

Combinatorial and Computational Approaches to Identify Interactions of Macrophage Colony-stimulating Factor (M-CSF) and Its Receptor c-FMS*

Received for publication, June 9, 2015, and in revised form, September 6, 2015. Published, JBC Papers in Press, September 10, 2015, DOI 10.1074/jbc.M115.671271

Lior Rosenfeld[‡], Jason Shirian[§], Yuval Zur^{‡¶}, Noam Levaot[¶], Julia M. Shifman^{§1}, and Niv Papo^{§‡2}

From the [‡]Department of Biotechnology Engineering and the National Institute of Biotechnology in the Negev, and the

[¶]Department of Physiology and Cell Biology, Ben-Gurion University of the Negev, Beer-Sheva 8410501 and the [§]Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem 9190401, Israel

Background: Identifying residues crucial for M-CSF-c-FMS binding remains difficult.

Results: Using a combination of experimental and computational methods, we identified mutations on M-CSF that reduce affinity to c-FMS.

Conclusion: Affinity-reducing mutations are located both inside and outside of the binding interface.

Significance: Knowledge of the critical residues will facilitate a better understanding of the M-CSF mechanism and facilitate drug design.

The molecular interactions between macrophage colony-stimulating factor (M-CSF) and the tyrosine kinase receptor c-FMS play a key role in the immune response, bone metabolism, and the development of some cancers. Because no x-ray structure is available for the human M-CSF-c-FMS complex, the binding epitope for this complex is largely unknown. Our goal was to identify the residues that are essential for binding of the human M-CSF to c-FMS. For this purpose, we used a yeast surface display (YSD) approach. We expressed a combinatorial library of monomeric M-CSF (M-CSF_M) single mutants and screened this library to isolate variants with reduced affinity for c-FMS using FACS. Sequencing yielded a number of single M-CSF_M variants with mutations both in the direct binding interface and distant from the binding site. In addition, we used computational modeling to map the identified mutations onto the M-CSF_M structure and to classify the mutations into three groups as follows: those that significantly decrease protein stability; those that destroy favorable intermolecular interactions; and those that decrease affinity through allosteric effects. To validate the YSD and computational data, M-CSF_M and three variants were produced as soluble proteins; their affinity and structure were analyzed; and very good correlations with both YSD data and computational predictions were obtained. By identifying the M-CSF_M residues critical for M-CSF-c-FMS interactions, we have laid down the basis for a deeper understanding of the M-CSF-c-FMS signaling mechanism and for the development of target-specific therapeutic agents with the ability to sterically occlude the M-CSF-c-FMS binding interface.

Ligand-receptor interactions are fundamental to many processes occurring in living organisms. The evolution of these protein-protein interactions (PPIs)³ is dependent on the development of highly specific sites of interaction within the ligand-receptor complex. A comprehensive understanding of how specific PPIs evolve requires the precise identification of the binding recognition sites on the protein agonists (1). Ligand-receptor recognition depends mainly on the physical and chemical properties of the binding interfaces of the two protein surfaces that become buried upon complex formation. A number of studies on the properties of the binding interfaces have revealed that non-covalent interactions, such as hydrogen bonding, burial of the hydrophobic area, and van der Waals interactions collectively determine the binding affinity between the proteins (2, 3). This binding affinity may be affected by mutations in one or both of the binding partners. In addition to residues lying in direct binding interfaces, mutations of more distant residues can also influence the binding strength of ligand-receptor complexes, either through allosteric effects (4) or through changes in the folding of the ligand protein.

Identification of the residues most crucial for binding of a particular ligand-receptor complex is important not only for understanding the evolution of inter-molecular interactions but also for the design of therapeutic molecules directed at inhibiting these interactions. However, the identification of such residues is frequently a difficult task, especially when a high resolution structure of the ligand-receptor complex is not available. In the absence of a high resolution structure, such residues have to be identified experimentally (*i.e.* through epitope mapping). One possible technique for such identification is alanine scanning mutagenesis, in which each residue, in turn, is replaced with Ala, and the change in the free energy of binding ($\Delta\Delta G_{\text{binding}}$) is measured (5, 6). However, such a pro-

* This work was supported in part by the European Research Council "Ideas Program" ERC-2013-StG Contract Grant 336041 (to N.P.). The authors declare that they have no conflicts of interest with the contents of this article.

¹ Supported by Israel Science Foundation Grant 1372/10.

² To whom correspondence should be addressed: Dept. of Biotechnology Engineering, Ben-Gurion University of the Negev, 1 Ben Gurion Ave., Beer-Sheva 8410501, Israel. Tel.: 972-8-6428863; Fax: 972-8-6428875; E-mail: papo@bgu.ac.il.

³ The abbreviations used are: PPI, protein-protein interaction; M-CSF_M, monomeric variant of M-CSF; PE, phycoerythrin; FITC, fluorescein isothiocyanate; RU, response unit; SPR, surface plasmon resonance; YSD, yeast surface display; amp, ampicillin.

cess is time-consuming and laborious, involving protein production and purification and binding affinity measurements for each mutant. Better alternatives lie in approaches based on various display technologies that allow for a quick exploration of all possible mutations and qualitative mapping of the contribution of each residue to the binding affinity (7–9).

In this study, we chose to perform epitope mapping by using the yeast surface display (YSD) technology (10–12), because this technique facilitates both rapid and quantitative library screening by fluorescence-activated cell sorting (FACS) and the minimization of artifacts resulting from host-expression bias through concurrent expression labeling (12, 13). For epitope mapping, in particular, there are two important advantages of YSD as follows: there is no need to produce and purify protein variants (11), and in contrast to phage display, the technique is suitable for eukaryotic proteins, preserving their folds and allowing glycosylation, albeit through slightly different sugar compositions compared with those of humans (15, 16). Indeed, epitope mapping by YSD has been successfully performed for epidermal growth factor receptor (17) and gp120 (18) as model systems and has also been used to map the neutralizing antibodies of botulinum neurotoxin (19).

Alongside experimental approaches, binding epitopes have also been mapped by computational means. Recently, Sharabi *et al.* (20, 21) developed a computational saturation mutagenesis protocol that introduces all possible single mutations into the binding interface and predicts $\Delta\Delta G_{\text{binding}}$ due to such mutations. This computational protocol, although successful for predicting the effect of mutations on experimental $\Delta\Delta G_{\text{binding}}$ in several PPIs, requires a high resolution three-dimensional structure of the ligand-receptor complex and can predict only the effect of residues within the ligand-receptor interface (22–24). In addition to the protocol of Sharabi *et al.* (20, 21), there are a number of other protocols that can predict changes in the thermodynamic stability of proteins due to mutations, which may result in ligand unfolding and consequently in a reduction in target binding (25–27). Nevertheless, computational and experimental combinatorial epitope mapping approaches are not mutually exclusive, and their combination is a powerful strategy for studying PPIs. In this study, our combinatorial approach was complemented with computational modeling with the aim to analyze and explain the effects of mutations identified by experimental YSD.

For the purpose of this study, we chose to investigate the important complex between the ligand, macrophage colony-stimulating factor (M-CSF), and the tyrosine kinase receptor, c-FMS, whose epitope is not known due to the absence of the x-ray structure of the complex. The interactions between M-CSF and c-FMS play key roles in the immune response and in bone metabolism. M-CSF is the most pleiotropic macrophage growth factor, regulating macrophage survival, proliferation, differentiation, and chemotaxis (28–30). Furthermore, M-CSF together with the receptor activator of the NF- κ B ligand are both essential and sufficient to induce the differentiation of mononuclear monocytes into osteoclasts, the bone-resorbing cells (31, 32). In this context, the significance of the interaction of M-CSF with c-FMS is clearly evident in both M-CSF and c-FMS-deficient mice, which suffer from retarded

skeletal growth and osteopetrosis as a result of diminished osteoclast function (33–35). Moreover, the M-CSF-c-FMS complex plays an important role in cancer (36). It was found that M-CSF is linked to macrophage recruitment in tumors, thus promoting their survival (37–39). Indeed, the M-CSF-c-FMS complex was shown to be involved in the development and survival of many cancer types, including breast cancer (39, 40), prostate cancer (41), melanoma (42), and bone metastasis (39).

The human M-CSF ligand consists of a signal peptide, a 149-amino acid receptor-binding domain, a variable spacer region, and a transmembrane region with a cytoplasmic tail (29). The human M-CSF receptor (c-FMS) is a 972-amino acid glycosylated protein that includes an extracellular region composed of five immunoglobulin-like domains (D1–D5), a transmembrane region, and an intracellular autocatalytic kinase domain (29). The bivalent binding of the soluble M-CSF covalent homodimer to the D2 and D3 extracellular domains of c-FMS results in receptor homodimerization through the membrane-proximal D4 and D5 domains of c-FMS and activation of the M-CSF-c-FMS signal transduction cascade, resulting in rapid stimulation of gene transcription, protein translation, and cytoskeletal remodeling (28, 29, 35, 41, 43).

Identifying the M-CSF residues that are crucial for its binding to c-FMS is important for understanding the mechanism of the diseases caused by misregulation of the M-CSF-c-FMS signaling pathway (29, 34, 44) and could assist in the development of drugs targeting this interaction. In this study, we used YSD to map M-CSF positions where mutations result in a significant decrease in the affinity of M-CSF for c-FMS. With purified proteins and computational modeling, we were able to further explain the mechanism by which the identified mutations are predicted to affect the binding affinity of the complex.

Experimental Procedures

Constructing a DNA Library in Yeast—The gene encoding for the secreted isoform of monomeric human M-CSF (M-CSF_M), which is 474 nucleotides long, was purchased from Genscript. To produce M-CSF_M, a mutation was inserted at position 31 to convert cysteine to serine so as to prevent dimerization of the monomers. Homology sites to pCTCON (45) and recognition sites for the restriction enzymes EcoRI and AvrII were inserted in the beginning and the end of the gene, respectively. The gene was amplified using a PCR with Phusion DNA polymerase (New England Biolabs) with primers that fit pCTCON homology (pCTCON forward, 5'-TAAGGACAATAGCTCGACG-ATTGAAG-3', and pCTCON reverse, 5'-CAGATCTCGAG-CTATTACAAGTCCTCTTC-3'). Thereafter, the gene was transformed into a competent EBY100 *Saccharomyces cerevisiae* yeast strain, along with a linear pCTCON plasmid, by using a MicroPulser electroporator (Bio-Rad) for the sake of homologous recombination, as described previously by Wittrup and co-workers (11); the resultant yeast contained the gene for the monomeric version of M-CSF, designated M-CSF_M. The plasmid was extracted from the yeast using ZymoprepTM yeast plasmid miniprep I kit (Zymo Research) and was amplified again using a PCR with Phusion enzyme (primers: SGA forward 5'-CAGTAACGTTTG-TCAGTAATTGCG-3', and SGA reverse 5'-GTGTAAAGTTG-

Identifying Residues Important in M-CSF-c-FMS Interactions

GTAACGGAACG-3'). The PCR product was amplified and mutated using the GeneMorph II random mutagenesis kit (Agilent Technologies). The conditions of the PCRs for constructing the three libraries, having different mutation frequencies, were set according to the kit protocol, using DNA polymerase, dNTP mix, and reaction buffer that were supplied with the kit. The primers for the reaction were pCTCON forward and reverse. Different amounts of template DNA (500, 750, and 1000 ng) were inserted into each 50- μ l PCR, which had 30 amplification cycles with the aim to obtain different mutation rates in the libraries. The gene libraries were amplified, using another PCR (primers: pCTCON inner forward, 5'-GTTCCAGACTACGCTCTGCAGG-3', and pCTCON reverse) using Phusion DNA polymerase, to produce ~ 30 μ g of DNA insert and were transformed into competent *S. cerevisiae* yeast strain EBY100. The transformed yeast was grown on SDCAA plates (0.54% Na₂HPO₄, 0.856% Na₂HPO₄·H₂O, 18.2% sorbitol, 1.5% agar, 2% dextrose, 0.67% yeast nitrogen base, 0.5% bacto casamino acids), with dilution plating to determine library size, and in SDCAA medium (2% dextrose, 0.67% yeast nitrogen base, 0.5% bacto casamino acids, 1.47% sodium citrate, 0.429% citric acid monohydrate, pH 4.5) (11). The libraries were combined to produce a library with 6×10^4 variants. To determine whether there were enough single point mutation variants to cover all possibilities, 90 different plasmids were extracted from the bacteria using HiYield plasmid mini kit (RBC Bioscience, Taiwan) and then sequenced (sequencing laboratory of the National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel). Out of the sequenced plasmids, 29% contained a single point mutation; thus, the library contained 17,400 single point mutation variants at the DNA level, *i.e.* more than an order of magnitude larger than all the possibilities for single point mutations, which are three times the number of nucleotides in the gene. The M-CSF_M gene library was inserted into the pCTCON plasmid, containing all the genetic information for expression of proteins in the YSD system (10), which was subsequently cultivated in SGCAA medium (2% galactose, 0.67% yeast nitrogen base, 0.5% bacto casamino acids, 1.47% sodium citrate, 0.429% citric acid monohydrate) (11), which induces the expression of the YSD system (45) (Fig. 1A).

Flow Cytometry Screening and Sorting—The yeast containing the gene for M-CSF_M was grown in SGCAA medium overnight, until the culture reached $A_{600} = 10$ (10^8 cells per ml). Cells were then collected and washed with 1 ml of PBSA (phosphate-buffered saline (PBS) + 1% of bovine serum albumin (BSA)). The expression of the displayed proteins was detected by labeling the cells with 9E10 mouse anti-c-Myc antibody (Abcam) in a 1:50 ratio, followed by sheep anti-mouse antibody conjugated to phycoerythrin (PE) (Sigma) in a 1:50 ratio. The binding of displayed M-CSF_M to the five extracellular domains of the human c-FMS receptor was detected with 1 nM soluble c-FMS conjugated to human Fc (R&D Systems), followed by goat anti-human Fc conjugated to fluorescein isothiocyanate (FITC) in a 1:50 dilution (Sigma). The labeling process was performed according to the protocol previously described by Boder and Wittrup (10). The yeast displaying the labeled proteins were screened using an Accuri C6 flow cytometry analyzer (BD Biosciences). The horizontal axis provides an indication of the

expression of the displayed protein (PE stain) and the vertical axis the binding of the presented protein to the target receptor (FITC stain). The same labeling procedure was used for the M-CSF_M library and for the individual selected clones, with the latter experiment being repeated three times. For the flow cytometry M-CSF_M library sorting process, a desired population (see details below) was enriched using a diagonal sorting gate with iCyt Synergy FACS apparatus (Sony Biotechnology). For the YSD-based titration curves, the same labeling process was implemented, and the percent of the binding population of M-CSF_M and the variants V78A, S84P, and L85S was measured at receptor concentrations of 10, 50, 100, and 500 pM, 1, 10, 50, 100, and 500 nM, and 1 μ M. Each measurement was normalized to the same clone's highest binding percent, which was set as 1, to receive a plot of normalized binding percent against receptor concentrations. A non-linear specific binding fit was implemented and all the curves using GraphPad Prism (GraphPad Software). The values of K_D for each protein were calculated using the fit.

DNA Sequencing—To perform DNA sequencing of the yeast library, the plasmid was first extracted from the yeast by using the ZymoprepTM yeast plasmid miniprep I kit (Zymo Research). The extracted plasmid was transformed into competent *Escherichia coli* cells via electroporation with a MicroPulser Electroporator (Bio-Rad). The transformed bacteria were grown overnight on 100 μ g/ml LB-ampicillin (amp) agar plates. The bacterial colonies were picked out and grown in LB-amp medium, and the plasmid was extracted from them with the HiYield plasmid mini kit (RBC Bioscience, Taiwan). The purified plasmids were then sequenced using the Sanger sequencing method (sequencing lab, National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev). The results were analyzed using Geneious R7 (Biomatters, New Zealand).

Protein Expression and Purification—Several variants of M-CSF_M were produced. Proteins were expressed in *Pichia pastoris* strain GS115 (46). The DNA encoding for the M-CSF_M variants was restricted using EcoRI and AvrII enzymes (New England Biolabs) and then ligated into a linear pPIC9K plasmid (Invitrogen), containing the AOX1 promoter for the expression of the gene (47, 48). A FLAG tag was added onto the N terminus of the protein and a His tag onto the C terminus. The ligated plasmid was transformed into *E. coli*, as described under "DNA Sequencing," and plated on LB-amp plates. Colonies were picked out and sequenced. A colony with the correct sequence was grown in LB-amp medium, and the plasmid was extracted from the bacteria. The plasmids were then linearized with SacI restriction enzyme (New England Biolabs) and transformed into electrocompetent *P. pastoris* GS115 cells by using the Multi-Copy *Pichia* expression kit protocol (Invitrogen). The transformed cells were plated on RDB plates (18.6% sorbitol, 2% agar, 2% dextrose, 1.34% yeast nitrogen base, 4×10^{-5} % biotin, and 5×10^{-3} % each of L-glutamic acid, L-methionine, L-leucine, L-lysine, and L-isoleucine) (49) for 48 h, collected, and plated on YPD-G418 (4 mg/ml G418) (49) plates for 48 h for multiple copy screening (50). Colonies were picked out of the plates and grown on 5 ml of BMGY medium (2% peptone, 1% yeast extract, 0.23% K₂H(PO₄), 1.1812% KH₂(PO₄), 1.34% yeast

nitrogen base, $4 \times 10^{-5}\%$ biotin, 1% glycerol) (51) overnight and in 5 ml of BMMY medium (2% peptone, 1% yeast extract, 0.23% $K_2H(PO_4)$, 1.1812% $KH_2(PO_4)$, 1.34% yeast nitrogen base, $4 \times 10^{-5}\%$ biotin, 0.5% methanol) (51) for 3 days, with 0.5% methanol being added each day. The cells were suspended, and overexpression of the desired protein was examined using the Western blot method, with 1:1000 ratio of mouse anti-FLAG primary antibody (Sigma), 1:5000 ratio of anti-mouse secondary antibody conjugated to alkaline phosphatase (Jackson ImmunoResearch), and 2 ml of 5-bromo-4-chloro-3-indolyl phosphate reagent for signal development (Sigma). The culture with the highest recombinant protein expression was grown in 50 ml of BMGY medium overnight and then in 0.5 liters of BMMY medium for 3 days, with 0.5% methanol being added each day. The cell suspension was subjected to two rounds of centrifugation and filtration as follows. The suspension was centrifuged at $3800 \times g$ for 10 min; the supernatant was filtered off using a 0.22- μm vacuum filter; the pH was adjusted to 8.0; and 300 mM NaCl and 5 ml of 1 M imidazole were added for 1 h at 4 °C to give a final concentration of 10 mM imidazole. The protein was then purified on nickel-nitrilotriacetic acid-Sepharose beads (Invitrogen) that bind to the His tag within the protein and eluted with 20 ml of a solution containing 50 mM Tris, pH 7, 300 mM NaCl, 250 mM imidazole, and 10% glycerol. The protein buffer was replaced with PBS by using Vivaspin with a 3-kDa cutoff (Vivaproducts). Sugars linked via N-glycosylation sites were removed from half of the concentrated protein by using 1:10 G5 buffer and 2 μl of endoglycosidase HF enzyme (New England Biolabs). Both glycosylated and non-glycosylated proteins were further purified using AKTA pure 150 FPLC with a Superdex 200 16/600 size-exclusion column (GE Healthcare). The yield of each of the purification processes was 10–27 mg per 0.5 liters of yeast culture, as determined using a NanoDrop spectrophotometer (Thermo Scientific) (extinction coefficient $10,345 \text{ M}^{-1} \text{ cm}^{-1}$), based on protein absorbance at 280 nm. Protein samples were also subjected to mass spectrometry analysis (Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev).

Circular Dichroism Analysis—Structural analysis of the purified proteins was performed using a J-815 CD spectrometer (Jasco, Tokyo, Japan) with a 1-mm path length quartz cuvette. Spectra of 5 μM of purified protein in 400 μl of PBS were obtained at room temperature. The average of three spectra was normalized to obtain ellipticity (degree $\times \text{cm}^2/\text{dmol}$). Data points with a diode voltage $>1000 \text{ V}$ were excluded. For M-CSF_M, the CD spectra were obtained for both glycosylated and non-glycosylated proteins.

Surface Plasmon Resonance (SPR) Spectroscopy—The binding constants between M-CSF_M and the soluble variants and c-FMS receptor, and between M-CSF_M and itself (dimerization assay) were determined by SPR spectroscopy on a ProteOn XPR36 (Bio-Rad). Glycosylated M-CSF_M and non-glycosylated M-CSF_M and its variants were immobilized on the surface of the chip by using the amine coupling reagents sulfo-NHS (0.1 M N-hydroxysuccinimide) and EDC (0.4 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide, Bio-Rad). The proteins (1 μg) were each covalently immobilized on the chip in 10 mM sodium

acetate buffer, pH 4.0, to give 1507, 1960, 2204, 1670, and 2486 response units (RU) for M-CSF_M-glycosylated, M-CSF_M-non-glycosylated, and the V78A, L85S, and S84P variants, respectively. BSA (3 μg ; 4514 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 human c-FMS receptor (Sino Biological, China) was then allowed to flow over the surface-bound M-CSF at concentrations of 4.375, 8.75, 17.5, 35, and 70 nM and a flow rate of 50 $\mu l/\text{min}$ for 6 min 50 s. While the analyte was flowing over the surface, the interactions between M-CSF and c-FMS 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 5 min. After each run, a regeneration step was conducted with 50 mM NaOH at a flow rate of 100 $\mu l/\text{min}$. 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 dissociation constant (K_D) was determined from the sensorgram of the equilibrium binding phase. Glycosylated M-CSF_M (31.25, 62.5, 125, 250, and 500 nM) was allowed to flow through the system to ensure that dimerization of the monomeric variants did not occur.

Construction of a Model for M-CSF-c-FMS Complexes—Because there was no x-ray structure available for human M-CSF in a complex with human c-FMS, we created a model for this complex based on available x-ray structures of each chain, either as an unbound monomer (M-CSF_M) or in a complex with another target (c-FMS). To this end, the structure of the c-FMS monomer was taken from chain C of the structure of human IL-34 bound to human c-FMS (Protein Data Bank code 4DKD (52)), and the structure of the human M-CSF dimer was taken from chains A and B of its x-ray structure (Protein Data Bank code 3UF2 (53)). These separate structures were then superimposed on the murine homolog of the M-CSF-c-FMS complex (Protein Data Bank code 3EJJ (54)) with the aid of PyMOL software. This preliminary superimposed structure was then further refined using the Rosetta docking protocol of the ROSIE server (55, 56). The structure with the lowest total score and the smallest root-mean-square deviation with respect to the murine complex structure was chosen.

Calculation of Changes in Binding Free Energy and Protein Stability—For selected mutants, the effect of the mutation on the local stability and change in free energy of binding was determined *in silico*. For these calculations, we utilized the ORBIT protein design software (26) using the model of the M-CSF_M-c-FMS complex as an input. The binding interface was defined and limited to those residues on M-CSF_M with atoms located within 5 Å from the c-FMS chain. All other residues in M-CSF_M were divided into three groups: core, surface, and boundary, depending on their proximity to the protein surface (26). Core residues are those whose α -carbons are at a distance of 5 Å or greater from the protein surface and β -carbons at a distance of 2 Å or greater from the surface. Surface residues are those residues whose sum of α -carbon and β -carbon distances from the protein surface is less than or equal to 2.7 Å. Boundary residues are all other residues. For calculating the free energy of binding for mutations in the interface, we

Identifying Residues Important in M-CSF·c-FMS Interactions

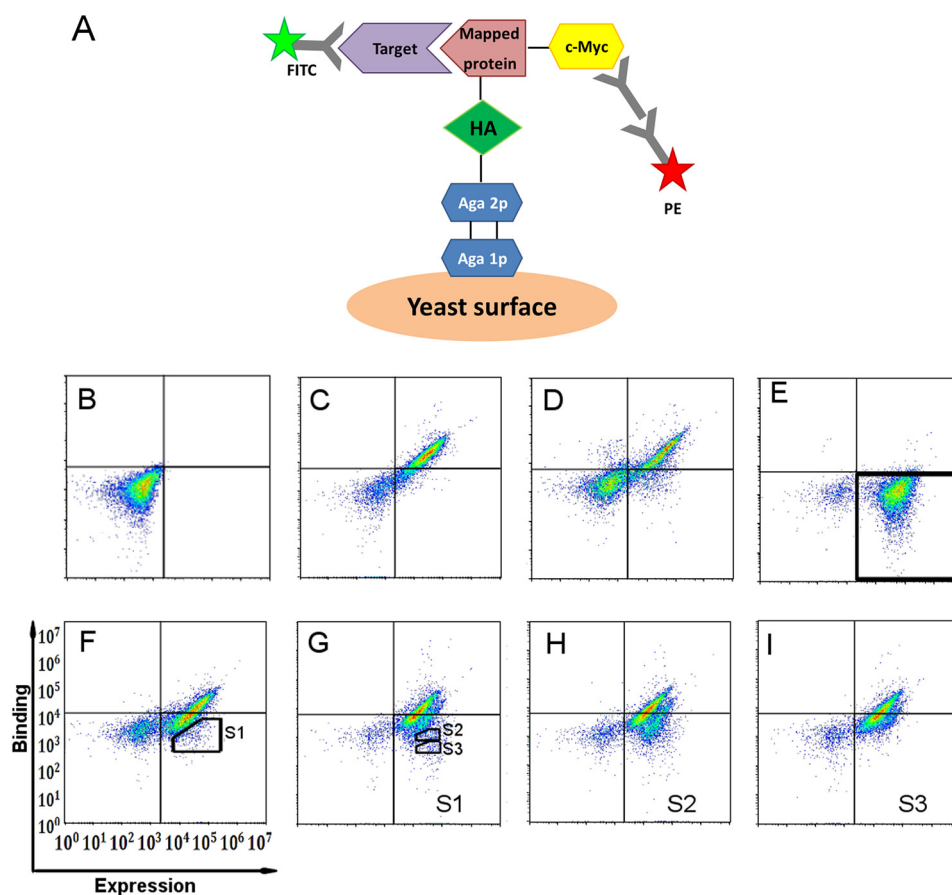


FIGURE 1. YSD and the flow cytometry sorting process. A, schematic representation of the YSD system. The mapped protein (M-CSF_M) is displayed on the yeast, and the soluble target is c-FMS. Flow cytometry screening of M-CSF_M populations throughout the sorting process is presented in B–I. B, negative control, non-labeled cells. C, M-CSF_M bound to the c-FMS target, labeled with both FITC and PE for binding and protein expression analysis, respectively. D, M-CSF_M library, bound to the c-FMS target, labeled with both FITC and PE for protein expression analysis. E, library labeled only with PE for protein expression analysis; enriching the population expressing the YSD proteins. The black square represents the gate used to collect the clones. F, clones were collected inside the black gate to enrich the low binding population, S1. G, clones were collected inside the black gates to enrich two low-binding gated populations, S2 and S3. H, flow cytometry analysis of population S2. I, flow cytometry analysis of population S3. All screening was performed against 1 nM c-FMS receptor.

utilized the energy function optimized by Sharabi *et al.* (57). In the course of the calculation, mutation was introduced into M-CSF_M, a shell around the mutated residue consisting of both the interface residues, and the residues in direct contact with the interface were allowed to repack, and the free energy of the M-CSF_M·c-FMS complex was calculated for the WT and for the mutant complex. The two chains were then separated, and the free energy of each single chain in the WT and the mutant complex was calculated. The change in free energy (ΔG) of the complex was calculated by subtracting the free energy of the WT complex from the free energy of the mutant complex. For calculating ΔG of each individual chain, the free energy of the respective WT chain was subtracted from the free energy of the respective mutant chain. The value of $\Delta\Delta G_{\text{binding}}$ was determined by subtracting the ΔG value of the individual chains from the ΔG value of the complex as a result of the mutation. For calculating changes in the thermodynamic stability of the M-CSF_M mutants, the standard ORBIT energy function (optimized for design of monomeric proteins) was utilized (26); only the M-CSF_M chain was used in the calculation, and residues within 2 Å of the mutant residue were allowed to repack. ΔG was calculated by subtracting the energy of the M-CSF_M struc-

ture from that of the mutant structure. The following variants were excluded from this analysis: E1V, which is not visible in the crystal structure of M-CSF; all the mutations to/from cysteine that were assumed to have changed the protein conformation due to altered disulfide bond formation; and mutations to proline, which cannot be accurately modeled with this method. The energy functions consist of terms that describe van der Waals attractive and repulsive interactions, hydrogen bond interactions, electrostatic interactions, and surface area-based solvation (see Sharabi *et al.* (57)). Rotamer libraries were used, which were based on the backbone-dependent library of Dunbrack and Karplus (58) with additional rotamers expanded by one standard deviation around their mean χ_1 and χ_2 values. The lowest energy rotameric conformation was found using the dead-end elimination theorem (59, 60).

Results

Sorting and Flow Cytometry Analysis of the M-CSF Library and Identification of Individual Variants—To identify the residues critical for binding of M-CSF to c-FMS, we utilized a YSD setup in which M-CSF or its variants were expressed on the surface of the yeast cells, although c-FMS was present in solu-

tion (Fig. 1A). First, we constructed a M-CSF single mutant in which Cys-31 was replaced by Ser to prevent covalent dimerization of the protein on the yeast surface. This mutant, designated M-CSF_M, served as the base for all additional mutations studied in this work. We then confirmed, by SPR spectroscopy of the purified M-CSF_M protein, that M-CSF_M does not associate with itself at concentrations of up to 500 nM, as described under "Production and Characterization of Selected M-CSF_M Variants" of the M-CSF_M proteins. Next, M-CSF_M expression on the yeast surface was monitored by labeling the protein with PE, and binding of M-CSF_M to the receptor was monitored by labeling c-FMS with FITC. To verify the expression and the correct folding of M-CSF_M on yeast, we first analyzed the FACS signal for the interaction between M-CSF_M and c-FMS. Our results showed that the binding interaction between the YSD M-CSF_M and 1 nM soluble c-FMS is indeed detected in the YSD setup with respect to the negative control, in which M-CSF_M is not labeled (Fig. 1, B and C). We next constructed a library that contained predominantly single mutation M-CSF_M variants by using the error-prone PCR method (11). The library was sequenced to verify correct incorporation of mutations and showed that 29% of the sequences indeed contained only single mutations within the gene, and the majority of the remaining sequences contained no mutations. The library diversity of 17,400 single point mutations exceeded the number of 1422 corresponding to all possible single M-CSF_M mutants by more than 10-fold, thus possibly containing most, if not all, point mutants. As a first step of our selection experiments, we screened the constructed M-CSF_M library for binding to c-FMS to verify that the binding signal detected for the M-CSF_M·c-FMS interaction was indeed retained (Fig. 1D). To eliminate the M-CSF_M mutants that were not well expressed on the yeast surface, we pre-sorted the library by collecting only the top 41% of the population with a good expression signal (Fig. 1E). We subsequently sorted the library twice more, each time collecting populations of M-CSF_M variants that exhibited reduced binding signal to the receptor compared with that of the parental M-CSF_M. The first sort was conducted against 1 nM c-FMS, yielding a population enriched with mutants exhibiting high expression and a low binding signal (6.7% of the screened population), compared with that of M-CSF_M (Fig. 1F, S1). Using the selected population as an input, we then performed another sort for decreased binding in the same concentration. In the second sort, we collected the two populations S2 (1.01% of the screened population) and S3 (2.02% of the screened population), with slightly different levels of binding, where binding in S3 is weaker than that in S2 (Fig. 1, H and I, respectively). We subsequently sequenced ~100 clones from each population pool (S1, S2, and S3). Sequencing results showed that population S1 (Fig. 1F) included 59 different clones carrying a total of 41 different mutations, with each clone containing a single unique mutation in the M-CSF_M gene. Population S2 contained only eight variants carrying a single mutation, and population S3 contained 16 point variants (data not shown). The remaining sequences either contained a stop codon or multiple mutations or were missing a part of the gene and were excluded from further analysis. Table 1 summarizes all identified single mutations from S1 population that resulted in decreased binding to

c-FMS together with the associated change in the chemical nature of the residue. As can be seen in Table 1, several of the identified mutations were either to proline or to/from cysteine. It is quite likely that the reduction in binding signal of such M-CSF_M mutants to the receptor was due to a substantial conformational change, to protein misfolding, or to dimerization through an additional disulfide bond (61). Fig. 2A maps the location of the identified affinity-decreasing mutations on the M-CSF_M structure, dividing them into the following four categories according to the homology-modeled structure of the human M-CSF·c-FMS complex: receptor binding interface, surface, boundary, and core (see "Experimental Procedures"). We further screened all the identified M-CSF_M clones individually against 1 nM c-FMS in the YSD setup and plotted their binding signals to c-FMS normalized to that of the parental M-CSF_M. Fig. 2B shows that the majority of the interface, boundary, and core mutations exhibited a greatly decreased YSD binding strength to c-FMS compared with that of M-CSF_M, although surface mutations exhibited a more moderate decrease. In addition, most of the mutations that decrease the affinity to c-FMS were located in the interface and the core of M-CSF_M. We further varied the receptor concentration and performed full titration experiments on YSD for four proteins, M-CSF_M and three of its variants as follows: V78A, S84P, and L85S (Fig. 3). Fitting of the data showed that M-CSF_M binds the c-FMS receptor with the best affinity ($K_D = 0.23$ nM), followed by V78A ($K_D = 1.83$ nM), L85S ($K_D = 8.27$ nM), and S84P ($K_D = 36.43$ nM). These results correlate well with the results shown in Fig. 2.

Computational and Structural Analysis of the M-CSF_M Variants—To investigate whether the decrease in c-FMS binding to the M-CSF_M mutants (shown by the YSD data) was due to the elimination of favorable intermolecular interactions between the two proteins or to destabilization and local or global unfolding of M-CSF_M upon mutation, we computationally modeled the two possibilities for the mutations identified in the experimental YSD setup as crucial for M-CSF_M·c-FMS binding. Because an x-ray structure of the human M-CSF·c-FMS complex is not available, we constructed a model for such a complex structure using the homologous complex for the murine M-CSF_M and c-Fms proteins and the structures of the unbound human M-CSF_M and c-FMS. Applying a previously described protocol (20, 57) on the model of the M-CSF_M·c-FMS complex, we defined the M-CSF_M binding interface and calculated the $\Delta\Delta G_{\text{binding}}$ due to each experimentally identified mutation in the binding interface, excluding S13P (see "Experimental Procedures") (Table 2). Our computational results predicted that most of the identified interfacial mutations will reduce M-CSF_M binding affinity to the receptor. To explore whether the remaining mutations identified by YSD (that do not belong to the binding interface) could lead to significant M-CSF_M destabilization and misfolding, we calculated the change in M-CSF_M stability due to single mutations ($\Delta G_{\text{M-CSF}}$), according to the previously published protocol (see "Experimental Procedures") (57). A mutation was defined as being significantly destabilizing if the $\Delta G_{\text{M-CSF}}$ value was greater than 2 kcal/mol. To obtain a clear picture of the effect of each mutation on M-CSF_M stability, we examined three classes of muta-

Identifying Residues Important in M-CSF_M-c-FMS Interactions

TABLE 1

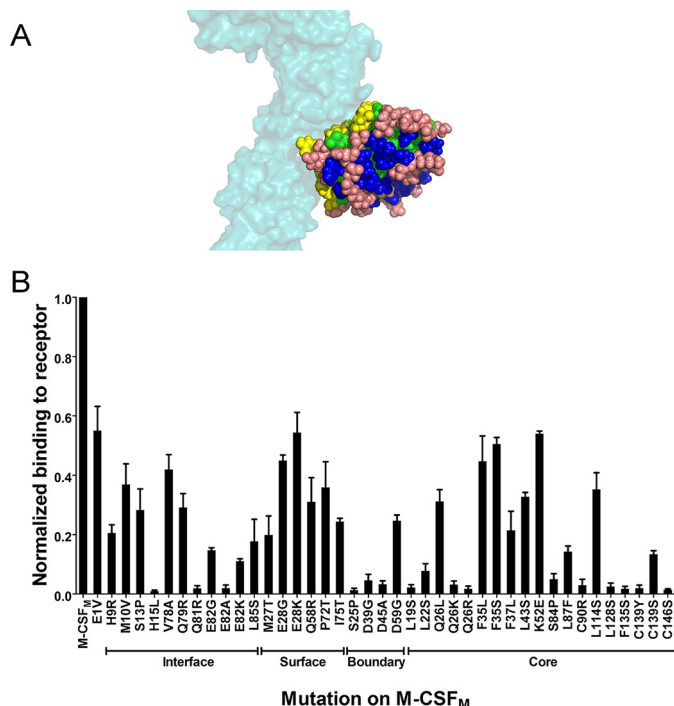
Identified M-CSF_M variants with decreased binding to c-FMS

Position	M-CSF _M residue	Mutated residue	Chemical nature of mutation	No. of repetitions
1	Glu	Val	Negative charge/non-polar	1
9	His	Arg	Positive charge/positive charge	2
10	Met	Val	Non-polar/non-polar	1
13	Ser	Pro	Proline	1
15	His	Leu	Positive charge/non-polar	1
19	Leu	Ser	Non-polar/polar	1
22	Leu	Ser	Non-polar/polar	1
25	Ser	Pro	Proline	2
26	Gln	Lys	Polar/positive charge	1
26	Gln	Leu	Polar/non-polar	1
26	Gln	Arg	Polar/positive charge	1
27	Met	Thr	Non-polar/polar	3
28	Glu	Gly	Negative charge/non-polar	1
28	Glu	Lys	Negative charge/positive charge	1
35	Phe	Ser	Non-polar/polar	1
35	Phe	Leu	Non-polar/non-polar	1
37	Phe	Leu	Non-polar/non-polar	1
39	Asp	Gly	Negative charge/non-polar	1
43	Leu	Ser	Non-polar/polar	1
45	Asp	Ala	Negative charge/non-polar	1
52	Lys	Glu	Positive charge/negative charge	3
58	Gln	Arg	Polar/positive charge	3
59	Asp	Gly	Negative charge/non-polar	2
72	Pro	Thr	Proline	3
75	Ile	Thr	Non-polar/polar	1
78	Val	Ala	Non-polar/non-polar	3
79	Gln	Arg	Polar/positive charge	1
81	Gln	Arg	Polar/positive charge	2
82	Glu	Gly	Negative charge/non-polar	1
82	Glu	Ala	Negative charge/non-polar	1
82	Glu	Lys	Negative charge/positive charge	1
84	Ser	Pro	Proline	2
85	Leu	Ser	Non-polar/polar	2
87	Leu	Phe	Non-polar/non-polar	1
90	Cys	Arg	Cysteine	1
114	Leu	Ser	Non-polar/polar	2
128	Leu	Ser	Non-polar/polar	1
135	Phe	Ser	Non-polar/polar	1
139	Cys	Ser	Cysteine	2
146	Cys	Ser	Cysteine	1
146	Cys	Arg	Cysteine	1

tions separately: surface, boundary, and core. Stability changes for surface mutations were minor except for some positions on the M-CSF_M binding interface (Fig. 4A), although moderate destabilization ($\Delta G_{\text{M-CSF}} = 0-2$ kcal/mol) was observed for most M-CSF_M boundary positions (Fig. 4B). Fig. 4C shows that the majority of the experimentally identified core mutations led to highly predicted M-CSF_M destabilization, supporting the conclusion that the identified core mutations lead to protein unfolding. We then mapped the experimentally identified mutations to the M-CSF_M structure and color-coded them according to their surface accessibility and their effect on protein stability (Fig. 5). For the interface residues, we found that mutations in positions in which the residue was buried more deeply (*i.e.* H15L and M10V) resulted in a higher decrease in stability (Fig. 5A). None of the surface or boundary non-interfacial mutations were found to be highly destabilizing (Fig. 5, B and C). In contrast, most identified mutations located in the core of the protein had marked destabilizing effects on the protein (Figs. 4C and 5D).

Production and Characterization of Selected M-CSF_M Variants—To validate the YSD and computational results, we produced and purified the parental M-CSF_M protein and some representative M-CSF_M variants. Because of the complexity and labor intensity of the production protocol, we selected only three M-CSF_M variants for full characterization, namely L85S,

V78A, and S84P, in which the mutations were located on the helix that directly interacts with the receptor according to the complex model (interface mutations) and that were predicted to affect the binding affinity directly or indirectly through protein misfolding (Fig. 6A). L85S carries a mutation from a hydrophobic to a smaller polar residue and was predicted to decrease the affinity between the two proteins without destabilizing M-CSF_M. V78A was also predicted to exhibit weaker binding to the receptor due to the creation of an unfavorable cavity in the binding interface. The S84P mutation is located in the middle of the helix and is hence likely to damage the structure of M-CSF_M. We first developed the expression and purification protocol using the parental M-CSF_M and then used the same protocol for preparation of the variants. To examine the effects of glycosylation on the structure and binding properties of M-CSF_M, the parental M-CSF_M preparation was divided into two fractions. In the first fraction, the *N*-glycosylations were removed, and in the second fraction, they were allowed to remain intact. The glycosylated and non-glycosylated proteins were then purified by size-exclusion chromatography (Fig. 6B). MS analysis of the non-glycosylated parental M-CSF_M protein gave a molecular mass of 21.6 kDa (Fig. 6C), which is similar to the calculated value of 21.1 kDa for the non-glycosylated protein, while the MS spectra of the glycosylated protein contained peaks corresponding to one glycosylation and two glycosyla-


TABLE 2

Predicted change in binding free energy of M-CSF_M to c-FMS for identified binding interface mutations

Positive $\Delta\Delta G_{\text{binding}}$ values represent affinity-reducing mutations.

Mutation	$\Delta\Delta G_{\text{binding}}$ kcal/mol
V78A	0.55
Q79R	-2.0
M10V	0.83
E82K	-0.32
L85S	1.2
E82A	0.69
Q81R	1.0
H15L	0.48
H9R	-1.0
E82G	1.0

FIGURE 2. Classification and binding analysis of individual M-CSF_M variants to c-FMS. A, M-CSF_M classification to surface (pink, sphere representation), boundary (blue, surface representation), and core (green, mesh representation) as observed in the model of the M-CSF_M-c-FMS complex. The c-FMS receptor is presented in light blue. B, normalized percent of binding in the screening of isolated M-CSF_M variants that express the protein and bind 1 nM c-FMS. All values were normalized to that of M-CSF_M. Values are means $\pm$ S.E., $n = 3$.

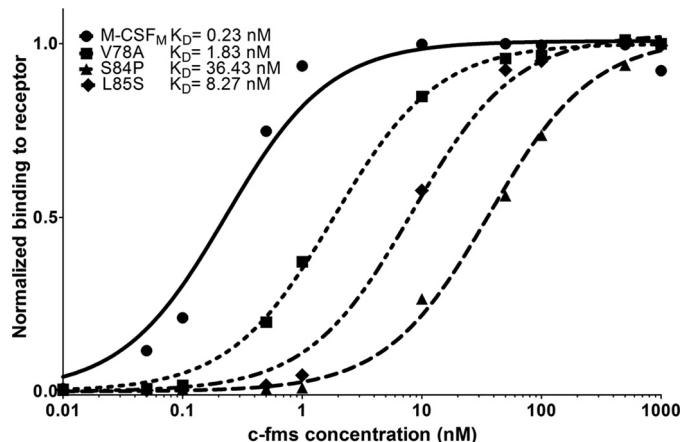


FIGURE 3. YSD titration curves for M-CSF_M and its variants. The binding of M-CSF_M (circles, solid line) and three of its variants V78A (square, dashed line), S84P (triangle, dashed line), and L85S (diamond, dashed line) was measured using FACS at the receptor concentrations of 10 pM to 1 μ M. The binding percent of each sample was normalized to the highest binding signal exhibited by the same variant. The data were fit to a 1:1 binding model to yield the K_D value for each protein.

tions present on the recombinant protein (data not shown). SDS-PAGE confirmed the same size of $\sim$ 21 kDa for both preparations (Fig. 6D). CD spectroscopy of both glycosylated and non-glycosylated proteins (Fig. 7) revealed that the glycosylations did not affect the protein secondary structure. Therefore, we did not remove the glycosylations in the preparation of the M-CSF_M variants. Comparison of the CD spectra of the three M-CSF_M variants to the spectrum of M-CSF_M showed that only

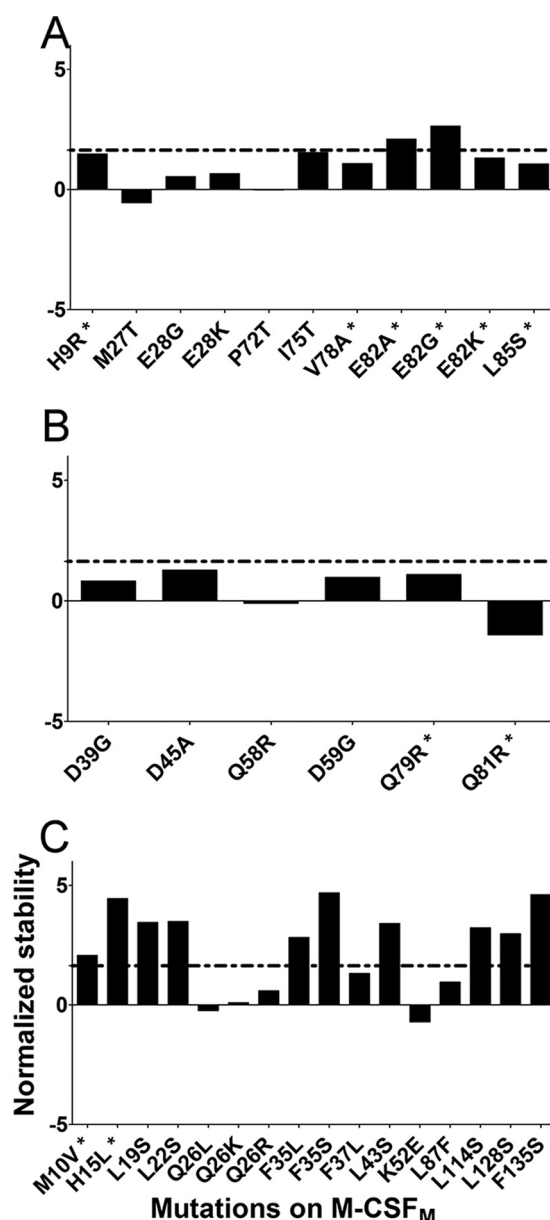


FIGURE 4. In silico affinity and stability analysis of individual clones. All M-CSF_M positions were classified into surface (A), boundary (B), and core (C) sub-domains. The change in M-CSF_M stability due to mutation, ΔG , for each variant is displayed. Any variant with a ΔG value of >2.0 is considered to exhibit significant destabilization. Mutations in the interface are labeled with an asterisk.

Identifying Residues Important in M-CSF_M-c-FMS Interactions

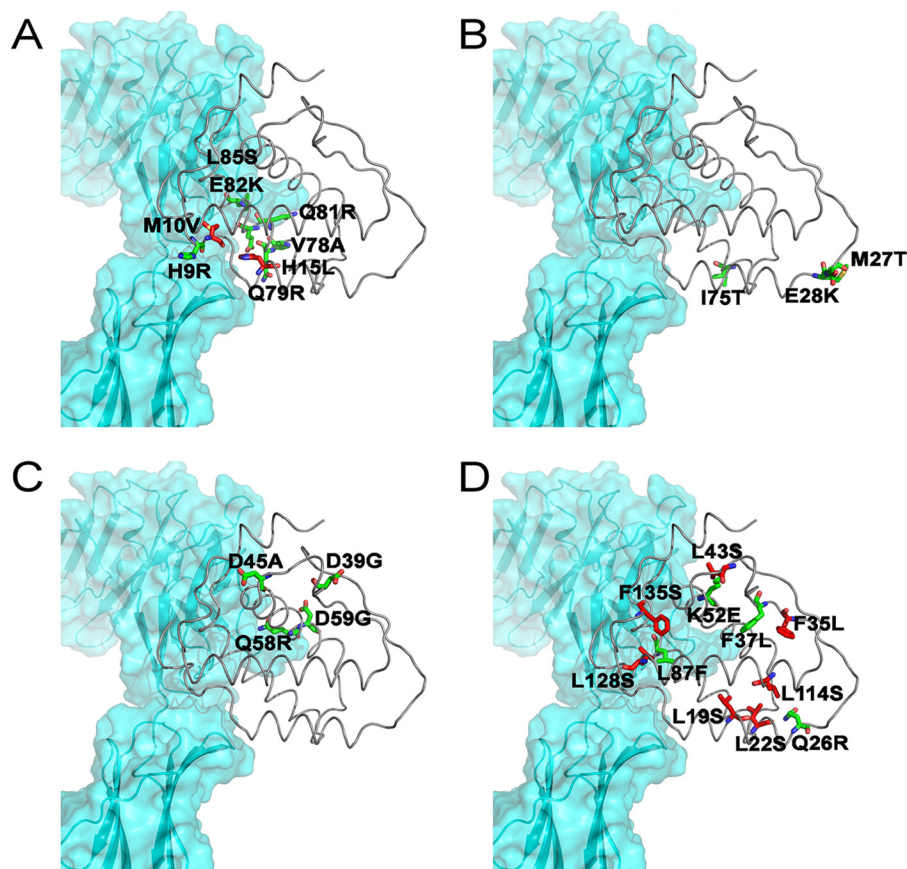


FIGURE 5. M-CSF_M structure showing the identified mutations and their influence on the stability of the protein. In each panel, M-CSF_M is shown in gray schematic, and the identified mutations in the interface (A), surface (B), boundary (C), and core (D) are shown in sticks. c-FMS receptor is colored in light blue. Red residues indicate predicted destabilizing mutations. Green residues indicate predicted structure maintaining mutations. For example, H15L is in the interface and is destabilizing, and H9R is in the interface but does not destabilize the structure.

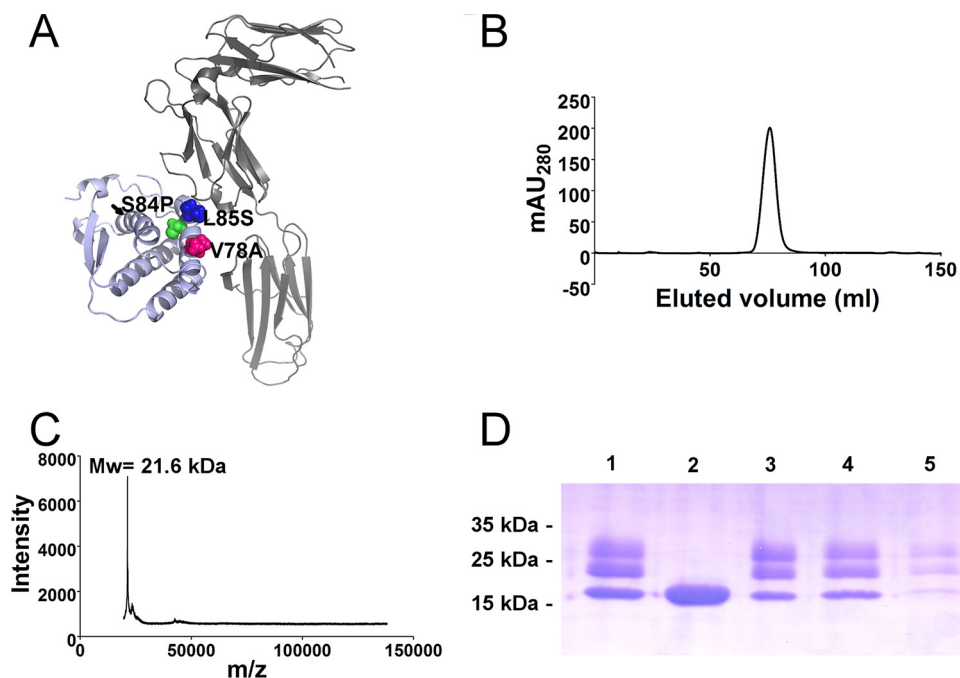


FIGURE 6. **Protein production and purification.** A, three proteins that were chosen for production are displayed on the three-dimensional model for the M-CSF_M-c-FMS complex. The residues of the mutated amino acids in those proteins are represented by spheres. Pink, V78A; green, S84P; blue, L85S. B, size-exclusion chromatography analysis for non-glycosylated M-CSF_M. The same chromatogram was obtained for the glycosylated protein (data not shown). C, mass spectrometry results for non-glycosylated M-CSF_M. D, SDS-PAGE for all produced variants: M-CSF_M-glycosylated (lane 1), M-CSF_M-non-glycosylated (lane 2), V78A (lane 3), L85S (lane 4), and S84P (lane 5).

one variant, L85S, completely retained the structure of M-CSF_M (Fig. 7). V78A, although mostly folded, showed the presence of some unfolded species, as indicated by a more pronounced peak at 210 nm in the CD spectrum. Not unexpectedly, S84P was at least partially unfolded (Fig. 7). To verify that M-CSF_M was indeed a monomer, the binding affinity between two monomers of M-CSF_M was analyzed using SPR at concentrations of 31.25–500 nM. No binding signal was detected (Fig. 8A). Then, to verify that the binding affinities of the produced

mutants were weaker than the binding affinity of M-CSF_M, we measured M-CSF_M-c-FMS receptor interaction using SPR. In such experiments, M-CSF_M (in both glycosylated and non-glycosylated forms) and its variants were immobilized on the surface of the sensor chip, and the c-FMS receptor was added in solution at concentrations of 4.375–70 nM (Fig. 8, B–F). SPR experiments showed that the binding affinities of the glycosylated and non-glycosylated forms of M-CSF_M for c-FMS were very similar, being 42.6 ± 0.92 and 31.6 ± 1.11 nM, respectively (Table 3). The affinity of V78A was reduced by about 2-fold to 77 nM and that for L85S was reduced by about 4-fold to 174 nM. The affinity of S84P could not be measured due to the absence of the binding SPR signal under the conditions of the experiment. Thus, all M-CSF_M variants identified by YSD exhibited reduced binding to the receptor in experiments involving soluble M-CSF_M proteins. In this context, we note that the experimentally measured K_D value correlated well with the differences in the binding percentile by YSD (Fig. 2B), with the K_D values measured by YSD (Fig. 3), and with the computationally predicted values of $\Delta\Delta G_{\text{binding}}$ (Table 3).

Discussion

This work demonstrates a fast approach to identify residues important for the interaction of human M-CSF with its receptor c-FMS, which is a crucial interaction in several diseases, including osteoporosis and some cancers. The demonstrated method, based on a YSD setup, facilitates the identification of all the residues that are important for binding, not only those in the binding interface but also those that are distant from the interaction site yet reduce the binding affinity either through allosteric effects or as a result of protein unfolding. These more

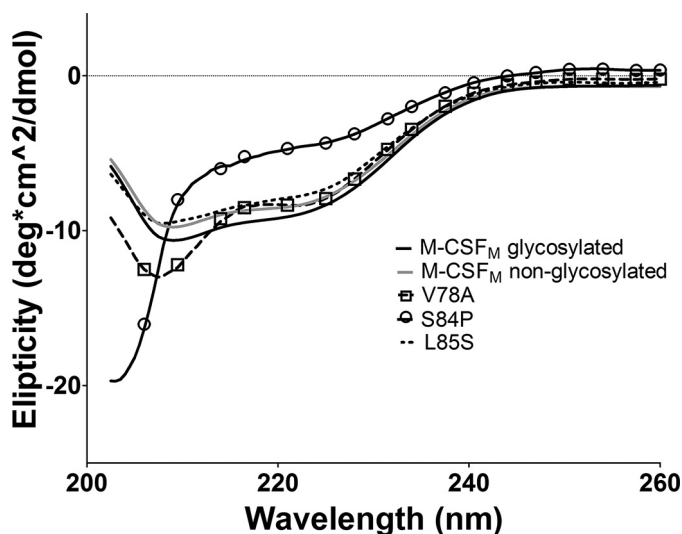


FIGURE 7. CD spectra for the produced proteins. M-CSF_M-glycosylated is shown by black solid line; M-CSF_M-non-glycosylated is shown by gray solid line; V78A is shown by squares and broken line; S84P is shown by circles, and L85S is shown by dotted line. All measurements were performed with 5 μ M protein in PBS, pH 7.4, 25°C.

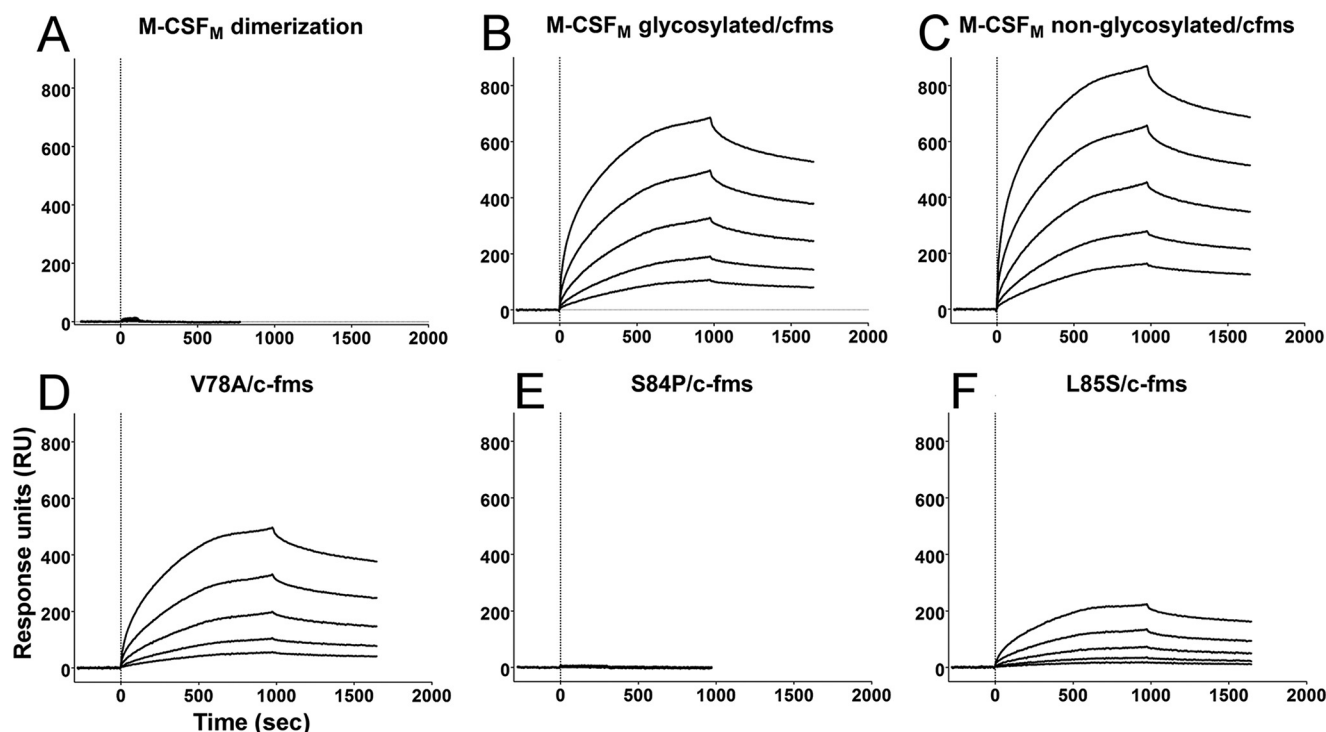


FIGURE 8. SPR binding experiments. Dimerization assay for M-CSF_M-glycosylated is presented in A. The association and dissociation curves of the produced M-CSF_M proteins with c-FMS are presented. M-CSF_M-glycosylated/c-FMS is shown in B; M-CSF_M-non-glycosylated/c-FMS is shown in C; V78A/c-FMS is shown in D; S84P/c-FMS is shown in E; and L85S/c-FMS is shown in F.

TABLE 3

Binding affinities and $\Delta\Delta G_{\text{binding}}$ for the purified M-CSF_M and variants to c-FMS

Variant	K_D (nM) measured by SPR ^a	-Fold decrease in binding compared with soluble glycosylated M-CSF _M measured by SPR	-Fold decrease in binding compared with glycosylated M-CSF _M in YSD binding assay	$\Delta\Delta G_{\text{binding}}$ kcal/mol (SPR) ^b	Predicted $\Delta\Delta G_{\text{binding}}$ kcal/mol ^c
M-CSF _M glycosylated	42.6 ± 0.92	1	1		
M-CSF _M non-glycosylated	31.6 ± 1.11	0.74	ND ^d	-0.18	
V78A	77.5 ± 2.21	1.82	2.38	0.35	0.55
S84P	ND	ND	20	ND	ND
L85S	174 ± 17.2	4.17	5.56	0.83	1.21

^a Measurements were repeated three times and the standard deviation is given.

^b $\Delta\Delta G_{\text{binding}}$ was calculated from experimentally measured K_D as $\Delta\Delta G_{\text{binding}} = -RT \ln(K_D)$.

^c $\Delta\Delta G_{\text{binding}}$ was predicted using our computational protocol outlined under "Experimental Procedures."

^d ND means not determined.

distant residues cannot be identified by other techniques, such as peptide epitope mapping, alanine scanning, or similar approaches that use both combinatorial and rational library design (62). Furthermore, we confirmed the utility of the technique by showing that the YSD results correlate well with the binding affinities measured by SPR for purified proteins in solution (Table 2). The described method may thus be applied for the identification of crucial residues in other PPIs, particularly for protein complexes such as the human M-CSF_M-c-FMS complex, for which there are no high resolution structures or for which laborious and time-consuming expression and purification protocols are required.

The computational modeling performed in this work complemented the experimental results, as it allowed us to predict the binding interface residues and to better understand the nature of the mutations that lead to affinity reduction highlighting the importance of the wild-type residue interaction. The group of mutations located in the binding interface (M10V, H15L, V78A, Q81R, E82A, E82G, and L85S) is predicted to reduce the free energy of binding by destroying some favorable intermolecular interactions with the receptor. The effects produced by these mutations, such as replacing hydrophobic residues with polar residues and vice versa, replacing large residues with smaller ones, and introducing unfavorable electrostatic interactions, have been previously shown to reduce binding affinities in proteins (63, 64). The effects of some of the interface mutations can be explained through structural analysis based on the modeled structure of the complex. The H15L mutation, for example, may cause a loss of a potential polar contact with the M-CSF backbone, destabilizing the interface. In the V78A mutation, the replacement of one of the hydrophobic side chains with a shorter one may reduce hydrophobic interactions with the receptor. The Q81R mutation is clearly unfavorable, as it would position an arginine residue opposite an arginine residue at position 150 of c-FMS. The E82K mutation may create a clash with an arginine or histidine residue at positions 150 and 151, respectively, on c-FMS. The polar-to-hydrophobic E82A mutation may be unfavorable, as the local environment on c-FMS is mainly polar. The E82G mutation in the middle of an α -helix is likely to disturb the local folding of the protein. It is therefore clear that it is critical for residue 82 to be either polar or negatively charged. In the L85S mutant, a hydrophobic residue is replaced with a polar residue opposite a hydrophobic patch on c-FMS. Mutations H9R and Q79R on M-CSF_M are shown to reduce binding affinity experimentally, but computational binding energy evaluations identify them as

affinity-enhancing. We attribute this discrepancy to possible inaccuracies in the three-dimensional orientation of the binding partners in our model of the M-CSF_M-c-FMS complex. According to this model, His-9 and Gln-79 on M-CSF_M are at the periphery of the binding interface and do not directly contact c-FMS. Mutation to large residues such as arginine at these positions is predicted to introduce direct interaction with the receptor, explaining the computationally calculated improvement in binding affinity. These mutations may be able to reduce binding in reality if the binding partners were closer to one another, where such mutations would present steric clashes and would therefore corroborate the experimental results.

Although we cannot predict the effects of mutations distant from the binding site, we speculate that they affect binding through an allosteric mechanism, as was previously observed in various studies (4). The mutations at positions 26, 27, 39, and 45 might have caused such an effect and will be the subject of a follow-up biophysical and structural study.

Our YSD method was able to identify residues in M-CSF that are energetically important for c-FMS receptor binding without prior knowledge of the structure of the M-CSF-c-FMS complex, and this information directed the production of several key M-CSF_M purified mutants to confirm our YSD results. For the two mutants tested *in vitro* (V78A and L85S), changes in the measured binding affinities correlated well with the predicted values. In addition to these mutations, we observed a number of mutations that are very likely to lead to protein unfolding (Fig. 4). Such mutations include all substitutions from cysteine, because they destroy the disulfide bonds crucial for M-CSF_M stability and mutations to/from proline. One such mutation (S84P) did indeed result in at least partial protein unfolding (as determined by CD for the purified protein). In addition, we identified a large number of mutations that were located in the core of M-CSF_M and were predicted to significantly destabilize the M-CSF_M protein, resulting in either global unfolding or local conformational changes that were propagated to the binding site (Fig. 4C). We also observed a number of charge-altering mutations that, although distant from the binding site, would be likely to alter the long range electrostatic interactions and could thus reduce binding by changing the interactions in the encounter complex (65). In addition, it is possible that some of the identified mutations created new glycosylation sites on the protein variants, for example mutants P72T, I75T, L128S, and F135S. However, these glycosylation sites are not in proximity to the receptor, according to our model of the M-CSF_M-c-FMS complex and are hence not likely to affect binding affinity.

Because the generation of the combinatorial library was based on random mutagenesis, mutations that require a change at two consecutive nucleotides bases (for example, Gly/Leu) were less likely to be included in the library. In a set-up such as the one used here, we might have also missed some mutations by sequencing only ~100 clones from each selection pool. In the future, next generation sequencing might help us to obtain a more comprehensive picture of binding hot spots (66).

In a previous site-directed mutagenesis study, Taylor *et al.* (14) discovered four positions that were regarded as important for the M-CSF-c-FMS interaction: 9, 15, 20, and 78. In this study, we confirmed three of these positions, namely 9, 15, and 78, as being the sites of affinity-reducing mutations. At one of these positions, position 78, we confirmed the critical role of the V78A, which had already been identified as a binding hot spot in the alanine scanning mutagenesis experiment of Taylor *et al.* (14). The fact that we were able to isolate the majority of the previously identified positions encourages us to believe that YSD is indeed as effective as and less laborious than alternative methods for epitope mapping.

Our findings regarding the identity of the residues that are important for the M-CSF-c-FMS interaction strongly suggest that therapeutic drugs that bind to binding and non-binding site residues on M-CSF might be beneficial. Such drugs would potentially facilitate disruption of the M-CSF-c-FMS interaction, or alternatively promote a M-CSF conformational change, without the need to sterically occlude the M-CSF-c-FMS binding interface. Given that the crystal structure of the human M-CSF-c-FMS complex is not available and that, according to our findings, the M-CSF interface residues are not the only contributors to ligand binding, the information produced in this study could constitute the first step in the design of therapeutics targeting not only the M-CSF-c-FMS interface but also other regions within the ligand. The identification of these surface, boundary, and core residues important for M-CSF-c-FMS interactions may also aid in the understanding of M-CSF-c-FMS mechanisms of action and suggest new targets for M-CSF inhibition.

Author Contributions—L. R., J. S., J. M. S., and N. P. designed the research. L. R. performed and analyzed the experiments shown in Figs. 2, 3, and 6–8. L. R. and Y. Z. performed and analyzed the experiments shown in Fig. 1. J. S. generated the computational data shown in Fig. 4. L. R. in consultation with J. S. generated Fig. 5. N. L. contributed to writing the Introduction. L. R., J. S., J. M. S., and N. P. wrote the paper and all authors reviewed and approved the final version of the manuscript.

Acknowledgments—We thank Dr. Alon Zilka for technical assistance. FACS and Proteon experiments were performed at the proteomics unit of National Institute for Biotechnology in the Negev.

References

- Zhao, Y., Muir, T. W., Kent, S. B., Tischer, E., Scardina, J. M., and Chait, B. T. (1996) Mapping protein-protein interactions by affinity-directed mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4020–4024
- Chakrabarti, P., and Janin, J. (2002) Dissecting protein-protein recognition sites. *Proteins Struct. Funct. Genet.* **47**, 334–343
- Erijman, A., Rosenthal, E., and Shifman, J. M. (2014) How structure de-

- finer affinity in protein-protein interactions. *PLoS ONE* **9**, e110085
- Murciano-Calles, J., McLaughlin, M. E., Erijman, A., Hooda, Y., Chakravorty, N., Martinez, J. C., Shifman, J. M., and Sidhu, S. S. (2014) Alteration of the C-terminal ligand specificity of the Erbin PDZ domain by allosteric mutational effects. *J. Mol. Biol.* **426**, 1–9
- Lefèvre, F., Rémy, M. H., and Masson, J. M. (1997) Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function. *Nucleic Acids Res.* **25**, 447–448
- Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A., and Sidhu, S. S. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8950–8954
- Cortese, R., Felici, F., Galfre, G., Luzzago, A., Monaci, P., and Nicosia, A. (1994) Epitope discovery using peptide libraries displayed on phage. *Trends Biotechnol.* **12**, 262–267
- Fack, F., Hügler-Dörr, B., Song, D., Queitsch, I., Petersen, G., and Bautz, E. K. (1997) Epitope mapping by phage display: random *versus* gene-fragment libraries. *J. Immunol. Methods.* **206**, 43–52
- Pál, G., Kouadio, J. L., Artis, D. R., Kossiakoff, A. A., and Sidhu, S. S. (2006) Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J. Biol. Chem.* **281**, 22378–22385
- Boder, E. T., and Wittrup, K. D. (2000) Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol.* **328**, 430–444
- Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768
- Gai, S. A., and Wittrup, K. D. (2007) Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **17**, 467–473
- Boder, E. T., and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557
- Taylor, E. W., Fear, A. L., Bohm, A., Kim, S. H., and Koths, K. (1994) Structure-function studies on recombinant human macrophage colony-stimulating factor (M-CSF). *J. Biol. Chem.* **269**, 31171–31177
- Tanner, W., and Lehle, L. (1987) Protein glycosylation in yeast. *Biochim. Biophys. Acta* **906**, 81–99
- Brooks, S. A. (2004) Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system. *Mol. Biotechnol.* **28**, 241–255
- Cochran, J. R., Kim, Y. S., Olsen, M. J., Bhandari, R., and Wittrup, K. D. (2004) Domain-level antibody epitope mapping through yeast surface display of epidermal growth factor receptor fragments. *J. Immunol. Methods* **287**, 147–158
- Mata-Fink, J., Kriegsman, B., Yu, H. X., Zhu, H., Hanson, M. C., Irvine, D. J., and Wittrup, K. D. (2013) Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. *J. Mol. Biol.* **425**, 444–456
- Levy, R., Forsyth, C. M., LaPorte, S. L., Geren, I. N., Smith, L. A., and Marks, J. D. (2007) Fine and domain-level epitope mapping of botulinum neurotoxin type A neutralizing antibodies by yeast surface display. *J. Mol. Biol.* **365**, 196–210
- Sharabi, O., Erijman, A., and Shifman, J. M. (2013) Computational methods for controlling binding specificity. *Methods Enzymol.* **523**, 41–59
- Sharabi, O., Shirian, J., and Shifman, J. M. (2013) Predicting affinity- and specificity-enhancing mutations at protein-protein interfaces. *Biochem. Soc. Trans.* **41**, 1166–1169
- Filchtinski, D., Sharabi, O., Rüppel, A., Vetter, I. R., Herrmann, C., and Shifman, J. M. (2010) What makes Ras an efficient molecular switch: a computational, biophysical, and structural study of Ras-GDP interactions with mutants of Raf. *J. Mol. Biol.* **399**, 422–435
- Aizner, Y., Sharabi, O., Shirian, J., Dakwar, G., Abvraham, O., Risman, M., and Shifman, J. M. (2014) Mapping the binding landscape of a picomolar protein-protein complex through computation and experiment. *Structure* **22**, 1–10
- Sharabi, O., Shirian, J., Grossman, M., Lebendiker, M., Sagi, I., and Shifman, J. (2014) Affinity- and specificity-enhancing mutations are frequent in multispecific interaction between MMP14 and its inhibitor TIMP2. *PLoS ONE* **9**, e93712

25. Schymkowitz, J. W., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F., and Serrano, L. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10147–10152
26. Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) *De novo* protein design: toward fully automated sequence selection. *J. Mol. Biol.* **273**, 789–796
27. Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531
28. Bartelmez, S. H., Bradley, T. R., Bertoncello, I., Mochizuki, D. Y., Tushinski, R. J., Stanley, E. R., Hapel, A. J., Young, I. G., Kriegler, A. B., and Hodgson, G. S. (1989) Interleukin 1 plus interleukin 3 plus colony-stimulating factor 1 are essential for clonal proliferation of primitive myeloid bone marrow cells. *Exp. Hematol.* **17**, 240–245
29. Douglass, T. G., Driggers, L., Zhang, J. G., Hoa, N., Delgado, C., Williams, C. C., Dan, Q., Sanchez, R., Jeffes, E. W., Wepsic, H. T., Myers, M. P., Koths, K., and Jadus, M. R. (2008) Macrophage colony stimulating factor: not just for macrophages anymore! A gateway into complex biologies. *Int. Immunopharmacol.* **8**, 1354–1376
30. Stanley, E. R., and Chitu, V. (2014) CSF-1 receptor signaling in myeloid cells. *Cold Spring Harb. Symp. Quant. Biol.* **6**, 1–21
31. Tushinski, R. J., Oliver, I. T., Guilbert, L. J., Tynan, P. W., Warner, J. R., and Stanley, E. R. (1982) Survival of mononuclear phagocytes depends on a lineage-specific growth factor that the differentiated cells selectively destroy. *Cell* **28**, 71–81
32. Yasuda, H., Shima, N., Nakagawa, N., Yamaguchi, K., Kinoshita, M., Mochizuki, S., Tomoyasu, A., Yano, K., Goto, M., Murakami, A., Tsuda, E., Morinaga, T., Higashio, K., Udagawa, N., Takahashi, N., and Suda, T. (1998) Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3597–3602
33. Wiktor-Jedrzejczak, W., Bartocci, A., Ferrante, A. W., Jr., Ahmed-Ansari, A., Sell, K. W., Pollard, J. W., and Stanley, E. R. (1990) Total absence of colony-stimulating factor 1 in the macrophage-deficient osteopetrotic (op/op) mouse. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4828–4832
34. Yoshida, H., Hayashi, S., Kunisada, T., Ogawa, M., Nishikawa, S., Okamura, H., Sudo, T., Shultz, L. D., and Nishikawa, S. (1990) The murine mutation osteopetrosis is in the coding region of the macrophage colony stimulating factor gene. *Nature* **345**, 442–444
35. Dai, X. M., Ryan, G. R., Hapel, A. J., Dominguez, M. G., Russell, R. G., Kapp, S., Sylvestre, V., and Stanley, E. R. (2002) Targeted disruption of the mouse CSF-1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies and reproductive defects. *Blood* **99**, 111–120
36. Chockalingam, S., and Ghosh, S. S. (2014) Macrophage colony-stimulating factor and cancer: a review. *Tumour Biol.* **35**, 10635–10644
37. Mantovani, A., Sozzani, S., Locati, M., Allavena, P., and Sica, A. (2002) Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. *Trends Immunol.* **23**, 549–555
38. Laoui, D., Van Overmeire, E., De Baetselier, P., Van Ginderachter, J. A., and Raes, G. (2014) Functional relationship between tumor-associated macrophages and macrophage colony-stimulating factor as contributors to cancer progression. *Front. Immunol.* **5**, 1–15
39. Lin, E. Y., Gouon-Evans, V., Nguyen, A. V., and Pollard, J. W. (2002) The macrophage growth factor CSF-1 in mammary gland development and tumor progression. *J. Mammary Gland Biol. Neoplasia.* **7**, 147–162
40. Murray, L. J., Abrams, T. J., Long, K. R., Ngai, T. J., Olson, L. M., Hong, W., Keast, P. K., Brassard, J. A., O'Farrell, A. M., Cherrington, J. M., and Pryer, N. K. (2003) SU11248 inhibits tumor growth and CSF-1R-dependent osteolysis in an experimental breast cancer bone metastasis model. *Clin. Exp. Metastasis.* **20**, 757–766
41. Escamilla, J., Schokrpur, S., Liu, C., Priceman, S. J., Moughon, D., Jiang, Z., Pouliot, F., Magyar, C., Sung, J. L., Xu, J., Deng, G., West, B. L., Bollag, G., Fradet, Y., Lacombe, L., Jung, M. E., Huang, J., and Wu, L. (2015) CSF1 receptor targeting in prostate cancer reverses macrophage-mediated resistance to androgen blockade therapy. *Cancer Res.* **75**, 950–962
42. Sluijter, M., van der Sluis, T. C., van der Velden, P. A., Versluis, M., West, B. L., van der Burg, S. H., and van Hall, T. (2014) Inhibition of CSF-1R supports T-cell mediated melanoma therapy. *PLoS ONE* **9**, e104230
43. Gronski, P., Badziong, W., Habermann, P., List, W., Müllner, H., Neurohr, K. J., Tripiet, D., and Seiler, F. R. (1988) *E. coli* derived human granulocyte-macrophage colony-stimulating factor (rh GM-CSF) available for clinical trials. *Behring Inst. Mitt.* **83**, 246–249
44. Liu, H., Leo, C., Chen, X., Wong, B. R., Williams, L. T., Lin, H., and He, X. (2012) The mechanism of shared but distinct CSF-1R signaling by the non-homologous cytokines IL-34 and CSF-1. *Biochim. Biophys. Acta.* **1824**, 938–945
45. Colby, D. W., Kellogg, B. A., Graff, C. P., Yeung, Y. A., Swers, J. S., and Wittrup, K. D. (2004) Engineering antibody affinity by yeast surface display. *Methods Enzymol.* **388**, 348–358
46. Cregg, J. M., Barringer, K. J., Hessler, A. Y., and Madden, K. R. (1985) *Pichia pastoris* as a host system for transformations. *Mol. Cell. Biol.* **5**, 3376–3385
47. Daly, R., and Hearn, M. T. (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* **18**, 119–138
48. Macauley-Patrick, S., Fazenda, M. L., McNeil, B., and Harvey, L. M. (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* **22**, 249–270
49. Luo, Y., Yuan, Z., Luo, G., and Zhao, F. (2008) Expression of secreted His-tagged S-adenosylmethionine synthetase in the methylotrophic yeast *Pichia pastoris* and its characterization, one-step purification, and immobilization. *Biotechnol. Prog.* **24**, 214–220
50. Cregg, J. M., Cereghino, J. L., Shi, J., and Higgins, D. R. (2000) Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* **16**, 23–52
51. Shi, X., Karkut, T., Chamankhah, M., Altling-Mees, M., Hemmingsen, S. M., and Hegedus, D. (2003) Optimal conditions for the expression of a single-chain antibody (scFv) gene in *Pichia pastoris*. *Protein Expr. Purif.* **28**, 321–330
52. Ma, X., Lin, W. Y., Chen, Y., Stawicki, S., Mukhyala, K., Wu, Y., Martin, F., Bazan, J. F., and Starovasnik, M. A. (2012) Structural basis for the dual recognition of helical cytokines IL-34 and CSF-1 by CSF-1R. *Structure* **20**, 676–687
53. Elegheert, J., Bracke, N., Pouliot, P., Gutsche, I., Shkumatov, A. V., Tarbouriech, N., Verstraete, K., Bekaert, A., Burmeister, W. P., Svergun, D. I., Lambrecht, B. N., Vergauwen, B., and Savvides, S. N. (2012) Allosteric competitive inactivation of hematopoietic CSF-1 signaling by the viral decoy receptor BARF1. *Nat. Struct. Mol. Biol.* **19**, 938–947
54. Chen, X., Liu, H., Focia, P. J., Shim, A. H., and He, X. (2008) Structure of macrophage colony stimulating factor bound to FMS: diverse signaling assemblies of class III receptor tyrosine kinases. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18267–18272
55. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299
56. Lyskov, S., Chou, F. C., Conchúir, S. Ó., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D., Sripakdeevong, P., Borgo, B., Havranek, J. J., Kuhlman, B., Kortemme, T., Bonneau, R., Gray, J. J., and Das, R. (2013) Serverification of molecular modeling applications: the Rosetta Online Server that includes everyone (ROSIE). *PLoS ONE* **8**, e63906
57. Sharabi, O., Dekel, A., and Shifman, J. M. (2011) Triathlon for energy functions: Who is the winner for design of protein-protein interactions? *Proteins* **79**, 1487–1498
58. Dunbrack, R. L., Jr., and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* **230**, 543–574
59. Desmet, J., De Maeyer, M., Hazes, B., and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539–542
60. Gordon, D. B., Hom, G. K., Mayo, S. L., and Pierce, N. A. (2003) Exact rotamer optimization for protein design. *J. Comput. Chem.* **24**, 232–243
61. Chao, G., Cochran, J. R., and Wittrup, K. D. (2004) Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. *J. Mol. Biol.* **342**, 539–550

62. Van Blarcom, T., Rossi, A., Foletti, D., Sundar, P., Pitts, S., Bee, C., Melton Witt, J., Melton, Z., Hasa-Moreno, A., Shaughnessy, L., Telman, D., Zhao, L., Cheung, W. L., Berka, J., Zhai, W., *et al.* (2015) Precise and efficient antibody epitope determination through library design, yeast display, and next generation sequencing. *J. Mol. Biol.* **427**, 1513–1534
63. Thompson, L. C., Walters, J., Burke, J., Parsons, J. F., Armstrong, R. N., and Dirr, H. W. (2006) Double mutation at the subunit interface of glutathione transferase rGSTM1–1 results in a stable, folded monomer. *Biochemistry* **45**, 2267–2273
64. Schreiber, G., and Fersht, A. R. (1995) Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486
65. Selzer, T., Albeck, S., and Schreiber, G. (2000) Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.* **7**, 537–541
66. Doolan, K. M., and Colby, D. W. (2015) Conformation-dependent epitopes recognized by prion protein antibodies probed using mutational scanning and deep sequencing. *J. Mol. Biol.* **427**, 328–340