SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation

Justina Jankauskaite, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio and Iain H. Moal*

1Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius LT-10257, Lithuania, 2Barcelona Supercomputing Center (BSC), Barcelona 08034, Spain, 3Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Utrecht, the Netherlands, 4Institut de Biològica Molecular de Barcelona (IBMB), CSIC, Barcelona 08028, Spain and 5European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge CB10 1SD, UK

*To whom correspondence should be addressed.

Abstract

Motivation: Understanding the relationship between the sequence, structure, binding energy, binding kinetics and binding thermodynamics of protein–protein interactions is crucial to understanding cellular signaling, the assembly and regulation of molecular complexes, the mechanisms through which mutations lead to disease, and protein engineering.

Results: We present SKEMPI 2.0, a major update to our database of binding free energy changes upon mutation for structurally resolved protein–protein interactions. This version now contains manually curated binding data for 7085 mutations, an increase of 133%, including changes in kinetics for 1844 mutations, enthalpy and entropy changes for 443 mutations, and 440 mutations, which abolish detectable binding.

Availability and implementation: The database is available as supplementary data and at https://life.bsc.es/pid/skempi2/.

Contact: moal@ebi.ac.uk

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Protein–protein interactions are central to almost all biological processes, from cellular signal transduction and the assembly of mesoscopic structures such as myofilaments, to viral adhesion and the immune response. Consequently the effects of changes in protein sequence on the structure, thermodynamics and kinetics of protein–protein interactions has wide implications for constraining the permissible substitutions that accrue over the course of evolution, and for understanding the molecular etiology of disease. Methods which measure, predict or optimize these changes have applications in designing de novo interactions (Fleishman et al., 2011), designing combinatorial protein libraries (e.g. Guntas et al., 2010), uncovering the effects of pathological mutations (e.g. Tidow et al., 2006), locating druggable binding sites (e.g. Stevers et al., 2017) and binding hotspots for drug design (Guo et al., 2014), altering binding kinetics (Cohen-Khait and Schreiber, 2016; Rosenfeld et al., 2017), protein–protein docking (e.g. EPA et al., 2013), and characterizing transition states (e.g. Wu et al., 2002), binding pathways (Plattner et al., 2017), and sequence-affinity landscapes (Aizner et al., 2014).

SKEMPI is a manually curated database of mutations in structurally characterized protein–protein interactions and the effect of those mutations on binding affinity and other parameters (Moal and...
The first release has been used as a basis for many further studies, including the development of energy functions (Moal and Fernandez-Recio, 2013; Moal et al., 2015b) which were subsequently implemented in the CCharPPI web server for characterizing protein–protein interaction motifs (Moal et al., 2015a), as well as being used for ranking docked poses (Barradas-Bautista et al., 2017; Moal et al., 2013; Moal et al., 2017; Pleifffenberger et al., 2017). SKEMPI has also been used to study human disease (Das et al., 2014; Peng and Alexov, 2016; Petukh et al., 2015a), assessing the role of dynamics on binding (Sumbul et al., 2015), exploring the conservation of binding regions (Hu et al., 2016), evaluating experimental affinity measurement methods (Geng et al., 2016), as well serving as a data source for models which predict dissociation rate changes upon mutation (Agius et al., 2013), pathological mutations (Gossage et al., 2014), hotspot residues (e.g. Hwang et al., 2014; Liu et al., 2015; Melo et al., 2016; Simoes et al., 2017) and changes in binding energy (e.g. Barlow et al., 2018; Berliner et al., 2014; Dehouck et al., 2013; Dourado and Flores, 2014; Lai et al., 2017; Li et al., 2016; Moretti et al., 2013; Pallara et al., 2013; Pantazes et al., 2015; Petukh et al., 2015b; Pires et al., 2014; Xiong et al., 2017; Yan et al., 2017; Zhao et al., 2014).

Here we present a major update to the benchmark in terms of the number of mutations in the database and the number of different systems included (Table 1). We now also include details of the experimental method for all entries, based on the categories of Geng et al., 2016, as well as mutations which abolished detectable binding or for which only an upper or lower affinity limit could be ascertained for the wild-type or mutant.

2 Materials and methods

2.1 Data sources

Just over two fifth of the data come from the previous version of SKEMPI (Moal and Fernandez-Recio, 2012), comprised mostly of data found in literature sources which came to the authors’ attention, in some cases during the data collection for the structural affinity benchmark (Kastritis et al., 2011) and following references therein. Some entries in SKEMPI 1.1 were found by checking the references in the ASEDb (Thorn and Bogan, 2001) and PINT (Kumar and Gromiha, 2006) databases, although not all the data passed the checks required for inclusion (see Section 2.2). Similarly, most of the new data in SKEMPI 2.0 was found by searching the literature, partly in tandem with the literature search for the more recent structural affinity benchmark (Vreven et al., 2015). During data collection three other relevant databases were published: ABind (Sirin et al., 2016), PROXiMATE (Jemimah et al., 2017) and dBMPIKT (Liu et al., 2017). Their references were checked if they were not already included. Data from these sources comprise 4%, 3% and 6% of SKEMPI 2.0 respectively. As with ASEDb and PINT, none of the data were directly copied into SKEMPI. Moreover, the cited papers were read and data entered using the same checks and procedures as other entries.

2.2 Data collection

Each entry was found in the literature and manually vetted. To ensure quality, a number of stringent checks were applied. Firstly, we ensured that the structure and the paper reporting the affinities refer to the same protein in the same species, and that structural and affinity data matched in terms of cofactors, ancillary chains and post-translational modifications. For instance, we distinguish between RasGTP and RasGppNHp, as the nucleotide modulates the affinity of Ras with its effectors. Where the full-length protein was not used, we checked to ensure that the fragment in the crystal structure matched that for which affinities are reported.

Once the checks are passed, the data is collected, including the PDB file, the chains of the interacting subunits, the mutation, the wild-type and mutant affinities (kₜₜ, M), the reference, the names of the proteins, the temperature at which the experiment is performed (T, K), the experimental method used (an extension of the category of Geng et al., 2016), notes on the entry and, when available, the association rate (kₜₜ, M⁻¹s⁻¹), dissociation rate (kₜₜ, s⁻¹), enthalpy (ΔH, kcal.mol⁻¹) and entropy (ΔS, cal.mol⁻¹K⁻¹). For cases where multiple PDB entries are available, the higher resolution structure is chosen. Where affinities or kinetic or thermodynamic parameters are reported in different units, these are converted to the units specified above. In some cases, when not reported directly, kₜₜ, kₜₜ, kₜₜ, ΔH and ΔS were calculated using the relationships

\[ ΔG = ΔH - TΔS = RTln(Kₜₜ) \quad \text{and} \quad Kₜₜ = 1/Kₜₜ = kₜₜ/kₜₜ. \]

To ensure consistency, the residue numbering in SKEMPI is the same as that reported in the PDB file. Thus, the numbering is often shifted or altered compared to that in the cited paper such that, for instance, if a crystal structure of an antibody is reported in the Kabat numbering scheme but the mutation data is not, then the mutation data is converted before entry into SKEMPI. For all entries it is the case that the affinity reported in the “wild-type” column corresponds to that of the PDB file, and the affinity in the “mutant” column is that after applying the specified mutation to the protein in the PDB file. Thus, where there are cases in which the PDB reports a mutant form and the entry corresponds to the reverse mutation back to the wild-type, the affinity of the former appears in the “wild-type” column and the latter in the “mutant” column. Such cases are noted in the database.

In addition to checking new entries, we reappraised the papers cited in SKEMPI 1.1 to collect data that were not collected previously, specifically to find mutants which abolish binding and to classify the experimental method used when not already included in the subset of SKEMPI covered in Geng et al., 2016. It is worth noting that often an author’s decision to report an interaction of affinity below the detection threshold as either non-binding, or as less affine than the weakest affinity presented in the paper, is arbitrary. Thus, those wishing to use the non-binding data as an inequality on the affinity may do so. We also corrected entries for five wild-type and four mutant affinities identified by Geng et al., 2016.

2.3 Post-processing and annotation

In addition to the above data, SKEMPI 2.0 also provides data on the location of the mutated residues, the homology between interactions in the dataset, and processed PDB files, which can be easily parsed.

**Residue location:** Each mutated residue is classified according to the scheme proposed by Levy (2010); residues at the interface are classified as support (mostly buried when unbound and entirely

### Table 1. Comparison with previous version

<table>
<thead>
<tr>
<th>SKEMPI 1.1</th>
<th>SKEMPI 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entries</td>
<td>3047</td>
</tr>
<tr>
<td>Unique entries</td>
<td>2792</td>
</tr>
<tr>
<td>kₜₜ and kₜₜ</td>
<td>713</td>
</tr>
<tr>
<td>ΔH and ΔS</td>
<td>127</td>
</tr>
<tr>
<td>Inequalities/no binding</td>
<td>0</td>
</tr>
<tr>
<td>Number of interactions</td>
<td>87</td>
</tr>
<tr>
<td>Number of PDB entries</td>
<td>158</td>
</tr>
<tr>
<td>Number of papers cited</td>
<td>66</td>
</tr>
</tbody>
</table>
buried upon binding), core (mostly solvent exposed when unbound but buried upon binding), and rim (partly buried upon binding), while residues away from the binding site are classified as interior or surface. Solvent exposed surface area was calculated using CCP4 (Winn et al., 2011).

Processed PDB files: The PDB files for the interactions, as downloaded from the Protein Data Bank (Berman et al., 2000), often contain multiple copies of the interacting proteins in the unit cell or other chains irrelevant to the interaction. In one instance, the binding of dimeric myostatin to follistatin-like 3, the myostatin dimer must be created by tessellating the unit cell. Further, some PDB files contain features that are not readily parsed by some software, such as residue insertion codes or negative residue numbers. To help users we provide “cleaned” PDB files which contain only the chains of interest, renumbered from one, as well as waters and other molecules with a non-hydrogen atom within 5 Å of a non-hydrogen atom of any of the chains of interest. Consequently, each mutation is reported with both PDB numbering and renumbered.

Defining homologous interactions: Each entry also specifies which other entries are mutations to homologous interactions. Two interactions are deemed homologous if they have a shared binding partner or homologous binding partner and at least 70% of the corresponding interface residues are common to both interactions. We determine the homology between proteins using the GAP4 program (Huang and Brutlag, 2007), and define homologous proteins as those with a similarity score greater than 50 and at least 30% sequence identity. Interface residues are defined as those with a non-hydrogen atom within 10 Å of a non-hydrogen atom on the binding partner. Interactions falling within manually assigned clusters of homologous interactions are designated as pMHC/TCR, antibody/antigen or protease/inhibitor. While the names of these clusters have been chosen to reflect the predominant function of their constituent interactions, they reflect the homologies within the dataset and are not functional assignments. Thus, for instance, some nanobodies are classified as antibodies as they bind to the same site as cetuximab, 14.3.d is classified as TCR, even though it is only the β chain, and its binding partner, enterotoxin C3, is classified as a pMHC.

3 Results and discussion
3.1 Diversity, bias and interrelationships within SKEMPI
In total 7085 entries were collected, summarized in Table 1 and Figure 1A. These data were derived from the literature and consequently, while encompassing a broad range of residues, proteins, interactions and systems, are biased toward the interests and capabilities of the research community. These biases are evident in the composition of the database according to parameters shown in Figure 1. The ΔG values span a large range, but mostly fall within ~3 to 7 kcal mol⁻¹ (Fig. 1B), for both biophysical and technical reasons (see Section 3.3). Almost three quarters of the data correspond to single point mutations, and more than half of those are mutations to alanine (Fig. 1C). Charge swap mutations and mutations between aromatic residues are also over-represented. Most single point mutations are located at the binding site, and most of those are at the core of the interface. Similarly, most double mutations are both in the binding site, and most of those are both in the core. By far the most popular methods for measuring binding affinity were surface plasmon resonance and spectroscopic methods such as fluorescence. Large biases toward specific interactions and classes of interaction are also present, such as early studies into protease inhibition and immunological interactions such as antibody-antigen complexes, the recognition of peptides presented on cell surfaces by T-cell receptors, cytokine signaling and the complement system. Indeed, almost half of the data corresponds to protease-inhibitor, antibody-antigen and pMHC/TCR interactions alone. While many interactions within these classes share common binding sites or homologous binding sites (Fig. 2), there are also connections between these groups, for instance via inhibitory antibodies which bind to a protease active site, or due to common binding regions of proteins in the immunoglobulin superfamily, such as antibodies, TCRs, MHCs and β2-microglobulin. In addition, present in the data are smaller clusters of shared and homologous interactions, such as the Ras-effector cluster. These relationships are noted in the database and may be useful for avoiding overfitting when developing models or for validation and estimating generalization error, as described previously (Moal and Fernandez-Recio, 2012).

The entries also vary in the degree of structural order. While most correspond to interactions between folded domains, the database contains entries in which structuring occurs upon binding, such as protein–peptide interactions and, in the extreme case, the ACTR/NCBD interaction in which both binding partners become ordered upon binding (Jemth et al., 2014). Indeed, the requirements of having a structure in order to be included in the dataset, ipso facto biases the data, and means that there is no representation of “fuzzy” complexes in which a diffuse structural ensemble in the bound state prevents the formation of a resolvable crystal.

Another source of variation is the origins of the interacting proteins. While entries range from viral and bacterial to the higher eukaryotes, biases are evident in the over-representation of model organisms including humans. With the exception of the pMHC-TCR, antibody-antigen and protease-inhibitor classes, most of the interactions are endogenous. Nevertheless, the set also includes exogenous interactions ranging from those between proteins from different individuals within the same species, namely the sex fusion proteins Juno and Izumo1 from human sperm and egg respectively (Aydin et al., 2016), to host-pathogen interactions such as adenovirus and coronavirus interactions with human receptors during viral entry (Howitt et al., 2003; Seiradake et al., 2006), to the inhibition of acetylcholinesterase by the snake venom neurotoxin fasciculin (Aizner et al., 2014). For the pMHC-TCR interactions, there are a variety of presented antigens, including exogenous viral peptides and gluten, as well as endogenous autoimmunie and cancer peptides. The antibody interactions include pathogen antibodies, as well as antibodies raised and optimized to target extracellular therapeutic targets. The protease interactions are mostly exogenous, arising from their inherent cross-reactivity due to the convergent evolution of their canonical inhibitory loop.

3.2 Notable studies comprising SKEMPI 2.0
The investigations from which SKEMPI data is derived are diverse, spanning many biological processes and reported in 295 publications including systematic scans, alanine and homolog scanning, design studies including computational design and designs derived from phage display, double mutant cycle studies, antibody engineering, biologic drug design and the evaluation of pathological mutations. The largest contribution comes from the group of the late Michael Laskowski Jr., a systematic study of all possible mutations at selected sites in the turkey ovomucoid third domain and its inhibitory interactions with four proteases (Lu et al., 2001), as well as studies of interactions of the same domain in other bird species and the design of ultra-high affinity broad-spectrum inhibitors. Substantial data also come from investigations into the inhibitory
interactions of class \( \beta \)-lactamases from the groups of Gideon Schreiber (e.g. Reichmann et al., 2007) and Timothy Palzkill (e.g. Brown et al., 2013), as well as cytokine receptor interactions, in particular studies of type I interferons also from the Schreiber group (e.g. Roisman et al., 2001) and that of K. Christopher Garcia (Thomas et al., 2011), but also the study of the GM-CSF/GMRC interaction from the group of Michael W. Parker (Broughton et al., 2016). Other prominent sources of data are studies into hormone receptor interactions, in particular the human growth hormone receptor from the group of Jim Wells (e.g. Cunningham and Wells, 1991) and the prolactin receptor from the group of Michael E. Hodsdon (Kulkarni et al., 2010), as well as studies into antigen recognition including the combined computational and experimental design study to enhance affinity of the AQC2 antibody to integrin \( \alpha \)-1 from the group of Herman Van Vlijmen (Clark et al., 2006), the dissection of the interactions of broadly neutralizing antibodies targeting HIV gp120 (Clark et al., 2017) from the group of Richard A. Friesner, and various investigations from the group of Roy A. Mariuzza (e.g. Dall’Acqua et al., 1998). Also notable is the alanine scanning of the urokinase-type plasminogen activator and its receptor from the group of Michael Ploug (Gardsvoll et al., 2006), studies of Ras effector interactions from the group of Christian Herrmann (e.g. Kiel et al., 2004), investigations of pMHC/TCR interactions from the group of Brian Baker (e.g. Piepenbrink et al., 2013), and investigations into the cognate and non-cognate recognition of \textit{Escherichia coli} colicin DNase bacteriotoxins by their immunity proteins from the group of Colin Kleanthous (e.g. Li et al., 1997).

### 3.3 Range and error

**Range:** The changes in binding free energy upon mutation range from \(-12.4\) to \(12.4\) kcal mol\(^{-1}\), as in SKEMPI 1.1, with \(\Delta \log_{10} k_{on}\) ranging from \(-3.6\) to \(2.4\), \(\Delta \log_{10} k_{off}\) ranging from \(-6.0\) to \(6.8\), \(\Delta \Delta H\)
Fig. 2. Overview of the interactions in SKEMPI. Nodes indicate proteins, scaled by the number of mutations of that protein and coloured according to category. Edges show direct interactions, as well as relationships between proteins that share a common or homologous binding site.
ranging from -18.3 to 26.5 kcal mol\(^{-1}\), and \(\Delta\Delta S\) ranging from -61 to 80 cal mol\(^{-1}\) K\(^{-1}\). Around 60 mutants are very destabilizing, reducing binding energy by 8 kcal mol\(^{-1}\) or more. These are all in enzyme/inhibitor complexes such as the inhibition of acetylcholinesterase by the snake venom fasciculin, or the inhibition of enzymes which would be detrimental should they unbind and become active in the wrong location, such as nucleases (barnase/barstar, colicin E9 DNase/IIm9, RNase A/angiogenin) and proteases (such as trypsin/BPTI). These interactions tend to be around picomolar affinity and are at the upper limit of what can be detected, due to the time required to reach equilibrium and the low concentrations required by the mass action law to probe informative regions of the binding curve. These very destabilizing mutations reduce affinity into the micromolar range, near the lower limit of what can be quantified using standard methods. As a consequence of both mutant and wild-type affinities being near detection thresholds, errors in these entries are typically large. Further, while some mutations may cause changes in affinity larger than seven orders of magnitude, the absence of affinities for such mutations in the benchmark can be explained by the fact that such mutations would involve affinities beyond the upper or lower limit. Indeed, there are new entries in which single or double substitutions reduce binding from tens of picomolar to having no detectable binding. For many of the highly destabilizing mutations a crystal structure for the mutant has also been solved, and the 30 most stabilizing mutations in the database (\(\Delta\Delta G_{1 \rightarrow 3} \leq 5\) kcal mol\(^{-1}\)) consist of the reverse mutation applied to these structures. These are mostly single or double mutants, but include mutations to up to 27 residues of the non-cognate Colicin E2/Im9 complex, which move it toward the cognate E9/IIm9 in sequence space (Li et al., 1997).

Errors: Standard errors in \(K_D\) are typically reported in the order of 50%, around 0.25 kcal mol\(^{-1}\). These estimates are derived by repeat measurements using the same equipment, environment and protocol, and thus do not include errors arising from systematic bias. Such biases can, however, be estimated from pairs of entries in which the same mutation is evaluated by different groups or using different techniques. For 84% of 1741 such pairs, both entries give a \(\Delta\Delta G\) value within 1 kcal mol\(^{-1}\) of each other. For 704 pairs for which \(k_{on}\) is available for both, 80% have \(\Delta \log k_{on}\) within 0.5 of each other. For 702 pairs for which \(k_{eq}\) is available, 83% have \(\Delta \log k_{eq}\) within 0.5. For the 62 pair with both \(\Delta\Delta H\) and \(\Delta\Delta S\) values, 61% have \(\Delta\Delta H\) within 3.0 kcal mol\(^{-1}\) of each other and 58% have \(\Delta\Delta S\) within 10 cal mol\(^{-1}\) K\(^{-1}\) of each other.

### 3.4 Mutant cycles

Within SKEMPI, some entries can be combined to construct mutant cycles, which quantify the interactions between residues, the dependence of these interactions on other residues, and other higher order effects. The most common instances are double mutant cycles, where affinities are available for the wild type, A, B and AB mutations, of which there are 610 examples. Of these, 53 involve at least one mutant for which binding was not observed, or only an inequality is available, and 235 involve mutations reported in the same reference, and thus the affinities are likely to have been measured using the same technique and conditions. A further 218 double mutant cycles can be constructed in the background of a third mutation (i.e. C, AC, BC and ABC mutations are available), of which 209 are not composed of non-binding mutations or mutations with inequalities, and 131 involve affinities coming from the same reference. Of the 766 double mutant cycles containing neither inequalities nor non-binding mutants, a number of parameters can be calculated, including \(\Delta\Delta G_{\text{AB} \rightarrow \text{A}}, \Delta\Delta G_{\text{AB} \rightarrow \text{B}}\) and \(\Delta\Delta G_{\text{AB}}\), the binding free energy change of both single and the double mutation respectively, as well as \(\Delta\Delta G_{\text{AB} \rightarrow \text{A}}\) and \(\Delta\Delta G_{\text{AB} \rightarrow \text{B}}\), the energy of a single mutation within the context of the other mutation, and \(\Delta\Delta G_{\text{int}} = \Delta\Delta G_{\text{AB} \rightarrow \text{A}} - \Delta\Delta G_{\text{AB} \rightarrow \text{B}}\) the interaction energy of the two mutations (Horovitz and Fersht, 1990). From these, it can be deduced that 345 are additive (\(\Delta\Delta G_{\text{int}} < 0.5\) kcal mol\(^{-1}\)). Of the non-additive cycles, 293 exhibit tighter binding in the double mutant than the sum of the single mutants (positive epistasis), of which six result in even tighter binding than individual effects of two single mutations that strengthen the interaction (synergistic positive), while 273 correspond to double mutants which reduce binding by less than the sum of two single mutants which reduce binding (antagonistic positive). Similarly, 128 cycles have double mutants exhibiting weaker binding than the sum of the two single mutants (negative epistasis), of which 58 contain two destabilizing single mutations (synergistic negative) and 26 contain two stabilizing single mutations (antagonistic negative). The range of \(\Delta\Delta G_{\text{int}}\) values rarely fall outside of the -5 to 3 kcal mol\(^{-1}\) range. Of the 421 non-additive cycles, 151 show noticeable sign epistasis, in which the sign of the effect of either the A or B mutation flips depending on the presence or absence of the background mutation (i.e. for the A mutation, \(\Delta\Delta G_{\text{A} \rightarrow \text{A}} > 0.2\) kcal mol\(^{-1}\) and \(\Delta\Delta G_{\text{A} \rightarrow \text{A}} > 0.2\) kcal mol\(^{-1}\) and \(\Delta\Delta G_{\text{A} \rightarrow \text{A}} > 0.4\) kcal mol\(^{-1}\)). Of these, 38 correspond to mutations, which destabilize the complex in the presence of the background mutation, but stabilize in its absence (destabilizing sign epistasis), while 113 correspond to mutations, which stabilize the complex in the mutant background but otherwise destabilize the complex (stabilizing sign epistasis). Only eight cycles exhibit the more extreme reciprocal sign epistasis, which in six cases are where both single mutations are stabilizing (< -0.2 kcal mol\(^{-1}\)), but the double mutant is destabilizing (>0.2 kcal mol\(^{-1}\)) and the remaining two correspond to two destabilizing mutations (>0.2 kcal mol\(^{-1}\)) for which the double mutation is stabilizing (< -0.2 kcal mol\(^{-1}\)). The types of substitutions that can give rise to extreme effects such as stabilizing reciprocal sign epistasis can be illustrated with the Mcl-IIBGc interaction in E. coli (Nam et al., 2008), in which the removal of the F136 side-chain of MIC creates a large cavity at the binding interface, the addition of a phenylalanine at the A451 position of IIBGc creates a large clash, however the double mutation creates an interaction that is even more stable than the wild-type by creating an anchor residue across the binding interface in which the cavity in MIC is filled by the new side-chain of IIB.

Higher order interaction terms can be garnered from higher cycles, such as triple mutant cubes, constructed from the energies of the wild-type, three single mutants, three corresponding double mutants and the triple mutant (Horovitz and Fersht, 1990). In SKEMPI, 45 triple mutant cubes can be made, with 10 coming from the same reference. For these, third order interaction energies fall within the -1 to 1 kcal mol\(^{-1}\) range. For fourth order interactions, constructed from energies of the wild-type, four single mutants, six double mutants, four triple mutants and the quadruple mutant, 14 examples exist within SKEMPI. However, care should be taken in ascribing meaning to fourth order residue coupling energies due to the accumulations of errors, which in these cases are exacerbated by the affinities having been reported in different publications. No fifth or higher order interactions are present.

### 3.5 The SKEMPI website

The database is accessible online at https://life.bsc.es/pid/skempi2/, where the raw CSV (comma-separated values) file containing all the
data can be downloaded. The data can also be browsed online, ordered and searched by any field, such as the experimental method or the location of the mutation, or searching for a specific protein by its name or PDB code, and structures may be visualized. Other pages on the web site offer a summary of the data, an FAQ and help page, and a page for user contributions, which will be evaluated to appear in future releases.

Funding

This work has been supported by the European Molecular Biology Laboratory [L.H.M.]; Biotechnology and Biological Sciences Research Council [Future Leader Fellowship BB/N011600/1 to L.H.M.]; Spanish Ministry of Economy and Competitiveness (MINECO) [BIO2016-79930-R to J.F.R.]; Interreg POCTEFA [EFA086/15 to J.F.R.]; European Commission [H2020 grant 676566 (MuG)].

Conflict of Interest: none declared.

References


Broughton,S.E. et al. (2016) Conformational changes in the GM-CSF receptor suggest a molecular mechanism for affinity conversion and receptor signaling. Structure, 24, 1271–1281.


