributions, and indirect effects involving multiple residues, is difficult to obtain with alanine scanning. However, such interactions can be identified by substituting different types of residues at a particular position, as is done in phage [11] and yeast surface display (YSD) system used in this study [12–14]. YSD technology has been used to perform epitope mapping for therapeutic antibodies against epidermal growth factor receptor [15,16] and gp120 [17] and to map the neutralizing antibodies of botulinum neurotoxin [18].

In the YSD system, high-throughput screening of tens of millions of yeast-displayed mutants allows the rapid isolation of proteins with altered properties. In this technique, yeast-displayed protein libraries are stained with a fluorescently labeled ligand or receptor [14,19], followed by fluorescence-activated cell sorting (FACS) to screen the libraries for mutants with decreased binding affinity to a target protein. One of the benefits of yeast display over other protein engineering technologies, such as phage or mRNA display, is that multicolor FACS can be used to quantitatively discriminate between clones that differ in their binding affinities to the desired target by as little as twofold [20]. This sensitivity is obtained by normalizing yeast binding levels with expression levels, so that a clone expressing a large amount of protein with low ligand binding affinity can be easily distinguished from a clone expressing a lower amount of protein but with a higher ligand binding affinity. Other advantages of the technique are that it circumvents the need to produce and purify protein variants [13] and—in contrast to phage display (for example, in shotgun scanning [21])—it is applicable for eukaryotic proteins, preserving their folding and tertiary structure and allowing glycosylation [22,23]. Taken together, these features facilitate both rapid exploration of all possible mutations and quantitative mapping of the contribution of each protein residue to the binding affinity.

While YSD can identify mutations leading to decreased PPI affinity, it cannot explain how these mutations affect binding energetics. To better understand the contribution of each residue to the PPI binding affinity, one can use computational modeling. In one such approach, developed by Sharabi and colleagues, all possible mutations are introduced at all binding interface positions, the surrounding residues are repacked, and the change in free energy of binding due to mutations ($\Delta\Delta G_{bind}$) is calculated [24,25]. The protocol for this computational saturation mutagenesis technique is fast and produces results in good agreement with experimental work [26,27]. A complementary approach developed by Kosloff *et al.* [28] for calculating per-residue energy contributions to PPIs in protein complexes combines continuum electrostatic calculations with charge perturbations and non-polar/hydrophobic energy contributions. The latter contributions are calculated by measuring the surface area buried in the complex for each residue in the protein [29]. This approach provides a quantitative measure of each residue's contributions to intra/intermolecular interactions in the PPI complex without explicit *in silico* mutagenesis, thereby bypassing computational errors that stem from the coupling of energy functions and conformation sampling [28]. The use of continuum electrostatics provides an accurate measure of electrostatic interactions and solvent effects in both wild-type and mutant proteins. In addition to calculating $\Delta\Delta G_{bind}$, changes in thermodynamic stability due to single mutations can be estimated using various protein design software packages [30–32]. Estimation of protein stability allows us to explain mutations that are likely to result in protein unfolding and consequent loss of binding.

A combination of the experimental YSD-based method for epitope mapping with computational approaches for determining $\Delta\Delta G_{bind}$ and stability calculations constitutes a powerful strategy for PPI characterization. Using such a combined experimental/computational approach, we recently mapped the binding epitope for macrophage colony-stimulating factor (M-CSF) interacting with its receptor c-FMS [33]. In the current study, we used a more comprehensive multi-method approach, combining complementary experimental and computational approaches, to map the epitope and analyze the effects of mutations in the complex between the c-Kit receptor tyrosine kinase, and its cognate stem cell factor (SCF) hematopoietic growth factor ligand [34–37].

The SCF/c-Kit interaction is important for the proliferation, adhesion, migration, differentiation, and survival in a variety of cells, including epithelial, endothelial, neuronal, and hemopoietic cells [38–40]. The controlled integration of these biological processes plays a critical role in organ formation during embryonic development [41]. Dysregulation of SCF/c-Kit signaling and gain-of-function c-Kit

mutations result in a phenotype of invasion, angiogenesis, and metastasis in many human tumors. Gain-of-function mutations in c-Kit have been identified in gastrointestinal stromal tumors (GIST), sinonasal lymphomas, testicular germ cell tumors, intracranial germinomas and papillary renal carcinomas, mast cell leukemia, and acute myeloid leukemia [42–44]. While only marginally expressed in healthy hematopoietic cells, SCF and c-Kit are overexpressed in certain leukemia cells, contributing significantly to the malignant phenotype [45,46]. Other tumor types that have been shown to simultaneously overexpress SCF and c-Kit (indicative of autocrine loops) include breast carcinomas, colorectal carcinomas, small cell lung carcinomas (SCLC), gynecological tumors, and neuroblastomas [47] (reviewed in Ref. [48]).

In light of the importance of the SCF/c-Kit interactions to disease, extensive work has been devoted to the biochemical and structural characterization of the SCF/c-Kit complex; for example, its X-ray structure has been solved (PDB ID: 2E9W) [49]. It is known that the c-Kit receptor consists of three domains, namely, an intracellular tyrosine kinase domain, a single transmembrane domain, and a ligand-binding extracellular domain composed of five immunoglobulin-like domains, D1–D5 [50]. Of these, only D1, D2, and D3 are involved in SCF/c-Kit interactions. For SCF, it is known that the soluble glycosylated form (165 aa) is a noncovalently bonded dimer, with a considerable secondary structure that includes four α-helices (αA, αB, αC, and αD) and two β-sheets (β1 and β2). This soluble SCF contains two intramolecular disulfide bonds between $Cys^{4}$–$Cys^{89}$ and $Cys^{43}$–$Cys^{138}$ [41,51]. Binding of SCF to the extracellular domain of monomeric c-Kit results in the dimerization of the receptor [52], which, in turn, activates a wide array of signaling proteins and pathways [53,54]. While these signal transduction pathways have been studied intensively using monoclonal antibodies or small-molecule kinase inhibitors, details regarding the role of SCF ligand dimerization in c-Kit receptor conformational change and activation are still lacking [51]. Moreover, the relationship between SCF dimerization and receptor binding is also unclear. The identification of the SCF residues (both within and outside the SCF/c-Kit binding interface) that are crucial for binding to c-Kit is needed for the better understanding of the molecular mechanism underlying the SCF-induced cancer cell activation. Moreover, identification of such residues would facilitate the development of new therapeutics that specifically disrupt SCF/c-Kit interactions, thereby providing the basis for patient-specific cancer treatment and disease management. To address these needs, we used YSD to map the SCF positions that reduce SCF binding affinity for c-Kit. Different approaches for computational modeling were used to interrogate the possible mechanistic effects of each identified mutation on the stability and residue-level interactions within the protein complex. Binding affinity measurements with selected purified SCF mutants confirmed these results.

## Results

### Screening a YSD SCF library for low-affinity variants against c-Kit

The SCF wild-type gene ($SCF_{WT}$) was introduced into *Saccharomyces cerevisiae* EBY100 yeast strain by using the YSD pCTCON20 vector. The $SCF_{WT}$ protein was thus displayed on the yeast cell surface conjugated to the c-Myc tag by a linker at the C terminus and to the HA tag on the N terminus (Fig. S1). To verify the expression and correct folding of the YSD $SCF_{WT}$ construct, we incubated yeast cells expressing $SCF_{WT}$ with the soluble c-Kit-Fc receptor (50 nM). Fluorescent labeling of anti-c-Myc antibody and anti-Fc antibody enabled the FACS monitoring of protein expression and protein binding, respectively. Figure 1a shows that $SCF_{WT}$ was expressed on the yeast surface and was able to bind to c-Kit. Next, we constructed a pCTSCF library of SCF mutants by using error-prone PCR. Sequencing of the library showed that SCF mutants contained, on average, one mutation incorporated in the open reading frame (data not shown). The above SCF library was then introduced and displayed on the surface of the EBY100 yeast strain. The size of the SCF library was estimated to be $1.05 \times 10^5$. Out of 14 randomly chosen and sequenced clones from the yeast-displayed SCF library, 2 contained a single-nucleotide substitution in the open reading frame, while the others were either wild-type or multiple mutants. This frequency of single mutations converts into 15,000 single-nucleotide substitutions in our whole library, accounting for more than 10-fold the maximum theoretical diversity for single-nucleotide mutations in the SCF gene (the theoretical maximum is 1305 possible substituting mutations for a gene that has 435 bp). When expressed on yeast, this SCF library showed a scattered population, with different subpopulations having different levels of expression and affinities to c-Kit, as determined by FACS (Fig. S1). To remove SCF clones that contain a stop codon or deletions/insertions, we performed a first sort in which we isolated SCF clones with a high expression level (designated SCF lib1). The SCF lib1 library was then screened against c-Kit at a concentration of 50 nM to identify SCF variants with reduced affinity to c-Kit (Fig. 1b), thereby generating a new library, designated SCF lib2 (Fig. 1c). SCF lib2 was then incubated with c-Kit at a concentration of 100 nM, and three subpopulations from SCF lib2 were collected (Fig. 1c), namely, SCF lib2.1 (lowest affinity), SCF lib2.2 (medium affinity), and SCF lib2.3 (highest affinity; Fig. 1d). For comparison, $SCF_{WT}$ was analyzed by

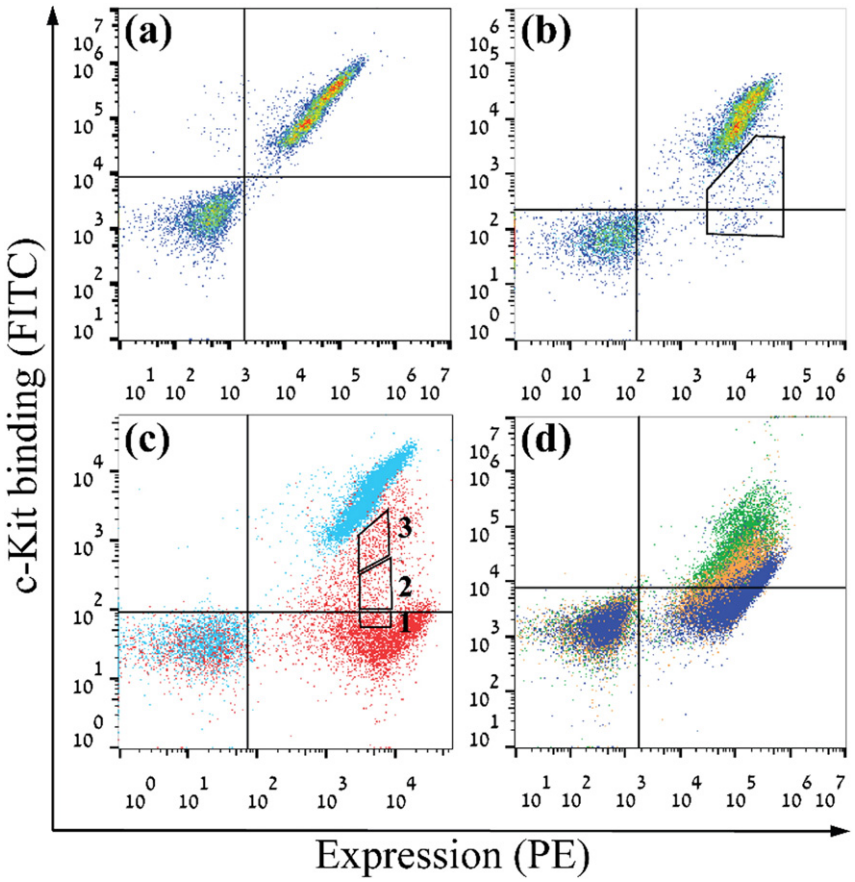


**Fig. 1.** Analysis of the affinity of YSD SCF variants to c-Kit (labeled with FITC) and YSD SCF expression levels by using anti c-Myc antibody labeled with PE. (a) $SCF_{WT}$ labeled with 50 nM c-Kit. (b) SCF library labeled with 50 nM c-Kit. The black pentagon outline shows the sorting gate of SCF lib1. (c) $SCF_{WT}$ (blue) and SCF lib1 enriched from sort 1 (red) labeled with 100 nM c-Kit. Closed shapes 1, 2, and 3 represent sorting gates for SCF lib2.1, lib 2.2, and lib 2.3, respectively. (d) FACS analysis of enriched SCF lib2.1 (purple), SCF lib2.2 (orange), and SCF lib2.3 (green), each labeled with 100 nM c-Kit.

FACS at the same receptor concentration used for screening in this experiment, that is, 100 nM (Fig. 1c).

### Identifying SCF point mutations that reduce SCF affinity for c-Kit

To identify point mutations in SCF that lead to a reduction in affinity for c-Kit, we have randomly selected approximately 100 single clones from each SCF lib2 library (SCF lib2.1, lib2.2, and lib2.3), and these were sequenced. Most of the clones from SCF lib2.1 contained undesired insertions or deletions that did not change the c-Myc tag but exhibited reduced affinity to the c-Kit receptor. Thus, we did not perform any further analysis of this library. From sequencing 100 clones from each of the other two libraries, SCF lib2.2 and lib2.3, we identified 26 different clones having a single amino acid mutation in SCF, while the rest of the sequences had multiple mutations. These single amino acid mutations originated from a single-nucleotide point mutation (Fig. 2 and Table S1). We divided these SCF-identified mutations into four groups, depending on their intermolecular and intramolecular environment in the SCF/c-KIT complex structure [49]: SCF/c-Kit interface, SCF/SCF dimerization interface, SCF core, and SCF surface (Fig. 2 and Table S1).

### Structural and binding affinity analysis of the identified SCF mutants

To quantify the extent of affinity reduction caused by each identified mutation, we expressed each clone in the YSD setup and measured its affinity to 10 nM c-Kit using FACS. The affinity was normalized to that of $SCF_{WT}$ (Fig. 3). Figure 3 shows that all identified SCF mutants caused a significant reduction in affinity to c-Kit. To determine whether a reduction in affinity was due to the conformational destabilization of SCF, we determined the affinity of each SCF clone to an anti-SCF structural antibody that could bind SCF only when SCF is correctly folded. This antibody recognizes a discontinuous SCF epitope and is termed “structural antibody” in this paper. We note here that

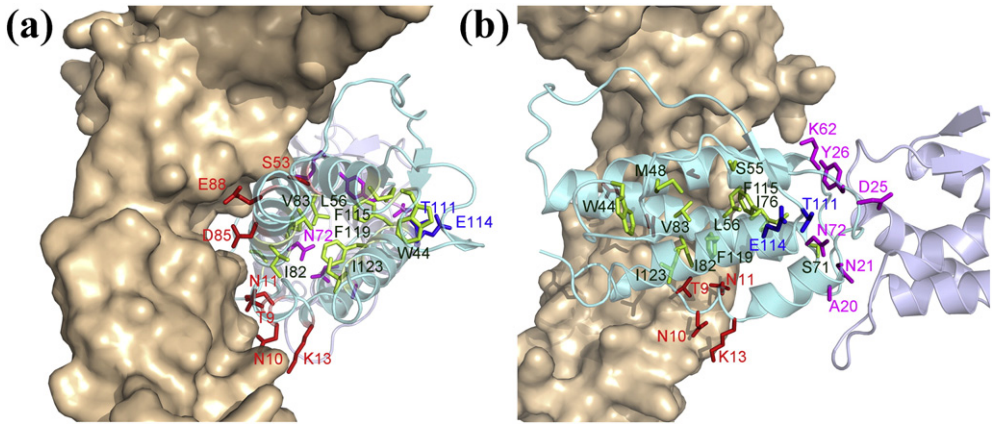


**Fig. 2.** The structure of SCF (cyan and purple) bound to c-Kit (brown) showing the positions of the affinity-reducing mutations identified in the YSD experiment. SCF residues were divided into four categories based on their location: SCF-Kit interface (T9, N10, N11, K13, S53, D85, and E88—red), SCF/SCF dimer interface (A20, N21, D25, Y26, K62, and N72—magenta), SCF core (W44, M48, S55, L56, S71, I76, I82, V83, F115, F119, and I123—green), and exposed/surface (T111 and E114—blue). (b) is identical to (a), but rotated 90° about the *y*-axis.

we had to consider the possibility that the structural antibody binds to the same or overlapping binding site on YSD SCF protein as the receptor, c-Kit. To exclude this possibility, we incubated YSD $SCF_{WT}$ with only one or both proteins, the phycoerythrin (PE)-labeled anti-SCF structural antibody and the FITC-labeled soluble c-Kit, and found almost no difference between the signal for single- and double-protein incubation (Fig. S2). Thus, we concluded that the structural antibody and c-Kit do not compete for the same binding site on $SCF_{WT}$. To confirm that the structural antibody can indeed distinguish between unfolded and folded SCF proteins, we compared the affinity of the anti-SCF structural antibody to a folded and unfolded YSD $SCF_{WT}$. The unfolded sample was generated by preheating YSD $SCF_{WT}$ to 95 °C for 5 min, which are the conditions shown to denature similar proteins in a previous study [55]. The displayed levels of the two YSD $SCF_{WT}$ samples (folded and unfolded) were the same (analyzed using PE-

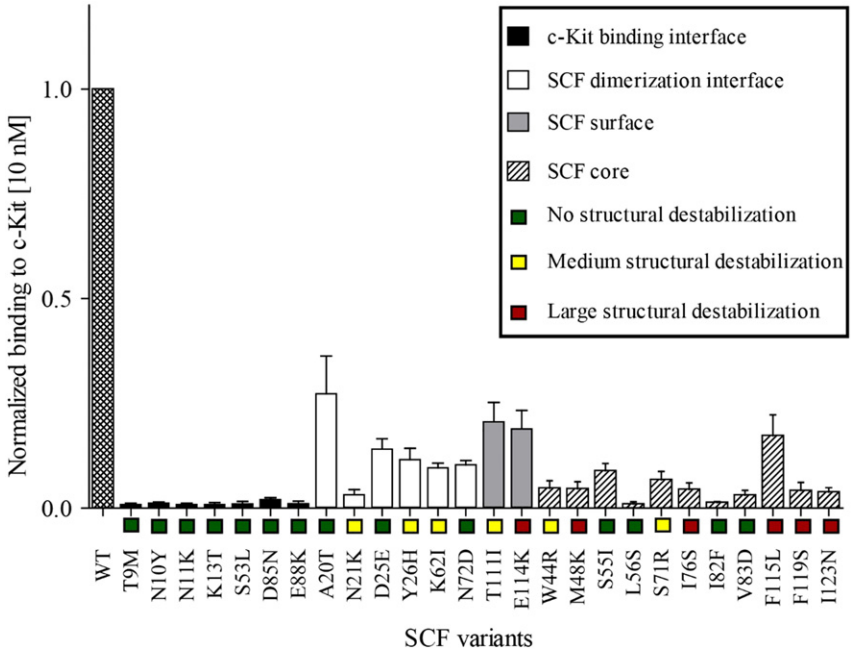


**Fig. 3.** c-Kit binding signals of individual SCF variants expressed on the yeast surface. The binding level of each SCF clone was normalized to its expression (via the cMyc tag) and to the binding signal of $SCF_{wt}$. The variants were grouped according to their structural positions (represented by column patterns) and structural stabilization properties (represented by square color tags). Each column value is a mean of triplicates ± standard error (SE).

labeled anti c-Myc), indicating no change in protein display upon structure destabilization. Importantly, we observed no binding of the structural antibody to the unfolded YSD $SCF_{WT}$ (data not shown). In order to further confirm that the structural antibody can indeed distinguish between unfolded and folded SCF proteins, we performed an ELISA assay with a purified $SCF_{WT}$ protein (see Materials and Methods section for $SCF_{WT}$ purification). In this experiment, a significant reduction in the binding of the structural antibody to unfolded $SCF_{WT}$ (denatured at 95 °C for 15 min) in comparison to folded $SCF_{WT}$ was observed (Fig. S3).

Next, we compared the affinity of each YSD SCF mutant for the anti-SCF structural antibody to that of YSD $SCF_{WT}$ (Fig. 4). The identified mutants were then divided into three categories: those that were indistinguishably folded in comparison to the wild type (i.e., showed no reduction in affinity to the structural antibody), those that were destabilized (showed some reduction in antibody affinity), and those that were unfolded (showed very low affinity to the antibody). Interestingly, all the identified mutations located at the SCF/c-Kit binding interface produced no structural destabilization of SCF. In contrast, 7 out of 11 mutations that were located in the SCF core led to structural destabilization or complete abolishment of structural antibody binding, suggesting that the reduction in affinity to c-Kit was probably a result of SCF unfolding and not of direct interaction with c-Kit. The two surface SCF mutations, T111I and E114K, resulted in medium and high structural destabilization, respectively. At the SCF/SCF dimerization site, three out of six mutations in SCF (N21K, Y26H, and K62I) caused medium structural destabilization. The other three mutations at the SCF/SCF dimerization site (A20T, D25E, and N72D) did not affect structural antibody binding, suggesting a direct correlation between SCF/SCF dimerization and SCF affinity to c-Kit when no SCF destabilization was observed (Figs. 3 and 4).

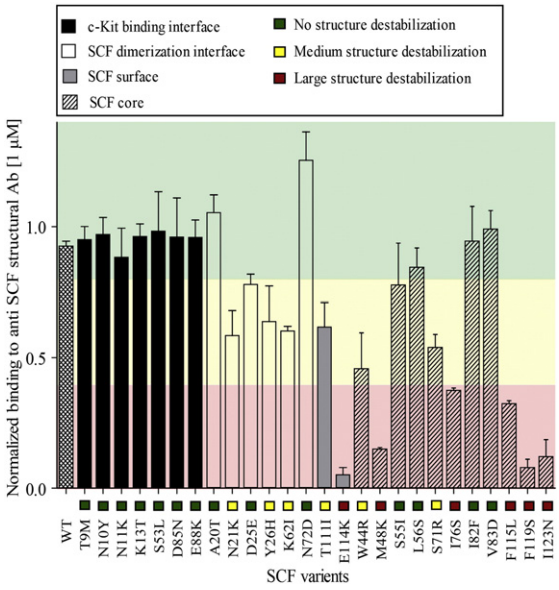


**Fig. 4.** SCF conformational stability. Flow cytometry data for the binding of monoclonal anti-SCF structural antibody to SCF wild type and variants displayed on yeast. Binding values of each YSD SCF variant normalized to its expression levels. Clones with high affinity to anti-SCF antibody, with values higher than 0.8, indicate no structural destabilization of the YSD SCF variants (marked in green). Clones showing low affinity to anti-SCF antibody, with values between 0.4 and 0.8, indicate medium structural destabilization of SCF variant YSD protein (marked in yellow). Clones showing no affinity to anti-SCF antibody, with values lower than 0.4, indicate extensive structural destabilization of YSD SCF protein (marked in red). Each column value is a mean of triplicates ± standard error (SE). We observed that in case the protein is completely unfolded, then it would not bind to the structural antibody (Fig. S3). However if there are mutations in the epitope (antigen binding site) of the antibody, there might also be complete or partial reduction in antibody affinity. Since we do not know where the epitope of the antibody is, this scenario cannot be ignored.

## Predicting changes in affinity to c-Kit due to identified SCF mutations

To predict the changes in SCF binding affinity to c-Kit due to mutations, we performed computational saturation mutagenesis on the SCF/c-Kit interface [25]. In this protocol, the X-ray structure of the SCF/c-Kit complex was used as input (PDB ID: 2E9W). We scanned 10 binding interface positions on SCF that were also identified by YSD as sites of affinity-reducing mutations with 17 aa (excluding proline, cysteine, and glycine) and we calculated the $\Delta\Delta G_{bind}$ for each of the mutations (Fig. 5a). Our results showed that 9 out of 10 experimentally identified affinity-reducing SCF mutations were predicted to decrease the affinity of the SCF/c-Kit interactions either slightly (yellow) or strongly (red; Fig. 5, circled). The only exception was position 13, where K13T was predicted to be slightly affinity-enhancing in contrast to experimental results. In addition to confirming the experimental YSD results for single mutations, our calculations show that most positions where YSD identified affinity-reducing mutations serve as hot spots for protein binding, containing a large number of mutations that destabilize SCF/c-Kit interactions.

Furthermore, we predicted the effect of mutations at the SCF dimerization interface on dimer formation by running the computational saturation mutagenesis protocol on the SCF/SCF dimerization interface. Here, six positions in the dimerization interface and one core position very close to the dimerization interface were scanned that were also identified by YSD as affinity-reducing in the SCF/c-Kit interface, and $\Delta\Delta G_{bind}$ was calculated (Fig. 5b). All seven experimentally identified mutants were also predicted

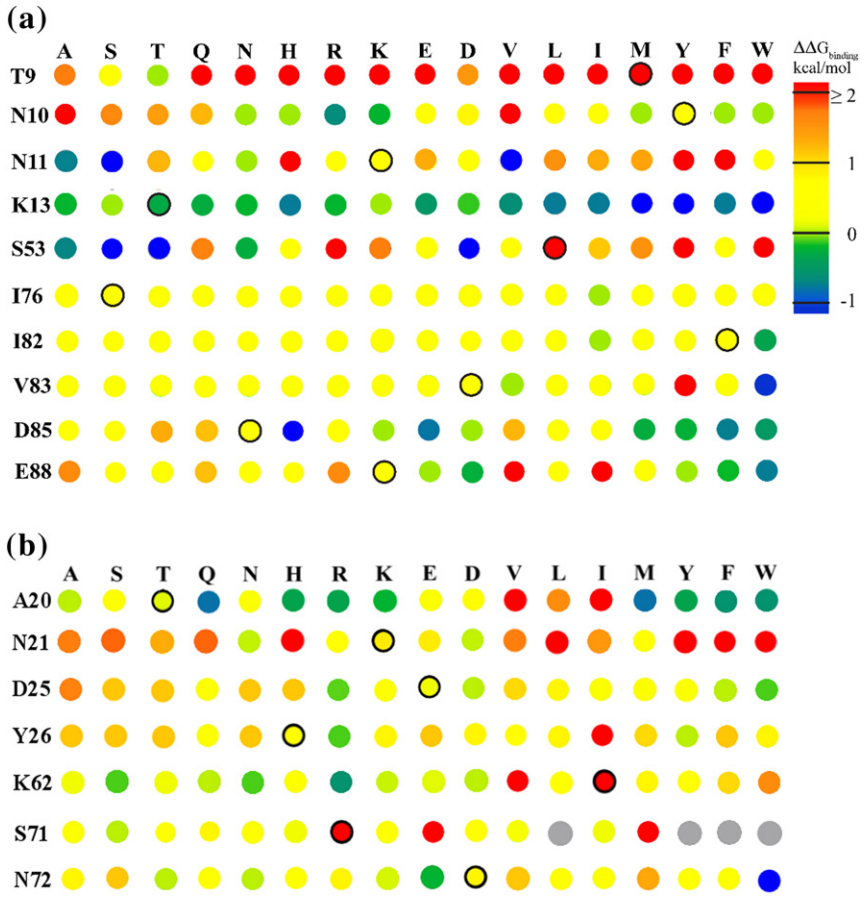


**Fig. 5.** Computational saturation mutagenesis probing (a) the SCF/c-Kit binding interface and (b) the SCF dimerization interface. SCF binding interface positions and their wild-type identity are shown on the left, and the mutated amino acids are shown on the top. The calculated $\Delta\Delta G_{bind}$ values are color coded from stabilizing (blue) through neutral (green) to destabilizing mutations (yellow and red). Gray dots represent calculations that failed to complete. Mutations identified by YSD experiments as affinity-decreasing are circled. All calculations were done using the computational saturation mutagenesis protocol described previously [24].

to reduce the dimerization ability of SCF. Furthermore, six out of seven positions were computationally identified as hot spots for monomer/monomer binding.

## Prediction of the effect of mutations on SCF stability

The structural antibody for SCF was used to experimentally determine how each mutation affects protein stability. However, mutations in the interface between this antibody and SCF may lead to reduced binding even though protein stability is not affected. To explore this scenario, we calculated the change in $\Delta G_{stability}$, namely, the change in global SCF stability due to all identified mutations. For this purpose, we used the RosettaBackrub algorithm [31], in which the mutation was introduced into the SCF, and the backbone and the surrounding side chains were allowed to adjust for the perturbation. $\Delta G_{stability}$ was calculated by subtracting the energy of the $SCF_{WT}$ from that of the mutant (Fig. 6). The mutations were grouped in four groups: c-Kit binding interface positions, SCF dimerization interface positions, SCF core positions, and surface positions. Figure 6 shows that all the mutations that fall in the direct c-Kit binding interface were predicted to have no effect on protein stability and were not destabilizing according to experimental results. Five out of six mutations belonging to the SCF/SCF dimerization interface were predicted to be slightly or moderately destabilizing, and three of them were found experimentally to be moderately destabilizing. Most mutations lying in the SCF core were predicted to significantly destabilize SCF and did indeed exhibit high destabilization experimentally. In one case—that of the core mutation I82F, there was a lack of agreement between computational and experimental results: the mutation was predicted to be highly destabilizing but did not influence stability experimentally. The high predicted destabilization for this

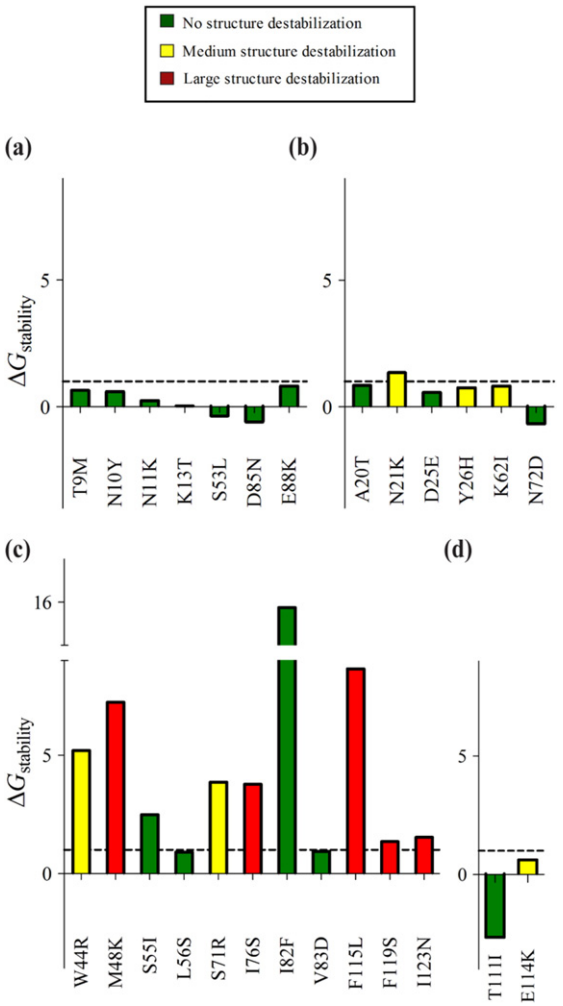


**Fig. 6.** Computationally predicted change in stability ($\Delta G_{stability}$) due to SCF single mutations for the experimentally identified mutations located (a) in the SCF/c-Kit binding interface, (b) in the SCF dimerization interface, (c) in the protein core, and (d) on the exposed surface. Each column is colored to show the experimental binding level of the structural antibody to the SCF mutant: green, no change in antibody binding with respect to WT SCF; yellow, some decrease in antibody binding; and red, large decrease in antibody binding. The dashed line at $\Delta G_{stability} = 1$ represents the stability limit above which improper folding of mutants might occur. All stability calculations were done using the RosettaBackrub suite [31].

mutation is likely due to the inability of the computational algorithm to model medium- and large-range backbone rearrangements required for accommodating a large residue in the site of the mutation. Some disagreement was also observed for the two surface mutations, T111I and E114K, which showed medium destabilization experimentally but were predicted to be either stabilizing or neutral. These two "near-neighbor" mutations might lie in the antigen binding site of the anti-SCF structural antibody, thus causing reduced antibody affinity but no actual protein unfolding.

## Electrostatic contributions of interface residues and indirect allosteric effects of core residues

To better quantify the electrostatic contributions of interface residues, we applied an approach based on the finite-difference Poisson–Boltzmann (FDPB) method [56] to calculate the net electrostatic and polar contributions ($\Delta\Delta G_{elec}$) of each SCF residue that belongs to the interface with c-Kit or with the other SCF monomer. This approach identified nine interfacial SCF residues that made significant electrostatic contributions to intermolecular interactions (Table 1). Seven of these residues were directly identified in the YSD screen. Our FDPB calculations showed that residues S53, D54, D77, and D84 participate directly in interactions with c-KIT, and residues D25, Y26, and K62 form an electrostatic/hydrogen-bond network across the SCF–SCF dimer interface (Fig. 7). Some of the mutations identified in the YSD screen affect these electrostatic interactions directly. For example, the D25E, Y26H, and K62I mutants obviously perturb the electrostatic/hydrogen-bond network that is central to the SCF–SCF dimer interface (Fig. 7a). Similarly, S53L perturbs a hydrogen-bond network with K203 and Y125 in c-Kit (Fig. 7b). However, our analysis also suggests a mechanistic basis for the reduction in binding affinity by additional mutations identified by YSD. Residues S55, L56, I76, I82, and V83 are all SCF core residues (Fig. 2) that are located beneath the SCF residues involved directly in intermolecular electrostatic interactions. The mutations that were identified in these positions (S55I, L56S, I76S, I82F, and V83D) drastically change the volume, geometry, and/or polarity of buried core residues and therefore most likely affect the local structure and hence the intermolecular interactions of the adjacent contributing residues—S53, S54, D77, and D84 (Fig. 7b and c). Since most of these mutations do not affect the binding of the structural antibody, we conclude that their effect is local and

**Table 1.** SCF interface residues with significant electrostatic contributions

| SCF interface residue | Electrostatic contributions |
|---|---|
| K24 | SCF–SCF dimer interface |
| D25 | SCF–SCF dimer interface |
| Y26 | SCF-SCF dimer interface |
| S53 | SCF–c-KIT interface |
| D54 | SCF–c-KIT interface |
| K62 | SCF–SCF dimer interface |
| S64 | SCF–SCF dimer interface |
| D77 | SCF–c-KIT interface |
| D84 | SCF–c-KIT interface |

SCF/c-Kit interface residues and SCF/SCF dimer interface residues were identified in details in the Materials and Methods section. Electrostatic contributions calculated by the FDPB method, with $\Delta\Delta G \leq -1$ kcal/mol considered as significant (see Materials and Methods).

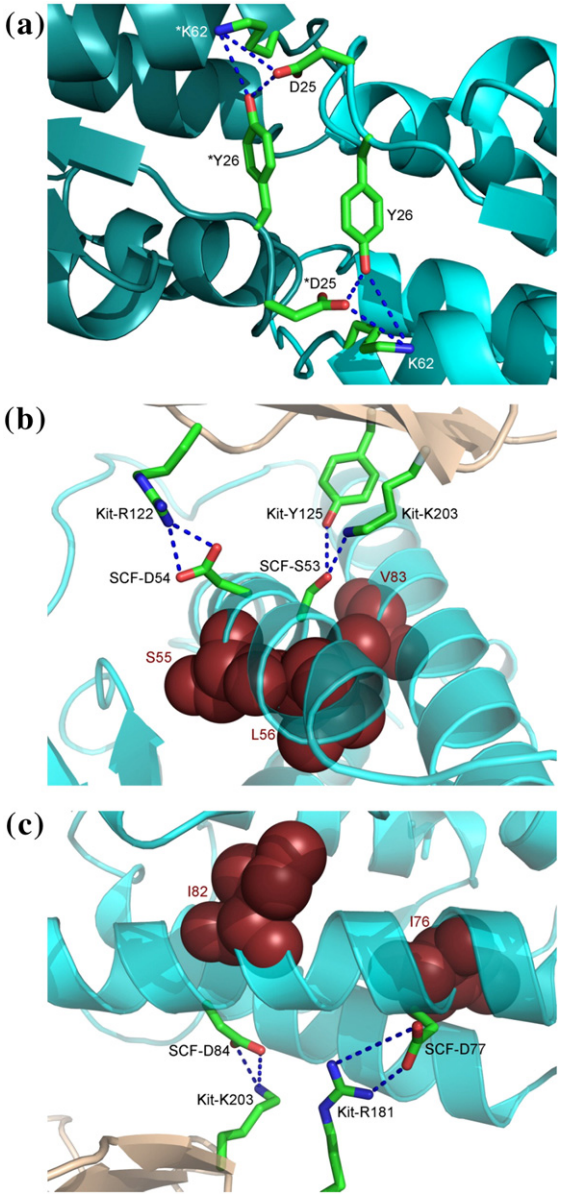


**Fig. 7.** SCF residues that significantly contribute electrostatically to intermolecular interactions. (a) SCF residues that form an electrostatic/hydrogen-bond network across the SCF–SCF dimer interface. (b and c) SCF residues that contribute electrostatically to c-Kit interactions (shown as sticks). These SCF residues are adjacent to SCF core residues (shown as dark red spheres) that are positioned to perturb the electrostatic contributions of the former upon the mutation of the latter core residues.

does not result in a major conformational change throughout the protein.

## Expression and *in vitro* characterization of selected SCF mutants

To validate our results, we chose 12 affinity-reducing SCF mutants for expression and biophysical characterization. The mutants were chosen to represent mutations in different structural elements: $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$ located at the SCF/c-Kit binding interface; $SCF_{A20T}$ located at the SCF/SCF dimerization interface; $SCF_{T111I}$ located at the SCF surface; and $SCF_{W44R}$, $SCF_{M48K}$, $SCF_{V83D}$, and $SCF_{F115L}$ located in the SCF core. $SCF_{WT}$ was also expressed for purposes of comparison. We chose to express the SCF mutants in *Pichia pastoris*, since such expression usually produces folded proteins supplied with post-translational modifications secreted into the growth medium. Western blot assay of the SCF mutants showed the clear expression of $SCF_{WT}$ and 9 out of the 12 chosen mutants $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{A20T}$, $SCF_{S53L}$, $SCF_{D85N}$, $SCF_{E88K}$, $SCF_{T111I}$, and $SCF_{F115L}$ but showed no expression of $SCF_{W44R}$, $SCF_{M48K}$, and $SCF_{V83D}$ (data not shown). Further attempts to optimize the expression conditions for the $SCF_{W44R}$, $SCF_{M48K}$, and $SCF_{V83D}$ mutants were unsuccessful. The lack of expression of the two core mutants $SCF_{W44R}$ and $SCF_{M48K}$ could be explained by their unfolded nature, which was manifested by a lack of affinity to a structural antibody and was confirmed by our calculations. $SCF_{T111I}$ and $SCF_{F115L}$ showed low expression levels, were identified only in the insoluble fraction, could not be purified as soluble proteins, and thus could not be further analyzed. A reduction in extracellular expression levels of unfolded proteins was previously shown in other studies and could be a result of proteolysis, which can occur either intra- or extracellular [57]. Other studies showed that even partially folded proteins and molten globules could be expressed and displayed on the yeast surface despite the yeast secretion quality control system [58,59]. Western blot results also showed that $SCF_{WT}$, $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{A20T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$ were expressed in two forms—with and without glycosylation (data not shown). All variants treated with Endo $H_f$ enzyme to remove any N-linked glycosylations showed only one band (at 15 kDa) (Fig. S5).

$SCF_{WT}$, $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{A20T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$ proteins (Fig. 8a) were purified on a Ni-NTA column, followed by gel filtration, and their mass was verified by MS/MS (Fig. S5). CD analysis showed that the glycosylated and the non-glycosylated $SCF_{WT}$ isoforms exhibited the same secondary structure, and surface plasmon resonance (SPR) experiments showed that the two SCF isoforms bind to c-Kit with comparable binding affinity ($K_D$ = 3.7 ± 0.7 nM for glycosylated and 3.1 ± 0.7 nM for non-glycosylated $SCF_{WT}$), indicating that there were no differences in the structure or the affinity for c-Kit between the glycosylated and non-glycosylated SCF (data not shown). We therefore report all further data only for glycosylated SCF protein variants. The CD spectra of the seven mutants $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{A20T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$ were very similar to the

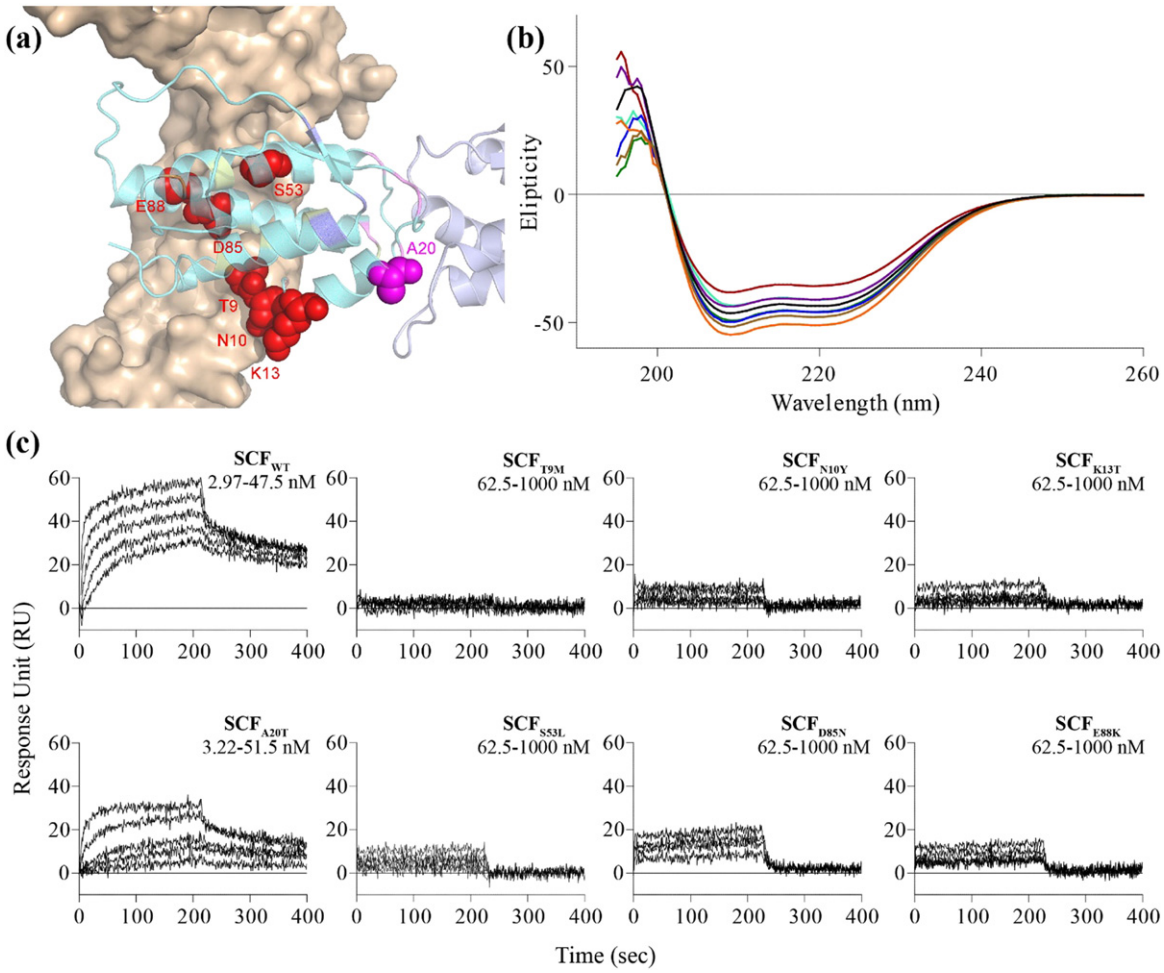


**Fig. 8.** Soluble proteins of SCF variants. (a) Residues chosen for further experimental study, in the same orientation and coloring as in Fig. 2b. (b) CD spectra of $SCF_{WT}$ (black), $SCF_{T9M}$ (purple), $SCF_{N10Y}$ (turquoise), $SCF_{T13K}$ (green), $SCF_{A20T}$ (red), $SCF_{S53L}$ (blue), $SCF_{D85N}$ (orange), and $SCF_{E88K}$ (brown) purified proteins. All measurements were performed with 20 μM protein in PBS at pH 7.4 and 25 °C. (c) SPR binding experiments. The association and dissociation curves of the purified SCF protein variants ($SCF_{WT}$, $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{T13K}$, $SCF_{A20T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$) with c-Kit are shown. Dissociation constant ($K_D$) for each complex was 3.74 nM for $SCF_{WT}$/c-Kit binding, 26.3 nM for $SCF_{A20T}$/c-Kit, and 120 nM for $SCF_{D85N}$/c-Kit; no binding was detected for other SCF/c-Kit complexes. Analyte concentrations were as follows: 2.97, 5.94, 11.88, 23.75, and 47.5 nM for $SCF_{WT}$; 3.22, 6.44, 12.88, 25.75, and 51.5 nM for $SCF_{A20T}$; 62.5, 125, 250, 500, and 1000 nM for $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{T13K}$, $SCF_{S53L}$, and $SCF_{E88K}$, using a flow rate of 100 μl/min for 216 s for the association phase and 200 s for the dissociation phase.

spectrum of $SCF_{WT}$ (Fig. 8b) and also to the spectrum of $SCF_{WT}$ observed in previous work [60], confirming that the mutants possess the same fold. These results are consistent with the lack of structural destabilization shown by the structural antibody assay and the computational results (Figs. 4 and 6, respectively). SPR spectra showed that $SCF_{A20T}$ and $SCF_{D85N}$ bound to c-Kit with reduced affinity ($K_D$ = 26.3 ± 4.8 nM and 120 ± 32.6 nM, respectively) compared to $SCF_{WT}$ ($K_D$ = 3.7 ± 0.7 nM;Fig. 8c). No binding was detected for $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{S53L}$, and $SCF_{E88K}$, indicating that their affinity to the receptor was reduced at least 1000-fold (Fig. 8c). These findings are in agreement with affinity levels of YSD $SCF_{WT}$, $SCF_{A20T}$, $SCF_{D85N}$, and the other SCF variants for soluble c-Kit, as detected by FACS (Fig. 3). In addition, the low affinity for six of the seven mutants (excluding $SCF_{K13T}$) was in agreement with our computational prediction of destabilization in $\Delta\Delta G_{bind}$ (Fig. 5a). In conclusion, the analysis of SCF variants in terms of soluble protein expression combined with CD and SPR spectra provides the confirmation of our experimental YSD and the computational results showing high structural destabilization for the $SCF_{W44R}$ and $SCF_{M48K}$ mutants, high reduction in the affinity for $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{S53L}$,

$SCF_{D85N}$, and $SCF_{E88K}$, and intermediate reduction in the affinity for $SCF_{A20T}$ mutant, respectively, compared with $SCF_{WT}$.

## Discussion

To identify the specific mutations that reduce the binding affinity of SCF to its c-Kit target receptor, the epitope mapping method used in this study combines experimental YSD with *in silico* saturation mutagenesis, structural stability calculations, and electrostatic calculations. This integrated approach facilitated the mapping of the amino acids that contribute to the binding energetics in the SCF/c-Kit interface, the SCF/SCF dimerization interface, and the SCF core. By distinguishing the contributions of these different SCF residue groups, we revealed not only the mutations in the SCF/c-Kit binding interface but also those that reduce target binding through other mechanisms, namely, SCF protein unfolding, the elimination of favorable SCF/SCF intermolecular interactions, or intramolecular (within SCF) or intermolecular (in SCF/c-Kit or SCF/SCF) allosteric and specific electrostatic effects. Thus, our integrated analysis methodology not only identified the residues that play a role in the binding process but also suggested a mechanistic basis for each loss-of-function mutation.

For mutations located in the SCF/c-Kit binding interface, we obtained very good agreement between the YSD results and computational predictions in that nine out of ten mutations identified experimentally as affinity-reducing (T9M, N10Y, N11K, S53L, I76S, I82F, V83D, D85N, and E88K) were also predicted to reduce affinity. Our modeling shows that the reduction in affinity for these mutations was achieved in a number of possible ways: by disrupting the side-chain packing by replacing small residues with larger ones (T9M, N10Y, S53L, and I82F), by introducing or removing a charge (N11K and D85N), by disrupting hydrophobic packing (I76S and V83D), by breaking hydrogen bonds (S53L and N11K), or by replacing a negatively charged residue with a positively charged residue (E88K).

We could not predict correctly the destabilization effect for one mutation K13T. A closer look shows that the error in prediction arises from the simplicity of the computational algorithm that does not repack side chains when calculating energies of the single chains (see Materials and Methods). While the K13T mutation is predicted to destabilize the total energy of the SCF/c-Kit complex (as observed), it is also predicted to considerably destabilize each single protein including the non-mutated c-Kit, resulting in a slightly negative (stabilizing) $\Delta\Delta G_{bind}$ values.

All the mutations located in the SCF dimerization interface and experimentally identified as affinity-reducing were computationally predicted to destabilize the SCF dimer, thereby suggesting that SCF dimerization is crucial to c-Kit binding. This is again in agreement with our experimental results that identified mutations in the dimeric interface lead to lower degree of dimer formation, which is important for the receptor binding.

In addition to mutations in the SCF/c-Kit binding interface and SCF/SCF dimerization interface, we identified a number of core and surface mutations that were predicted to lead to different degrees of SCF destabilization. Such mutations resulted in SCF mutants that could not be expressed by an established protocol, indicating that these mutants are significantly destabilized in solution. We thus identified a number of mutations that reduce SCF/c-Kit binding through the SCF unfolding mechanism.

In summary, our approach provides a comprehensive and unbiased way to identify mutations that significantly destabilize a particular PPI and to obtain a mechanistic explanation for the effect of each mutation. In particular for the SCF/c-kit interaction, we were able to assign the affinity-reducing mutations to three groups including those that remove favorable intermolecular interactions, which affect the ability of SCF to form a dimer, and those that significantly destabilize SCF protein. In the future, this type of study could be performed for other disease-associated PPIs as a means of elucidating the nature of disease-associated mutations, thereby paving the way for the engineering of PPI inhibitors directed toward binding hot spots.

## Materials and Methods

### Construction of the YSD $SCF_{wt}$ clone and the SCF library

The gene encoding $SCF_{WT}$ (SCF soluble domain, positions 26–166) [49] was cloned into the pCT (generously provided by Dane Wittrup laboratory, MIT, USA) YSD backbone plasmid (linearized by NheI and BamHI restriction enzymes). This was performed by homologous recombination via electroporation into *S. cerevisiae* EBY100 yeast strain [generously provided by A. Aharoni laboratory, Ben-Gurion University of the Negev (BGU), Israel] using the Gene Pulser Transfection Apparatus (Bio-Rad, CA, USA). Prior to ligation, the $SCF_{WT}$ gene was elongated using primers 1 and 2 (Supplementary Data) having pCT plasmid homology sequence, NheI and BamHI restriction sites, and a linker (TTGCCAGATAAACCATTGG CTTTCCAAGATC CATCT) located between the 3′ end of $SCF_{WT}$ gene and the 5′ end of c-Myc tag. The latter originated from the pCT plasmid YSD expression system (Fig. S1). The SCF gene library was constructed by PCR reaction with the GeneMorph® II random mutagenesis kit (Stratagene, CA, USA), according to the product

protocol for a low mutation rate, using 500 ng of $SCF_{WT}$ gene as the template and primers 1 and 2 (Supplementary data). $SCF_{WT}$ and the generated SCF gene library were amplified by PCR reaction with primers 1 and 2 using Phusion HF DNA polymerase (New England Biolabs, MA, USA) to reach 5 µg of DNA. Amplified $SCF_{WT}$ and the SCF library were cloned into the pCT plasmid by homologous recombination via electroporation into EBY100 yeast strain as described for $SCF_{WT}$ gene and were grown on SD-CAA plates [0.54% disodium phosphate (wt/vol), 0.856% monosodium phosphate monohydrate (wt/vol), 18.2% sorbitol, 1.5% agar, 2% dextrose, 0.67% yeast nitrogen base (wt/vol), and 0.5% bacto casamino acids (wt/vol)] with serial dilution plating for library size determination, resulting in $1.05 \times 10^5$ individual clones. To verify library diversity, we sequenced 14 random colonies as follows: EBY100 colonies were scraped from the SD-CAA plate into SD-CAA medium [2% dextrose (wt/vol), 0.67% yeast nitrogen base (wt/vol), 0.5% bacto casamino acids (wt/vol), 1.47% sodium citrate (wt/vol), and 0.429% citric acid monohydrate (wt/vol; pH 4.5)], and they were incubated at 30 °C with shaking at 300 rpm overnight. pCTCON plasmid containing the SCF clone was isolated from the EBY100 yeast strain using the Zymoprep™ Yeast Plasmid Miniprep kit (Zymo Research, CA, USA), according to the kit protocol. The DNA product (about 100 ng) from each colony was separately electroporated into electro-competent *Escherichia coli* bacteria using the Gene Pulser Transfection Apparatus, plated on Luria Broth-Ampicillin (LB-Amp) plates (1 mg ampicillin/L), and incubated at 37 °C for 24 h. One colony from each plate (carrying a single SCF clone) was scraped into 5 ml of LB-Amp medium and incubated at 37 °C, with shaking at 250 rpm for 24 h. Each SCF clone was isolated using HiYield plasmid mini kit (RBC Bioscience, Taiwan) and then sequenced [DNA microarrays and DNA sequencing unit of the National Institute for Biotechnology in the Negev (NIBN), BGU] using primers 3 and 4 (Supplementary Data, note 1). The transformed $SCF_{WT}$ and the SCF library were incubated in SD-CAA medium at 30 °C, with shaking at 300 rpm overnight until an $OD_{600}$ of 10 ($10^8$ cells/ml) was reached, and then stored at 4 °C until the next step.

### Affinity screening and analysis of SCF variants using flow cytometry

A YSD SCF library was generated (as described in the previous section) to identify single-point mutations in SCF that reduce the binding affinity of SCF to the extracellular domain of the c-Kit receptor. To this end, the YSD library was subjected to a selection process using FACS by labeling the YSD SCF library with a soluble fluorescently labeled c-Kit (positions 1–520). For expressing the YSD SCF library, cells from the SD-CAA medium were first induced in galactose-containing SG-CAA medium [2% galactose (wt/vol), 0.67% yeast nitrogen base (wt/vol), 0.5% bacto casamino acids (wt/vol), 1.47% sodium citrate (wt/vol), and 0.429% citric acid monohydrate (wt/vol)] at 30 °C, with shaking at 300 rpm until an $OD_{600}$ of 5 was reached. For screening, the expressed YSD SCF library was fluorescently labeled as follows: about $1 \times 10^6$ cells (or at least 10-fold more yeast cells than the library size after each sorting step) from the SG-CAA medium were washed with PBSA buffer [1% bovine serum albumin (BSA) in phosphate buffered saline] and then incubated with recombinant human c-Kit-Fc chimera (Abcam, MA, USA) and an anti-c-Myc antibody (9E10, Abcam, MA, USA) in PBSA for 1 h at room temperature. The cells were then washed with ice-cold PBSA and incubated with anti-mouse IgG-PE antibody (Sigma, MO, USA) and anti-human IgG (Fc specific) FITC antibody (Sigma, MO, USA) in PBSA at 4 °C for 20 min in the dark. Cells were washed again with ice-cold PBSA and resuspended in 400 µl of PBSA for flow cytometry analysis. To identify by FACS the SCF variants with a low affinity to c-Kit, c-Kit concentrations were increased from 10 nM in first screening cycle to 100 nM in the second cycle. The YSD expression level for each clone in the library was determined by c-Myc labeling and detection. The detection of both c-Myc and c-Kit provided an XY diagram, where the *x*-axis represented expression levels and the *y*-axis represented affinity levels to c-Kit. Sorting of clones having low binding affinity to c-Kit was performed by using diagonal sorting gates (as shown in Fig. 1b and c) to overcome avidity effects (i.e., by normalizing affinity levels to expression) and allow the enrichment of clones that showed low affinity levels to c-Kit (relative to main population) while having high expression levels. For cell sorting, we used the FACS Aria III (BD Biosciences, CA, USA) at the Scientific Equipment Center (Bar-Ilan University, Israel) and the SY3200™ (Sony Biotechnology Inc., CA, USA) at the Cytometry, Proteomic and Microscopy Unit (NIBN, BGU). We analyzed each sorting step by labeling YSD SCF variants using the same process as for the screening. Analysis was performed using an Accuri C6 flow cytometer (BD Biosciences, CA, USA), and data analysis using a FlowJo software (Treestar, Inc., CA, USA). In the last sorting step, we differentiated between three different gates having similar expression levels but different, but continuous, affinity levels. The three gates included populations having different degrees of affinity levels to c-Kit (all of them lower than $SCF_{WT}$). Using this method, we were able to screen and separate low, medium, and high affinity c-Kit-binding SCF populations. To reveal single-point mutations that led to SCF affinity reduction to c-Kit, 100 clones from each library were sequenced (as described in the previous section). Each YSD SCF variant with a single-point mutation

that was identified, except for mutations from or to proline, cysteine, and glycine, was analyzed for binding affinity to c-Kit and was compared to YSD $SCF_{WT}$. The assay was performed as follows: cells were induced and labeled as described earlier using 10 nM soluble c-Kit and an anti c-Myc Ab and using fluorescently labeled secondary antibodies and were analyzed in three independent repetitions using an Accuri C6 flow cytometer, with data analysis performed using FlowJo software. The geometric mean of the affinity level to c-Kit of each clone was normalized to its own geometric mean of expression, and the results were normalized to $SCF_{WT}$.

### Structural stability analysis of YSD SCF protein variants

YSD SCF clones were labeled separately for expression and for binding to the monoclonal anti human-SCF structural antibody (clone 2H8, Sigma, MA, USA). This structural antibody detects only SCF proteins having a similar tertiary structure to the wild-type variant, that is, it does not recognize unfolded or misfolded variants. The labeling was carried out in the same manner as described in the previous section. Cells were incubated with a 1:50 dilution of anti c-Myc antibody to test their expression levels or with SCF antibody (1 μM in PBSA) to test their folding. After washing, both samples were fluorescently labeled with 1:50 diluted anti-mouse IgG-PE antibody (Sigma, MO, USA) in PBSA. An Accuri C6 flow cytometer was used for cell screening, and data analysis was performed using FlowJo software. Each clone was labeled and analyzed in three independent replications.

### Recombinant expression and purification of selected SCF variants

$SCF_{WT}$ and selected variants ($SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{A20T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$) were purified according to previous protocols [61]. The following purification procedure was used: pCTCON plasmids containing SCF clones were isolated from EBY100 yeast strain using Zymoprep™ Yeast Plasmid Miniprep kit (Zymo Research, CA, USA), according to the kit protocol. SCF clones were amplified from the pCTCON plasmid using primers 5 and 6 (Supplementary Data) and were digested using EcoRI and AvrII restriction enzymes, according to the product protocol (New England Biolabs, MA, USA). The digested clones were ligated into the linear expression pPIC9K plasmid (digested with EcoRI and AvrII;Invitrogen, CA, USA) using Quick ligase (New England Biolabs, USA) according to the kit protocol. With this method, a FLAG-tag was added to the N terminus of the purified protein, and a His tag to the C terminus. The ligated product was electroporated into *E. coli*, incubated at 37 °C with shaking at 300 rpm for 1 h in LB medium [1% yeast extract (wt/vol), 2% peptone (wt/vol), and 2% dextrose (wt/vol)], plated onto LB-Amp plates (1 mg ampicillin/L), and incubated at 37 °C for 24 h. Several colonies were scraped from the plate, transferred into LB-Amp medium, and incubated at 37 °C with shaking at 300 rpm for 24 h. pPIC9K plasmids containing the SCF clones were isolated from each *E. coli* colony using HiYield plasmid mini-kit (RBC Bioscience, Taiwan) and were sequenced using primers 7 and 8 (Supplementary Data). pPIC9K plasmids containing SCF clones with the correct sequence were linearized using SacI, and about 30 μg of DNA was electroporated into *P. pastoris* strain GS115 (Invitrogen, CA, USA) for chromosomal incorporation by homologous recombination. Transformed cells were plated onto RDB plates [18.6% sorbitol (wt/vol), 2% agar (wt/vol), 2% dextrose (wt/vol), 1.34% yeast nitrogen base (wt/vol), $4 \times 10^{-5}$% biotin (wt/vol), and $5 \times 10^{-3}$% of L-glutamic acid, L-methionine, L-leucine, L-lysine, and L-isoleucine (wt/vol)] and incubated for 3 days at 30 °C. Cells were scraped from the plates using E-buffer [0.12% Tris base (wt/vol), 9.24% sucrose (wt/vol), and 0.02% MgCl2 (wt/vol; pH 7.5)] and plated onto YPD-G418 plates (4 mg/ml G418) for an additional 3 days. Colonies from each clone were harvested (about 10), seeded into 5 mL of BMGY medium [2% peptone (wt/vol), 1% yeast extract (wt/vol), 0.23% $K_2H(PO_4)$ (wt/vol), 1.1812% $KH_2(PO_4)$ (wt/vol), 1.34% yeast nitrogen base (wt/vol), $4 \times 10^{-5}$% biotin (wt/vol), and 1% glycerol (vol/vol)], and incubated at 30 °C with shaking at 300 rpm for 24 h. Each SCF clone was analyzed for expression. For expression, cells from the BMGY medium were resuspended in 5 mL of BMMY medium [2% peptone (wt/vol), 1% yeast extract (wt/vol), 0.23% $K_2H(PO_4)$ (wt/vol), 1.1812% $KH_2(PO_4)$ (wt/vol), 1.34% yeast nitrogen base (wt/vol), $4 \times 10^{-5}$% biotin (wt/vol), and 0.5% MeOH (vol/vol)] to reach an $OD_{600}$ of 1 and incubated at 30 °C with shaking at 300 rpm for 72 h, with 0.5% MeOH being added every 24 h. Protein expression and secretion to the media were analyzed by Western blot, using mouse anti-FLAG primary antibody (Sigma Aldrich, MO, USA) and alkaline phosphatase-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch, PA, USA). BCIP reagent (Sigma Aldrich, MO, USA) was used for protein expression signal analysis according to the product protocol. Individual clones with the strongest expression levels were selected and subjected to large-scale expression. Briefly, clones in 5 ml of BMGY medium at an $OD_{600}$ of 8–10 were added to 500 ml of BMGY medium at 30 °C with shaking at 300 rpm for 24 h. Cells were precipitated from the medium and induced for protein secretion in 500 ml of BMMY at 30 °C with shaking at 300 rpm for 72 h. MeOH (0.5%) was added every 24 h. After induction, the supernatant was filtered through a 0.22-μm vacuum filter. The filtrate was then subjected to the following workup: NaCl was added to reach a

final concentration of 300 mM, imidazole (Sigma, MO, USA) was added to give a final concentration of 10 mM, and the solution pH was adjusted to 8.0. After 1 h at 4 °C, the medium was filtered again, and the protein of interest was purified using 2 ml of Ni-NTA Sepharose beads column (Invitrogen, CA, USA). Protein was eluted from the Ni-NTA beads with 15 ml of buffer at pH 8.0 containing 50 mM sodium phosphate, 300 mM NaCl, and 250 mM imidazole. The eluted protein was concentrated to 0.5 ml using Vivaspin with a 3-kDa cutoff (Vivaproducts, MA, USA) and was then subjected to gel-filtration chromatography on a Superdex™ 75 10/300 GL column using an AKTA™ Pure (GE Biosciences, PA, USA) chromatography system. The gel-filtration column was pre-equilibrated with PBS, and the protein was eluted with PBS buffer. Protein concentration was calculated from protein absorbance at 280 nm (extinction coefficient of 13,200 $M^{-1}$ $cm^{-1}$ and calculated mass of 19.3 kDa) obtained using a NanoDrop spectrophotometer (Thermo Scientific, MA, USA). Protein samples were also subjected to mass spectrometry analysis (Ilse Katz Institute for Nanoscale Science Technology, BGU). The purified protein was stored at −80 °C.

### Structural detection analysis for anti-SCF structural Ab by Elisa

In order to validate that the anti-SCF structural antibody recognizes only the folded conformation of $SCF_{WT}$ protein, an ELISA assay was carried out as follows. A polystyrene, high binding 96 well plate (Microlon) was incubated with either 1:600 anti-SCF structural Ab (clone 2H8, Sigma, MA, USA) or 1:2000 anti-Flag tag Antibody for 1 h, washed three times with PBST buffer (0.05% tween in phosphate-buffered saline), incubated for blocking with 5% milk in PBS for 1 h, washed again, and incubated with 1 μM folded or unfolded (preheated at 95 °C for 15 min) purified soluble $SCF_{WT}$ protein (tagged with Flag and 6×His) in PBSA for 1 h. The plate was washed again and incubated with 1:5000 HRP anti-6×His-conjugated antibody (Sigma, MA, USA) for 1 h. After additional washing step, the plate was incubated with 100 μl of TMB substrate (Thermofisher, USA). $H_2SO_4$ (100 μl, 1 M) was added in order to terminate the reaction, and the absorption was detected at a wavelength of 450 nM. The assay was performed in three independent repetition, and the data were analyzed as a mean absorption value.

### Structural analysis of purified SCF variants by CD spectroscopy

To evaluate the secondary structure of each of the purified SCF proteins, we used a J-815 CD spectrometer (JASCO, Tokyo, Japan) with a 1-mm path length quartz cuvette. Spectra of 20 μM of purified protein in 400 μl of PBS were obtained at 25 °C, and the signal of the blank solvent (PBS) was subtracted. Ellipticity (degree × $cm^2$/dmol) was obtained from the normalized average of three spectra in the range of 195–260 nm. Data points with a diode voltage of >1100 V were excluded.

### Yeast-displayed SCF dimerization analysis

Dimerization levels of selected YSD SCF mutants, having mutations at the SCF dimerization site, were tested as follows. Each YSD SCF mutant clone was labeled for expression by incubation with diluted 1:50 mouse anti c-Myc antibody in 1% PBSA (1% BSA in PBS) for 1 h at room temperature, washed with ice-cold 1% PBSA, and then incubated with diluted 1:50 PE fluorescently labeled anti-mouse antibody in 1% PBSA for 20 min at 4 °C in the dark. In the same reaction mixtures, the same YSD SCF mutant clones were also labeled for dimerization by incubation with 1 μM of purified $SCF_{WT}$ in 1% PBSA at 37 °C for 1 h, washed with ice-cold 1% PBSA, and incubated with fluorescent APC-conjugated anti FLAG-tag antibody (Biolegend, CA, USA) at 4 °C for 20 min in the dark. For yeast cell fluorescence screening, an Accuri C6 flow cytometer was used, and data analysis was performed using FlowJo software. Each clone was labeled and analyzed in three independent replications. The binding level (geometric mean) with soluble $SCF_{WT}$ of each YSD SCF clone normalized to the latter level of expression (geometric mean) was normalized to the results obtained for YSD $SCF_{WT}$. The YSD clones that were tested in this assay were SCF A20T and SCF N72D, and the controls were $SCF_{WT}$, SCF T9M, and $SCF_{mut}$ (SCF mutant with two mutations, L63V and L49V, having reduced dimerization affinity as shown in Ref. [51]).

### SPR spectroscopy

The binding constants for binding between the purified SCF variants and the c-Kit receptor were determined by SPR spectroscopy on a ProteOn XPR36 (Bio-Rad) instrument. Recombinant human c-Kit (R&D systems, USA) was immobilized on the surface of the chip by using the amine coupling reagents 0.1 M *N*-hydroxysuccinimide (S-NHS) and 0.4 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC; Bio-Rad). c-Kit (3 or 5 μg) was covalently immobilized on the chip in 10 mM sodium acetate buffer (pH 4.5) to give 4638.81 response units (RU). BSA (2 μg; 4323.64 RU) was immobilized on the chip as a negative control. Unbound esters were deactivated with 1 M ethanolamine HCl at pH 8.5. Before each binding assay, the temperature was set at 25 °C. Soluble purified SCF variants were then allowed to

flow over the surface-bound c-Kit at concentrations of 47.5, 23.75, 11.88, 5.94, 2.97, and 0 nM for $SCF_{WT}$, concentrations of 51.5, 25.75, 12.88, 6.44, 3.22, and 0 nM for $SCF_{A20T}$, and concentrations of 1000, 500, 250, 125, 62.5 and 0 nM for $SCF_{T9M}$, $SCF_{N10Y}$, $SCF_{K13T}$, $SCF_{S53L}$, $SCF_{D85N}$, and $SCF_{E88K}$ at a flow rate of 100 μl/min for 216 s. While the SCF variants were flowing over the surface-bound c-Kit, the interactions between them were determined. The next step was to examine the dissociation of the proteins while allowing PBST (phosphate-buffered saline + 0.005% Tween) to flow over the surface for 10 min. After each run, a regeneration step was performed with 50 mM NaOH at a flow rate of 100 μl/min for 18 s. For each protein complex, a sensorgram was generated from the RUs measured during the course of the PPI minus the values of the BSA channel. The $K_D$ was determined from the sensorgram of the equilibrium binding phase.

### Identifying the interface/core/surface SCF residues and calculating the residue-level energy contributions

The atomic model of the SCF/Kit complex was taken from PDB entry 2E9W. SCF interface/core/surface residues were identified by measuring the solvent-accessible [3,62] surface area using the surfv program [63] as follows: SCF/c-Kit interface residues are those with $\geq 5$ Å$^2$ buried in the interface with c-Kit; SCF/SCF dimer interface residues are those with $\geq 5$ Å$^2$ buried in the homodimer interface; SCF core residues are those that expose $\leq 5$ Å$^2$ to the solvent in the SCF monomer, and surface SCF residues include all other residues.

Quantitative electrostatic calculations to determine per-residue contributions were performed in details in Ref. [56]. Prior to the calculations, hydrogen atoms were added using CHARMM, and the structures were subjected to conjugate gradient minimization with a harmonic restraint force of 50 kcal mol$^{-1}$ Å$^{-2}$ applied to the heavy atoms. We calculated the net electrostatic/polar energetic contributions ($\Delta\Delta G_{elec}$) of each residue to the interaction with the cognate protein partner using an unorthodox variant of *in silico* mutagenesis, perturbing the charges of each residue. We neutralized either a residue's backbone and side chain or a residue's side chain alone, thereby differentiating between side-chain *versus* main-chain energetic contributions. We calculated the energetic contributions ($\Delta\Delta G_{elec}$) of all SCF residues within 15 Å of the SCF/c-Kit or the SCF/SCF interfaces to interactions with their relevant protein partner (c-Kit or the second SCF monomer) by using the FDPB method as implemented in DelPhi software [64]. Since the PDB entry contains two copies of SCF and two copies of c-Kit, we repeated the calculations for every chain and compared the results so as to minimize false negatives.

### *In silico* saturation mutagenesis of the SCF/c-Kit interface and SCF/SCF interface

The positions of c-Kit bearing experimentally found mutations that were within 5 Å of SCF in the SCF/c-Kit complex structure [49] were used to computationally model the effect of single mutations on the affinity of the complex (positions 9, 10, 11, 13, 53, 82, 85, and 88). Chains B and D in the PDB entry 2E9W were used as input, and hydrogen atoms were added by MolProbity [65]. For each of the SCF binding interface positions, we defined a set of shell positions including the residues on SCF that were within 4 Å of the considered SCF position and the residues on c-Kit that were within 3 Å of the considered position. For each binding interface position, we then performed 18 calculations, when the considered position on SCF was either kept as wild type or replaced with another amino acid (except for proline, cysteine, or glycine). The *in silico* saturation mutagenesis protocol is described in detail in our previous work [24,25]. Briefly, during the calculation, the shell and the interface residues were repacked, and the energy of the SCF/c-Kit complex was calculated for the wild-type and mutated complex. We then separated the two chains and calculated the energy of each unbound chain without the further repacking of the side chains. The intermolecular energy was calculated by subtracting the energies of the single chains (SCF and c-Kit) from the total energy of the complex [see Eqs. (1) and (2)].

$$\Delta G = E_{\text{complex}} - E_{\text{single chains}} \quad (1)$$

$\Delta\Delta G_{bind}$ was calculated by subtracting the intermolecular energy of the wild-type complex from that of the mutant complex.

$$\Delta\Delta G_{\text{bind}} = \Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}} \quad (2)$$

Finally, the obtained $\Delta\Delta G_{bind}$ was normalized according to a linear equation obtained in our previous work [30]. Rotamer libraries were based on the backbone-dependent library of Dunbrack and Karplus [66] with additional rotamers expanded by one standard deviation around their mean χ1 and χ2 values.

For the SCF dimer, the same protocol was used. All the positions in the SCF dimerization interface bearing experimentally found mutations were used for computational modeling (positions 20, 21, 25, 26, 62, 71, and 72). Chains C and D in the PDB entry 2E9W were used as input.

### Computing protein stability

The web server of RosettaBackrub was used to model the stability of the c-Kit mutants. The web server uses the backrub method [31] to simulate backbone adjustments due to mutations. Chain D of PDB entry

2E9W [49] and a shell distance of 6 Å were used, and 20 models were generated for each mutation. Additionally, the wild-type amino acid for each position bearing a mutation was modeled. For calculating the protein stability, the model with the lowest Rosetta score was used. The Rosetta score of the wild type was subtracted from the score of the mutant. The stability of all the experimentally found mutations was modeled with this method.

## Acknowledgments

We are grateful to Jason Shirian for helpful comments on the manuscript. The authors thank Dr. Alon Zilka for his technical assistance. FACS and Proteon experiments were performed at the NIBN proteomics unit. This work was supported by the European Research Council "Ideas program" ERC-2013-StG (contract grant number: 336041) to Niv Papo. Julia M. Shifman acknowledges the support from the Israel Science Foundation (1873/15). Mickey Kosloff acknowledges the support by grants from the Israel Science Foundation (grant numbers 1454/13, 1959/13, and 2155/15) and from the Ministry of Science, Technology and Space, Israel, and the Ministry of Foreign affairs, Italy (3-10704).

**Conflict of interest**: The authors declare that they have no conflict of interest with respect to publication of this paper.

## Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.11.018.

*Received 1 September 2016;*
*Received in revised form 15 November 2016;*
*Accepted 22 November 2016*
Available online 25 November 2016

***Keywords:***
protein–protein interactions;
protein engineering;
combinatorial selection;
computational protein design;
binding affinity

***Abbreviations used:***
PPI, protein–protein interaction; YSD, yeast surface display; FACS, fluorescence-activated cell sorting; SCF, stem cell factor; $SCF_{WT}$, SCF wild-type; PE, phycoerythrin; SPR, surface plasmon resonance; FDPB, finite-difference Poisson–Boltzmann; LB-Amp, Luria Broth-Ampicillin; BSA, bovine serum albumin; RU, response unit.

## References

[1] G.A. Weiss, et al., Rapid mapping of protein functional epitopes by combinatorial alanine scanning, Proc. Natl. Acad. Sci. 97 (2000) 8950–8954.
[2] Y. Zhao, et al., Mapping protein–protein interactions by affinity-directed mass spectrometry, Proc. Natl. Acad. Sci. 93 (1996) 4020–4024.
[3] P. Chakrabarti, J. Janin, Dissecting protein–protein recognition sites, Proteins: Struct. Funct. Bioinf. 47 (2002) 334–343.
[4] A. Erijman, E. Rosenthal, J.M. Shifman, How structure defines affinity in protein–protein interactions, Proteins 47 (2014) 334–343.
[5] J. Murciano-Calles, et al., Alteration of the C-terminal ligand specificity of the Erbin PDZ domain by allosteric mutational effects, J. Mol. Biol. 426 (2014) 3500–3508.
[6] J. Fernández-Recio, Prediction of protein binding sites and hot spots, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 1 (2011) 680–698.
[7] I. Ezkurdia, et al., Progress and challenges in predicting protein–protein interaction sites, Brief. Bioinform. 10 (2009) bbp021.
[8] R.A. Jordan, et al., Predicting protein–protein interface residues using local surface structural similarity, BMC Bioinf. 13 (2012) 41.
[9] K.L. Morrison, G.A. Weiss, Combinatorial alanine-scanning, Curr. Opin. Chem. Biol. 5 (2001) 302–307.
[10] F. Lefèvre, M.-H. Rémy, J.-M. Masson, Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function, Nucleic Acids Res. 25 (1997) 447–448.
[11] F.F. Vajdos, et al., Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis, J. Mol. Biol. 320 (2002) 415–428.
[12] E.T. Boder, K.D. Wittrup, Yeast surface display for directed evolution of protein expression, affinity, and stability, Methods Enzymol. 328 (2000) 430–444.
[13] G. Chao, et al., Isolating and engineering human antibodies using yeast surface display, Nat. Protoc. 1 (2006) 755–768.
[14] S.A. Gai, K.D. Wittrup, Yeast surface display for protein engineering and characterization, Curr. Opin. Struct. Biol. 17 (2007) 467–473.
[15] J.R. Cochran, et al., Domain-level antibody epitope mapping through yeast surface display of epidermal growth factor receptor fragments, J. Immunol. Methods 287 (2004) 147–158.
[16] G. Chao, J.R. Cochran, K.D. Wittrup, Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display, J. Mol. Biol. 342 (2004) 539–550.
[17] J. Mata-Fink, et al., Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast, J. Mol. Biol. 425 (2013) 444–456.
[18] R. Levy, et al., Fine and domain-level epitope mapping of botulinum neurotoxin type A neutralizing antibodies by yeast surface display, J. Mol. Biol. 365 (2007) 196–210.
[19] E.T. Boder, K.D. Wittrup, Yeast surface display for screening combinatorial polypeptide libraries, Nat. Biotechnol. 15 (1997) 553–557.
[20] J.J. Van Antwerp, K.D. Wittrup, Fine affinity discrimination by yeast surface display and flow cytometry, Biotechnol. Prog. 16 (2000) 31–37.

[21] F. Fack, et al., Epitope mapping by phage display: random versus gene-fragment libraries, J. Immunol. Methods 206 (1997) 43–52.
[22] W. Tanner, L. Lehle, Protein glycosylation in yeast, Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes 906 (1987) 81–99.
[23] S.A. Brooks, Appropriate glycosylation of recombinant proteins for human use, Mol. Biotechnol. 28 (2004) 241–255.
[24] O. Sharabi, A. Erijman, J.M. Shifman, Computational methods for controlling binding specificity, Methods Enzymol. 523 (2013) 41–59.
[25] O. Sharabi, J. Shirian, J.M. Shifman, Predicting affinity- and specificity-enhancing mutations at protein–protein interfaces, Biochem. Soc. Trans. 41 (2013) 1166–1169.
[26] O. Sharabi, et al., Affinity- and specificity-enhancing mutations are frequent in multispecific interactions between TIMP2 and MMPs, PLoS One 9 (2014), e93712.
[27] Y. Aizner, et al., Mapping of the binding landscape for a picomolar protein–protein complex through computation and experiment, Structure 22 (2014) 636–645.
[28] M. Kosloff, E. Alexov, V.Y. Arshavsky, B. Honig, Electrostatic and lipid anchor contributions to the interaction of transducin with membranes: mechanistic implications for activation and translocation, J Biol Chem. 283 (45) (2008 Nov 7) 31197–31207, http://dx.doi.org/10.1074/jbc.M803799200 PMID: 18782760.
[29] F.B. Sheinerman, B. Al-Lazikani, B. Honig, Sequence, structure and energetic determinants of phosphopeptide selectivity of SH2 domains, J. Mol. Biol. 334 (2003) 823–841.
[30] O. Sharabi, A. Dekel, J.M. Shifman, Triathlon for energy functions: who is the winner for design of protein–protein interactions? Proteins: Struct. Funct. Bioinf. 79 (2011) 1487–1498.
[31] C.A. Smith, T. Kortemme, Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction, J. Mol. Biol. 380 (2008) 742–756.
[32] J. Schymkowitz, et al., The FoldX web server: an online force field, Nucleic Acids Res. 33 (2005) W382–W388.
[33] L. Rosenfeld, et al., Combinatorial and computational approaches to identify interactions of macrophage colony-stimulating factor (M-CSF) and its receptor c-FMS, J. Biol. Chem. 290 (2015) 26,180–26,193.
[34] N.S. Yee, I. Paek, P. Besmer, Role of kit-ligand in proliferation and suppression of apoptosis in mast cells: basis for radiosensitivity of white spotting and steel mutant mice, J. Exp. Med. 179 (1994) 1777–1787.
[35] D. Kent, et al., Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway, Clin. Cancer Res. 14 (2008) 1926–1930.
[36] B. Wehrle-Haller, The role of Kit-ligand in melanocyte development and epidermal homeostasis, Pigment Cell Res. 16 (2003) 287–296.
[37] P. Rossi, et al., Role of c-kit in mammalian spermatogenesis, J. Endocrinol. Investig. 23 (2000) 609–615.
[38] J. Matsui, et al., Stem cell factor/c-kit signaling promotes the survival, migration, and capillary tube formation of human umbilical vein endothelial cells, J. Biol. Chem. 279 (2004) 18,600–18,607.
[39] A.V. Das, et al., Identification of c-Kit receptor as a regulator of adult neural stem cells in the mammalian eye: interactions with Notch signaling, Dev. Biol. 273 (2004) 87–105.
[40] I. Ahmad, A.V. Das, J. James, S. Bhattacharya, X. Zhao, Neural stem cells in the mammalian eye: types and regulation, Semin. Cell Dev. Biol. 15 (1) (2004) 53–62 Review. PMID: 15036208.
[41] V.C. Broudy, Stem cell factor and hematopoiesis, Blood 90 (1997) 1345–1364.
[42] G. D'Amato, et al., Update on the biology and therapy of gastrointestinal stromal tumors, Cancer Control 12 (2005) 44–56.
[43] T. Hongyo, et al., Specific c-kit mutations in sinonasal natural killer/T-cell lymphoma in China and Japan, Cancer Res. 60 (2000) 2345–2347.
[44] Y. Sakuma, et al., c-kit gene mutations in intracranial germinomas, Cancer Sci. 95 (2004) 716–720.
[45] H. Ikeda, et al., Expression and functional role of the proto-oncogene c-kit in acute myeloblastic leukemia cells, Blood 78 (1991) 2962–2968.
[46] V.C. Broudy, et al., Blasts from patients with acute myelogenous leukemia express functional receptors for stem cell factor, Blood 80 (1992) 60–67.
[47] M.C. Heinrich, et al., Inhibition of c-kit receptor tyrosine kinase activity by STI 571, a selective tyrosine kinase inhibitor, Blood 96 (2000) 925–p. 932.
[48] M.C. Heinrich, et al., Inhibition of KIT tyrosine kinase activity: a novel molecular approach to the treatment of KIT-positive malignancies, J. Clin. Oncol. 20 (2002) 1692–1703.
[49] S. Yuzawa, et al., Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor, Cell 130 (2007) 323–334.
[50] M.A. Smith, C.J. Pallister, J.G. Smith, Stem cell factor: biology and relevance to clinical practice, Acta Haematol. 105 (2001) 143–150.
[51] Y.-R. Hsu, et al., The majority of stem cell factor exists as monomer under physiological conditions. Implications for dimerization mediating biological activity, J. Biol. Chem. 272 (1997) 6406–6415.
[52] P. Blume-Jensen, et al., Activation of the human c-kit product by ligand-induced dimerization mediates circular actin reorganization and chemotaxis, EMBO J. 10 (1991) 4121.
[53] P. Wollberg, et al., The adapter protein APS associates with the multifunctional docking sites Tyr-568 and Tyr-936 in c-Kit, Biochem. J. 370 (2003) 1033–1038.
[54] M. Kozlowski, et al., SHP-1 binds and negatively modulates the c-Kit receptor by interaction with tyrosine 569 in the c-Kit juxtamembrane domain, Mol. Cell. Biol. 18 (1998) 2089–2099.
[55] D.S. Jones, P.-C. Tsai, J.R. Cochran, Engineering hepatocyte growth factor fragments with high stability and activity as Met receptor agonists and antagonists, Proc. Natl. Acad. Sci. 108 (2011) 13,035–13,040.
[56] M. Kosloff, et al., Integrating energy calculations with functional assays to decipher the specificity of G protein–RGS protein interactions, Nat. Struct. Mol. Biol. 18 (2011) 846–853.
[57] S. Macauley-Patrick, et al., Heterologous protein production using the *Pichia pastoris* expression system, Yeast 22 (2005) 249–270.
[58] S. Park, et al., Limitations of yeast surface display in engineering proteins of high thermostability, Protein Eng. Des. Sel. 19 (2006) 211–217.
[59] T.A. Whitehead, et al., Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing, Nat. Biotechnol. 30 (2012) 543–548.
[60] T. Arakawa, et al., Glycosylated and unglycosylated recombinant-derived human stem cell factors are dimeric and have extensive regular secondary structure, J. Biol. Chem. 266 (1991) 18,942–18,948.

[61] D.R. Higgins, J.M. Cregg, Pichia Protocols, vol. 103, Springer, 1998.
[62] C. Chothia, J. Janin, Principles of protein–protein recognition, Nature 256 (1975) 705–708.
[63] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, Proteins 11 (1991) 281–296.
[64] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry, Science 268 (1995) 1144–1149.
[65] V.B. Chen, et al., MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. Sect. D: Biol. Crystallogr. 66 (2010) 12–21.
[66] R.L. Dunbrack, M. Karplus, Backbone-dependent rotamer library for proteins. Application to side-chain prediction, J. Mol. Biol 20 (1993) 543–574.