

# **Supporting Information – A Hybrid Knowledge-Based and Empirical Scoring Function for Protein-Ligand Interaction: SMoG2016**

*<sup>1</sup>Théau Debroise, <sup>1</sup>Eugene I. Shakhnovich and <sup>1</sup>Nicolas Chéron\**

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA 02138, USA; <sup>2</sup>Ecole Normale Supérieure, PSL Research University, UPMC Univ. Paris 06, CNRS, Département de Chimie, UMR 8640 PASTEUR, 24 rue Lhomond, 75005 Paris, France.

\*Correspondence: [nicolas.cheron@ens.fr](mailto:nicolas.cheron@ens.fr)

## Table of contents:

<b>1. Description of the SMoG2001 scoring function.....</b>	<b>2</b>
<b>2. Evaluation of different shells for the knowledge-based potential.....</b>	<b>4</b>
<b>3. Influence of the slope of the contact function .....</b>	<b>8</b>
<b>4. Influence of the <math>\alpha</math> parameter on the correlation for hybrid scores with <math>\Delta</math>SASA .....</b>	<b>9</b>
<b>5. List of complexes contained in the testing set (195 complexes).....</b>	<b>10</b>
<b>6. List of complexes contained in the training set (1,038 complexes) .....</b>	<b>11</b>

## 1. Description of the SMoG2001 scoring function

Knowledge-based scoring functions take advantage of structural information in a database of protein-ligand crystallographic structures. In the pairwise approximation, the total score  $S$  can be expressed as a sum over all contacts between the ligand and the protein atoms:

$$S = \sum_p \sum_l F(\sigma_p, \sigma_l) \quad (1)$$

where the index  $p$  runs over the protein atoms of type  $\sigma_p$ , the index  $l$  runs over the ligand atoms of type  $\sigma_l$  and  $F(\sigma_p, \sigma_l)$  is the free energy associated to a specific contact. The distance-dependent potential  $F(\sigma_p, \sigma_l)$  is usually derived from the inverse Boltzmann-like analysis:

$$\Delta F(\sigma_p, \sigma_l) = -RT \ln \frac{f(\sigma_p, \sigma_l)}{f_{ref}} \quad (2)$$

where  $f(\sigma_p, \sigma_l)$  is the frequency of the contact between the atoms of type  $\sigma_p$  and  $\sigma_l$  in the database. Since we are interested in differences of free energy, it is important to take into account a reference state where  $f_{ref}$  is the frequency of this contact. It has been shown that a correct choice of  $f_{ref}$  is crucial for the efficiency of the function<sup>1</sup>. Contact between atoms is usually defined with geometric criteria, for example if the two atoms are at less than 5 Å from each other. The inverse Boltzmann hypothesis assumes that the occurrence of a type of contact in the database is linked to its free energy contribution. For example, if CH<sub>2</sub> groups from the ligands are often found in contact with CH from valines in a database, then this should mean that a favorable free energy exists between the two atom types (be it electrostatic, dispersion, dipole, vibrational, etc...).

One of the first knowledge-based protein-ligand potential (SMoG96) was developed in 1996 by De Witte and Shakhnovich<sup>2</sup>. It was then improved in 2001 (SMoG2001) with the form<sup>1</sup>:

$$S = \sum_r \sum_p \sum_l \Delta F(\sigma_p, \sigma_l, r) \Delta(p, l, r) \quad (3)$$

where  $\Delta(p, l, r)$  is the contact function. In SMoG2001, two shells of contact were used: if the distance between two atoms was between 0 and 3.5 Å, a first type of contact was defined with a given contribution to the energy; if the distance between the two atoms was between 3.5 and 4.5 Å, a second type of contact was defined with another energy contribution; if the distance between the two atoms was greater than 4.5 Å, the two atoms were considered to be not in contact with no energy contribution. To derive the statistics, the following function was used:

$$\Delta F(\sigma_p, \sigma_l, r) = -\ln \left[ \frac{N(\sigma_p, \sigma_l, r)}{C(r) \Omega(\sigma_p, \sigma_l, r) \sum_{\sigma_p} \sum_{\sigma_l} N(\sigma_p, \sigma_l, r)} \right] \quad (4)$$

where  $N(\sigma_p, \sigma_l, r) / \sum_{\sigma_p} \sum_{\sigma_l} N(\sigma_p, \sigma_l, r)$  is the frequency of one type of contact in a given interval of distance (the term  $f(\sigma_p, \sigma_l)$  in equation (2)).  $C(r) * \Omega(\sigma_p, \sigma_l, r)$  is the probability of this contact in the reference state ( $f_{ref}$  in equation (2)) where atomic interactions are assumed to be zero.  $C(r)$  is a normalization constant and  $\Omega(\sigma_p, \sigma_l, r)$  is a term accounting for the composition of the database in a defined interval of distance:

$$\Omega(\sigma_p, \sigma_l, r) = N(\sigma_p, r)^\alpha N(\sigma_l, r)^\beta \quad (5)$$

where  $N(\sigma_p, r)$  and  $N(\sigma_l, r)$  are the number of atoms of type  $\sigma_p$  or  $\sigma_l$  in an interval of distances. It was shown that the best correlation is obtained when  $\alpha=\beta=0.9^1$ . The aim of this term is to remove the correlation between the frequency of observation of one contact and the abundance of the two atom types in the database.  $C(r)$  is derived from eq. 6 by taking the average over all the atom types:

$$\ln[C(r)] = \langle \ln \frac{N(\sigma_p, \sigma_l, r)}{\Omega(\sigma_p, \sigma_l, r) \sum_{\sigma_p} \sum_{\sigma_l} N(\sigma_p, \sigma_l, r)} \rangle + F_0(r) \quad (6)$$

where  $F_0(r)$  is a free parameter equal to the average free energy in a given interval of distance,  $F_0(r) = \langle F(\sigma_p, \sigma_l, r) \rangle$ . The detailed theory behind this function is well described in the SMoG96 and SMoG2001 papers<sup>1,2</sup> and we point the interested reader to these articles.

## References

- (1) Ishchenko, A. V.; Shakhnovich, E. I. SMall Molecule Growth 2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein–Ligand Interactions. *J. Med. Chem.* **2002**, *45*, 2770–2780.
- (2) DeWitte, R. S.; Shakhnovich, E. I. SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733–11744.

## 2. Evaluation of different shells for the knowledge-based potential

Table S1. Average correlations on 20 testing sets for the knowledge-based potentials trained on the corresponding 20 training sets with two shells.

Shells configuration	$\langle R \rangle$	Shells configuration	$\langle R \rangle$
<b>3.0-4.5</b>	.305	<b>4.5-5.5</b>	.348
<b>3.0-5.0</b>	.353	<b>4.5-6.0</b>	.365
<b>3.0-5.5</b>	.376	<b>4.5-6.5</b>	.374
<b>3.0-6.0</b>	.389	<b>4.5-7.0</b>	.383
<b>3.0-7.0</b>	.394	<b>4.5-7.5</b>	.391
<b>3.0-8.0</b>	.397	<b>4.5-8.0</b>	.393
<b>3.0-9.0</b>	.397	<b>4.5-9.0</b>	.397
<b>3.0-10.0</b>	.393	<b>4.5-10.0</b>	.396
<b>3.5-4.5</b>	.262	<b>5.0-7.0</b>	.388
<b>3.5-5.0</b>	.311	<b>5.0-8.0</b>	.396
<b>3.5-5.5</b>	.343	<b>5.0-9.0</b>	.400
<b>3.5-6.0</b>	.365	<b>5.0-10.0</b>	.398
<b>3.5-7.0</b>	.384	<b>5.5-6.5</b>	.385
<b>3.5-8.0</b>	.393	<b>5.5-7.0</b>	.390
<b>3.5-9.0</b>	.396	<b>5.5-7.5</b>	.395
<b>3.5-10.0</b>	.393	<b>5.5-8.0</b>	.397
<b>4.0-6.0</b>	.362	<b>5.5-9.0</b>	.400
<b>4.0-6.5</b>	.374	<b>5.5-10.0</b>	.398
<b>4.0-7.0</b>	.384	<b>6.0-8.0</b>	.397
<b>4.0-8.0</b>	.395	<b>6.0-9.0</b>	.400
<b>4.0-9.0</b>	.398	<b>6.0-10.0</b>	.398
<b>4.0-10.0</b>	.396		

Table S2. Average correlations on 20 testing sets for the knowledge-based potentials trained on the corresponding 20 training sets with three shells.

<b>Shells configuration</b>	<b>&lt;R&gt;</b>	<b>Shells configuration</b>	<b>&lt;R&gt;</b>	<b>Shells configuration</b>	<b>&lt;R&gt;</b>
<b>3.0-4.5-6.0</b>	.377	<b>3.5-4.5-9.0</b>	.396	<b>4.0-6.5-8.5</b>	.398
<b>3.0-4.5-6.5</b>	.384	<b>3.5-4.5-10.0</b>	.394	<b>4.0-6.5-9.0</b>	.399
<b>3.0-4.5-7.0</b>	.390	<b>3.5-5.0-6.0</b>	.360	<b>4.0-6.5-9.5</b>	.399
<b>3.0-4.5-7.5</b>	.396	<b>3.5-5.0-6.5</b>	.371	<b>4.0-6.5-10.0</b>	.397
<b>3.0-4.5-8.0</b>	.397	<b>3.5-5.0-7.0</b>	.381	<b>4.5-5.5-6.5</b>	.374
<b>3.0-4.5-9.0</b>	.398	<b>3.5-5.0-7.5</b>	.390	<b>4.5-5.5-7.0</b>	.382
<b>3.0-4.5-10.0</b>	.395	<b>3.5-5.0-8.0</b>	.393	<b>4.5-5.5-7.5</b>	.390
<b>3.0-5.0-7.0</b>	.393	<b>3.5-5.0-9.0</b>	.397	<b>4.5-5.5-8.0</b>	.393
<b>3.0-5.0-7.5</b>	.399	<b>3.5-5.0-10.0</b>	.395	<b>4.5-5.5-9.0</b>	.398
<b>3.0-5.0-8.0</b>	.399	<b>3.5-5.5-7.0</b>	.382	<b>4.5-5.5-10.0</b>	.396
<b>3.0-5.0-8.5</b>	.401	<b>3.5-5.5-7.5</b>	.390	<b>4.5-6.0-7.0</b>	.383
<b>3.0-5.0-9.0</b>	.400	<b>3.5-5.5-8.0</b>	.393	<b>4.5-6.0-7.5</b>	.390
<b>3.0-5.0-10.0</b>	.397	<b>3.5-5.5-8.5</b>	.396	<b>4.5-6.0-8.0</b>	.394
<b>3.0-5.5-8.0</b>	.400	<b>3.5-5.5-9.0</b>	.398	<b>4.5-6.0-8.5</b>	.397
<b>3.0-5.5-8.5</b>	.401	<b>3.5-5.5-10.0</b>	.396	<b>4.5-6.0-9.0</b>	.398
<b>3.0-5.5-9.0</b>	.401	<b>3.5-6.0-8.0</b>	.394	<b>4.5-6.0-10.0</b>	.397
<b>3.0-5.5-10.0</b>	.397	<b>3.5-6.0-8.5</b>	.397	<b>4.5-6.5-8.0</b>	.393
<b>3.0-6.0-9.0</b>	.401	<b>3.5-6.0-9.0</b>	.398	<b>4.5-6.5-8.5</b>	.396
<b>3.0-6.0-10.0</b>	.398	<b>3.5-6.0-10.0</b>	.396	<b>4.5-6.5-9.0</b>	.398
<b>3.0-6.5-9.0</b>	.400	<b>3.5-6.5-9.0</b>	.397	<b>4.5-6.5-9.5</b>	.398
<b>3.0-6.5-9.5</b>	.399	<b>3.5-6.5-9.5</b>	.397	<b>4.5-6.5-10.0</b>	.396
<b>3.0-6.5-10.0</b>	.397	<b>3.5-6.5-10.0</b>	.395	<b>4.5-7.0-9.0</b>	.398
<b>3.0-7.0-9.0</b>	.400	<b>3.5-7.0-9.0</b>	.397	<b>4.5-7.0-10.0</b>	.397
<b>3.0-7.0-10.0</b>	.397	<b>3.5-7.0-10.0</b>	.395	<b>5.0-7.0-9.0</b>	.400
<b>3.5-4.5-5.5</b>	.333	<b>4.0-6.0-8.0</b>	.395	<b>5.0-7.0-9.5</b>	.400
<b>3.5-4.5-6.0</b>	.356	<b>4.0-6.0-8.5</b>	.398	<b>5.0-7.0-10.0</b>	.398
<b>3.5-4.5-6.5</b>	.368	<b>4.0-6.0-9.0</b>	.399	<b>5.5-7.5-9.5</b>	.400
<b>3.5-4.5-7.0</b>	.379	<b>4.0-6.0-10.0</b>	.398	<b>5.5-7.5-10.0</b>	.398
<b>3.5-4.5-8.0</b>	.391	<b>4.0-6.5-8.0</b>	.394		

Table S3. Average correlations on 20 testing sets for the knowledge-based potentials trained on the corresponding 20 training sets with four shells.

<b>Shells configuration</b>	<b><math>\langle R \rangle</math></b>	<b>Shells configuration</b>	<b><math>\langle R \rangle</math></b>	<b>Shells configuration</b>	<b><math>\langle R \rangle</math></b>
<b>3.0-4.0-5.0-6.0</b>	.373	<b>3.0-4.5-6.5-8.5</b>	.399	<b>3.5-5.0-6.5-10.0</b>	.396
<b>3.0-4.0-5.0-6.5</b>	.381	<b>3.0-4.5-6.5-9.0</b>	.399	<b>3.5-5.0-7.0-9.0</b>	.398
<b>3.0-4.0-5.0-7.0</b>	.389	<b>3.0-4.5-6.5-10.0</b>	.397	<b>3.5-5.0-7.0-9.5</b>	.398
<b>3.0-4.0-5.0-7.5</b>	.396	<b>3.0-4.5-7.0-9.0</b>	.399	<b>3.5-5.0-7.0-10.0</b>	.396
<b>3.0-4.0-5.0-8.0</b>	.397	<b>3.0-4.5-7.0-9.5</b>	.399	<b>3.5-5.5-7.5-9.5</b>	.398
<b>3.0-4.0-5.0-9.0</b>	.399	<b>3.0-4.5-7.0-10.0</b>	.397	<b>3.5-5.5-7.5-10.0</b>	.396
<b>3.0-4.0-5.0-10.0</b>	.397	<b>3.0-5.0-7.0-9.0</b>	.401	<b>4.0-5.0-6.0-7.0</b>	.381
<b>3.0-4.0-5.5-6.5</b>	.382	<b>3.0-5.0-7.0-9.5</b>	.400	<b>4.0-5.0-6.0-7.5</b>	.390
<b>3.0-4.0-5.5-7.0</b>	.389	<b>3.0-5.0-7.0-10.0</b>	.398	<b>4.0-5.0-6.0-8.0</b>	.394
<b>3.0-4.0-5.5-7.5</b>	.396	<b>3.0-5.0-7.5-9.0</b>	.401	<b>4.0-5.0-6.0-9.0</b>	.399
<b>3.0-4.0-5.5-8.0</b>	.397	<b>3.0-5.0-7.5-9.5</b>	.400	<b>4.0-5.0-6.0-10.0</b>	.397
<b>3.0-4.0-5.5-9.0</b>	.400	<b>3.0-5.0-7.5-10.0</b>	.398	<b>4.0-5.0-6.5-8.0</b>	.393
<b>3.0-4.0-5.5-10.0</b>	.397	<b>3.0-5.5-8.0-10.0</b>	.398	<b>4.0-5.0-6.5-8.5</b>	.397
<b>3.0-4.0-6.0-8.0</b>	.398	<b>3.5-4.5-5.5-6.5</b>	.368	<b>4.0-5.0-6.5-9.0</b>	.399
<b>3.0-4.0-6.0-8.5</b>	.400	<b>3.5-4.5-5.5-7.0</b>	.378	<b>4.0-5.0-6.5-10.0</b>	.397
<b>3.0-4.0-6.0-9.0</b>	.400	<b>3.5-4.5-5.5-7.5</b>	.387	<b>4.0-5.0-7.0-9.0</b>	.399
<b>3.0-4.0-6.0-10.0</b>	.398	<b>3.5-4.5-5.5-8.0</b>	.391	<b>4.0-5.0-7.0-9.5</b>	.399
<b>3.0-4.0-6.5-8.5</b>	.399	<b>3.5-4.5-5.5-9.0</b>	.396	<b>4.0-5.0-7.0-10.0</b>	.398
<b>3.0-4.0-6.5-9.0</b>	.400	<b>3.5-4.5-5.5-10.0</b>	.395	<b>4.0-5.5-7.0-8.5</b>	.398
<b>3.0-4.0-6.5-10.0</b>	.397	<b>3.5-4.5-6.0-7.5</b>	.387	<b>4.0-5.5-7.0-9.0</b>	.399
<b>3.0-4.0-7.0-10.0</b>	.397	<b>3.5-4.5-6.0-8.0</b>	.391	<b>4.0-5.5-7.0-10.0</b>	.398
<b>3.0-4.5-5.5-6.5</b>	.383	<b>3.5-4.5-6.0-8.5</b>	.395	<b>4.5-5.5-6.5-7.5</b>	.390
<b>3.0-4.5-5.5-7.0</b>	.389	<b>3.5-4.5-6.0-9.0</b>	.397	<b>4.5-5.5-6.5-8.0</b>	.393
<b>3.0-4.5-5.5-7.5</b>	.396	<b>3.5-4.5-6.0-10.0</b>	.395	<b>4.5-5.5-6.5-9.0</b>	.398
<b>3.0-4.5-5.5-8.0</b>	.397	<b>3.5-4.5-6.5-8.5</b>	.395	<b>4.5-5.5-6.5-10.0</b>	.397
<b>3.0-4.5-5.5-9.0</b>	.399	<b>3.5-4.5-6.5-9.0</b>	.396	<b>4.5-5.5-7.0-8.5</b>	.397
<b>3.0-4.5-5.5-10.0</b>	.397	<b>3.5-4.5-6.5-10.0</b>	.395	<b>4.5-5.5-7.0-9.0</b>	.398
<b>3.0-4.5-6.0-7.5</b>	.396	<b>3.5-4.5-7.0-9.5</b>	.397	<b>4.5-5.5-7.0-10.0</b>	.397
<b>3.0-4.5-6.0-8.0</b>	.397	<b>3.5-4.5-7.0-10.0</b>	.395	<b>4.5-6.0-7.5-9.0</b>	.398
<b>3.0-4.5-6.0-8.5</b>	.399	<b>3.5-5.0-6.5-8.0</b>	.392	<b>4.5-6.0-7.5-10.0</b>	.397
<b>3.0-4.5-6.0-9.0</b>	.400	<b>3.5-5.0-6.5-8.5</b>	.396	<b>4.5-6.5-8.5-10.0</b>	.396
<b>3.0-4.5-6.0-10.0</b>	.397	<b>3.5-5.0-6.5-9.0</b>	.397		

Table S4. Average correlations on 20 testing sets for the knowledge-based potentials trained on the corresponding 20 training sets with five shells.

<b>Shells configuration</b>	<b>&lt;R&gt;</b>	<b>Shells configuration</b>	<b>&lt;R&gt;</b>
<b>3.0-4.0-5.0-6.0-7.0</b>	.388	<b>3.5-4.5-5.5-6.5-9.0</b>	.396
<b>3.0-4.0-5.0-6.0-7.5</b>	.395	<b>3.5-4.5-5.5-6.5-10.0</b>	.395
<b>3.0-4.0-5.0-6.0-8.0</b>	.397	<b>3.5-4.5-5.5-7.0-8.5</b>	.395
<b>3.0-4.0-5.0-6.0-9.0</b>	.400	<b>3.5-4.5-5.5-7.0-9.0</b>	.396
<b>3.0-4.0-5.0-6.0-10.0</b>	.397	<b>3.5-4.5-5.5-7.0-10.0</b>	.395
<b>3.0-4.0-5.0-6.5-8.0</b>	.397	<b>3.5-4.5-6.0-7.5-9.0</b>	.397
<b>3.0-4.0-5.0-6.5-8.5</b>	.399	<b>3.5-4.5-6.0-7.5-10.0</b>	.396
<b>3.0-4.0-5.0-6.5-9.0</b>	.400	<b>3.5-5.0-6.5-8.0-9.5</b>	.397
<b>3.0-4.0-5.0-6.5-10.0</b>	.397	<b>3.5-5.0-6.5-8.0-10.0</b>	.396
<b>3.0-4.0-5.0-7.0-9.0</b>	.400	<b>4.0-5.0-6.0-7.0-8.0</b>	.393
<b>3.0-4.0-5.0-7.0-10.0</b>	.397	<b>4.0-5.0-6.0-7.0-8.5</b>	.397
<b>3.0-4.0-5.5-7.0-8.5</b>	.400	<b>4.0-5.0-6.0-7.0-9.0</b>	.399
<b>3.0-4.0-5.5-7.0-9.0</b>	.400	<b>4.0-5.0-6.0-7.0-10.0</b>	.398
<b>3.0-4.0-5.5-7.0-10.0</b>	.398	<b>4.0-5.0-6.0-7.5-9.0</b>	.399
<b>3.0-4.0-5.5-7.5-9.0</b>	.401	<b>4.0-5.0-6.0-7.5-9.5</b>	.399
<b>3.0-4.0-5.5-7.5-10.0</b>	.398	<b>4.0-5.0-6.0-7.5-10.0</b>	.398
<b>3.0-4.0-6.0-8.0-10.0</b>	.398	<b>4.0-5.0-6.5-8.0-9.5</b>	.399
<b>3.0-4.5-6.0-7.5-9.0</b>	.400	<b>4.0-5.0-6.5-8.0-10.0</b>	.397
<b>3.0-4.5-6.0-7.5-9.5</b>	.399	<b>4.0-5.5-7.0-8.5-10.0</b>	.398
<b>3.0-4.5-6.0-7.5-10.0</b>	.397	<b>4.5-5.5-6.5-7.5-8.5</b>	.396
<b>3.0-4.5-6.0-8.0-10.0</b>	.397	<b>4.5-5.5-6.5-7.5-9.0</b>	.398
<b>3.0-4.5-6.5-8.5-10.0</b>	.397	<b>4.5-5.5-6.5-7.5-10.0</b>	.397
<b>3.5-4.5-5.5-6.5-7.5</b>	.387	<b>4.5-5.5-6.5-8.0-9.5</b>	.398
<b>3.5-4.5-5.5-6.5-8.0</b>	.391	<b>4.5-5.5-6.5-8.0-10.0</b>	.397
<b>3.5-4.5-5.5-6.5-8.5</b>	.394		

### 3. Influence of the slope of the contact function

We present below the average correlation of the KBP2016 scoring function with experimental data made on the 20 testing sets for different intervals  $\Delta r$  of the contact function (see Figure S1 below). This value defines the slope of the function. Only a little influence is observed and we have kept a value of  $\Delta r = 0.2 \text{ \AA}$ . Since this value is the same for all shells, two atoms that are separated by a distance between  $r_2 - \Delta r/2$  and  $r_2 + \Delta r/2$  still have a total contact number of 1 if  $r_2$  is a distance that separates two shells (for example, the two atoms can be at 60% in one shell and 40% in the other).

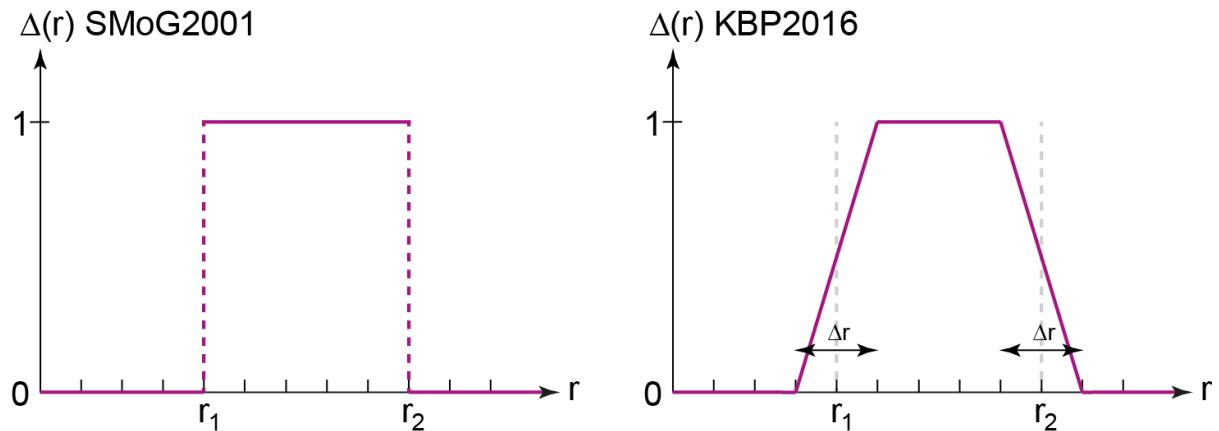


Figure S1. Contact function for the scoring functions.

Table S5. Influence of  $\Delta r$  on the average correlation of the scoring function made with the shell configuration 3.0-5.0-8.5  $\text{\AA}$ .

$\Delta r (\text{\AA})$	0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
$R$	0.400	0.401	0.400	0.400	0.399	0.399	0.398	0.398	0.399

#### 4. Influence of the $\alpha$ parameter on the correlation for hybrid scores with $\Delta\text{SASA}$

Table S6. Influence of the  $\alpha$  coefficient on the probability to find the correct rotamer with the  $\Delta\text{SASA}$  scoring function. For each target angle, we report the percentage of times the rotamer with the lowest score falls into the target angle range (%(success)). Amongst the rotamers which falls into the target angle range, we also report the fraction for which only 1 out of the 36 rotamers has the lowest score (%(g=1)).

$\alpha$	$\theta_{opt} = 0^\circ$		$\theta_{opt} \in \{-10^\circ, 10^\circ\}$		$\theta_{opt} \in \{-20^\circ, 20^\circ\}$	
	% (success)	% (g=1)	% (success)	% (g=1)	% (success)	% (g=1)
<b>0.00</b>	24	30	38	54	48	64
<b>0.05</b>	36	53	56	70	66	74
<b>0.10</b>	40	57	58	71	69	76
<b>0.15</b>	43	60	61	72	70	76
<b>0.20</b>	44	62	61	73	71	76
<b>0.25</b>	45	63	62	73	73	77
<b>0.30</b>	46	63	63	73	73	77
<b>0.35</b>	47	64	64	74	73	77
<b>0.40</b>	47	65	65	74	74	77
<b>0.45</b>	48	65	65	74	74	78
<b>0.50</b>	47	65	66	75	75	78
<b>0.55</b>	48	65	67	75	75	78
<b>0.60</b>	48	65	66	75	75	78
<b>0.65</b>	48	66	67	75	75	78
<b>0.70</b>	48	65	67	75	75	78
<b>0.75</b>	48	65	68	75	75	78
<b>0.80</b>	49	66	67	75	75	78
<b>0.85</b>	49	66	67	75	75	78
<b>0.90</b>	49	66	68	75	75	78
<b>0.95</b>	49	66	68	75	75	78
<b>1.00</b>	49	66	68	75	75	78
<b>1.05</b>	49	66	68	75	76	78
<b>1.10</b>	49	66	68	76	76	78
<b>1.15</b>	49	66	68	75	76	78
<b>1.20</b>	49	66	68	76	76	78
<b>1.25</b>	49	66	68	76	76	78
<b>1.30</b>	49	66	69	76	76	78
<b>1.35</b>	49	66	69	76	76	78
<b>1.40</b>	49	66	69	76	76	78
<b>1.45</b>	49	66	69	76	75	78
<b>1.50</b>	49	66	69	76	75	78

## 5. List of complexes contained in the testing set (195 complexes)

10gs	1o3f	1zea	2qbr	2xdl	3bfu	3g0w	3mfv	3ueu
1a30	1o5b	2brb	2qft	2xhm	3bkk	3g2n	3mss	3uem
1bcu	1os0	2cbj	2qmj	2xnb	3bpc	3g2z	3muz	3uo4
1e66	1oyt	2cet	2r23	2xy9	3cft	3gbg	3myg	3uri
1f8b	1p1q	2d1o	2v00	2xys	3cj2	3gcs	3n7a	3utu
1f8c	1ps3	2d3u	2v7a	2y5h	3coy	3ge7	3n86	3vd4
1f8d	1q8t	2fvd	2vl4	2yfe	3cyx	3gnw	3nox	3vh9
1gpk	1q8u	2g70	2vo5	2yge	3d4z	3gy4	3nq3	3zso
1h23	1qi0	2gss	2vot	2yki	3dd0	3huc	3nw9	3zsx
1hfs	1r5y	2hb1	2vvn	2ymd	3dxg	3i3b	3oe5	4de1
1hnn	1sln	2iwx	2vw5	2zcq	3e93	3imc	3ov1	4de2
1igj	1sqa	2j62	2w66	2zcr	3ebp	3ivg	3owj	4des
1jyq	1u1b	2j78	2wbg	2zjw	3ehy	3jvs	3ozt	4dew
1kel	1u33	2jdm	2wca	2zwz	3ejr	3k5v	3pe2	4djr
1lbk	1uto	2jdu	2weg	2zx6	3fl7	3kgp	3pww	4djv
1lol	1vso	2jdy	2wtv	2zxd	3f3a	3kv2	3pxf	4g8m
1loq	1w3k	2obf	2x00	3acw	3f3c	3kwa	3s8o	4gid
1lor	1w3l	2ole	2x0y	3ag9	3f3e	3l3n	3su2	4gqq
1mq6	1w4o	2p4y	2x8z	3ao4	3f80	3l4u	3su3	4tmn
1n1m	1xd0	2pcp	2x97	3b3s	3fcq	3l4w	3su5	
1n2v	1yc1	2pq9	2xb8	3b3w	3fk1	3l7b	3u9q	
1nvq	1z95	2qbp	2xbv	3b68	3fv1	3lka	3udh	

## 6. List of complexes contained in the training set (1,038 complexes)

4a7i	1upf	4loi	6abp	4iaw	1yon	4j4v	3k02	3q8h
3jup	1xz8	1x9d	3bzf	3uu1	3n9s	4ase	3e6y	2ce9
4hw3	4b0j	1aj7	4nku	4f7v	2wpb	3oyw	3u51	2xzw
1n8u	4f8j	3rx7	2wc4	1t13	1njj	1pzp	3iae	4mz6
3h78	966c	4kx8	4bqh	3hjo	3kku	4prg	4e8y	4fcf
5yas	1ngw	3rt8	4ly9	2wuf	4bkt	2b17	4cc7	2xpc
3gta	4his	3pd4	1yej	1sbr	2w8j	3zdv	3evf	2gv7
1ybo	4cgi	4c1u	1dfa	4btx	4ha5	2v11	2xg9	3ppr
4mjo	4dxg	2zpk	1pg2	1iup	4a0j	1utc	2i3z	3gsm
4mfl	1b55	1f2p	3smq	115s	3ijg	2x7x	1lgt	3e9i
4ewn	4fmu	4kln	1s5z	1y2g	2doo	1cze	4kne	2y1w
3o7u	1fl3	4ido	3rm9	1p0y	2rkn	1n3i	4igr	1j4r
2f10	4arb	4gue	3r1v	2ioa	3ps1	1lhw	4fzj	2g2r
2q89	2z3h	3fmz	2b2v	3aqt	1vea	2idk	2jk9	3pma
2a14	2xtn	3q2a	4i2z	2o4r	1x7a	4fkk	4gv8	3juk
1w2h	2v95	3n2c	1uxa	4dn0	1ceb	4agq	3hmp	6std
2yoh	1rpj	3q8d	2xm2	2qg0	3az9	3pgl	2vba	3kr2
1eoc	3d94	4hyb	2wva	1y01	1fpp	1pzi	1f9g	1fmb
2jg8	4bxk	2vnf	1pl0	1fl6	3ozp	2uy5	2o9r	4kao
1yvm	3srg	3b95	1p4u	3i9g	1lkl	4gah	3ipu	1p28
3hfv	1sm3	1d6s	1t5a	4i8n	1joc	3f6g	3gst	2oym
3tay	1c3x	1a0q	2xv1	2ivu	1moq	4mr4	1b74	1nlt
3poa	4da5	2nsj	3pjg	1oai	4kyk	3bex	2vi5	2h5a
2yek	1rdn	2i6b	2ayr	4lch	4fqo	3exe	2wk6	2p3a
4ad3	1h7a	4ipn	3njq	4fz3	4h85	3u7n	2i80	2q9y
2a31	4b76	4g01	3lc5	3h91	4gz	3uug	1css	3o4l
3nti	2ay9	4lko	2qhd	1pgp	1y20	1x39	1hp0	3pp7
1f74	3upk	4a6l	3zns	1sw1	4btm	3f51	1q91	4hpi
2z97	3mp6	3n7s	2cvd	830c	4b1j	3f8f	3stj	3lxk
3rl7	2wtx	2nnq	4lpb	3cz1	1d5r	6tim	2wf5	3e3c
3rjw	4b5w	3e8u	3at1	1upk	4hnn	2ygu	3tll	3bbb
2wq5	2zxg	2q2n	1aqj	4jwk	1x8j	3ies	1ajq	2mas
3ouj	1qm4	2izx	3u0p	1wur	2v77	4nj3	3e73	2nwl
1u0g	3qt7	2e95	3dla	3t2w	4dgb	1koj	4lmn	4irx
3qvv	4az5	3iqi	2jb6	3uzj	1oba	3wjw	4hv	3afk
3td4	4fab	3zhf	3vd7	2igw	8cpa	1yci	4i54	3rwq
1xap	1a5h	3uw4	4apo	1x07	1b57	2w0s	4b5d	2ra6
4fn5	3p7i	3l4z	4hzm	3u3z	3c39	3tvw	3ohi	3aqa
3shj	3r42	1nnu	4e3b	2w9h	2flr	2a2g	3eu5	4gwi
4edu	3qcf	1qsc	4hwb	5apr	1p6e	4lxd	3veh	4jdf
2itp	4axd	5tmp	4lrr	3pb7	1ro7	3i51	7kme	1qji
3nsq	3tfu	2oaz	1ctu	4n07	1dy4	1db1	4a7j	1af6
3zpr	1mmr	2w0z	4lkf	1y3p	1atl	1bdl	1ssq	1w1g
2wed	3suum	4m48	1jp5	1b56	2wm0	3mtw	3nin	2q8h
4j03	3r0y	3pn3	3nht	3mi2	4m0f	2hpa	1udu	3fhs
6cgt	1hsl	1byk	2e27	1y4z	2xgs	1etz	1drv	1jzs
3fsj	3upv	4db7	3t70	3e85	3t4p	1rgl	3r5t	3k05
2zxb	3f7g	2yz3	1gyy	2a58	4djh	2c1p	4app	1clu

1zc9	3chg	2nn7	4e28	3aaq	2hqu	2e7f	1odj	1h62
2c4w	4c1m	2yim	3bbt	3tl0	3tam	2imd	1ii5	3sm1
4bny	4e1k	4alx	3b5j	3tcp	2ya8	1wcq	2csn	2ccc
2bvd	2qr1	3b67	3qzv	1oth	2yfa	2xpk	1ui0	2pql
1x6u	4fxq	2uvm	3ewc	4dt2	1d1p	3o5x	2y4m	2jew
4igk	1pz5	4ebt	1pu8	1ecq	1x8d	4ge9	3o0g	3zv7
4l1u	4gby	3uw1	4dcy	3hav	4hu1	3f78	2yiv	3n7h
2iu0	4dsu	3ebb	3tsd	4ew3	2q7q	4mme	1e72	3esj
1kc7	1m2x	3fhb	4bc5	4lar	3unn	1w96	3g0e	3cke
1gt5	4bt5	3dne	2toh	2i0e	1awi	1nj5	3pwk	4b6q
2p7a	4i47	3hkw	3nyn	1td7	4eqj	4ewh	4jvb	4mo8
3dzt	3fed	4jr0	3lm1	4jin	3luo	3ozg	5tmn	2pfy
2ov4	3m3z	2pu1	2v59	3m89	1lnm	1ele	4dv8	1duv
3u15	6gpb	3s0e	1k03	3vw7	4hxw	2csm	4mss	1s50
3fqe	4as9	3gcu	4b7r	1ckb	2fqt	4glx	3f82	1cet
3feg	4bbh	2p53	3l3m	2o3z	3gx0	2rkm	4k5y	4lh3
4de3	3vjc	3m2u	4n70	1t7d	2a25	4kp0	16pk	4dfb
3ifo	4i9u	4bbg	1d2e	1hn4	3guz	1w70	4fht	1dub
1wm1	3k37	1v1j	1al8	3rti	1rm8	3own	3hb4	6upj
2w3o	2pmc	2o1c	2qwf	1c6y	3fj7	4gj3	2hf8	3zxe
3uxd	3eb1	3aho	1qpb	4ks5	2aow	1m13	1axs	2wq4
4ai5	3v31	3a2c	4kox	2hrm	1ibg	3kqr	3rdv	2ot1
1ct8	1ork	1u1w	4ayt	2az5	4hym	1qw7	1u9l	1uj6
5er2	3gp0	3h6z	2qfu	2oi2	3sud	2qdt	4ad6	2xyx
2qx0	4h5e	1amk	3s8n	3jrs	3c2u	3dri	1h2u	3a5y
4nks	3hfb	3p8z	1rlj	3t6b	3ai8	3te5	1ur9	4bi6
1osv	4hxj	3ik1	3vws	1qm5	3zy5	1ugp	4fyh	1k27
3dp9	3tc5	1fd0	4aro	2wp1	1akq	3k4q	3rux	2hxm
1kc5	3ivv	3ip8	3p2h	3sfg	3pty	3hzy	4nbn	4kcx
3rsr	3g91	2afx	4elh	4bqs	2gfd	1ik4	1ryh	1tyr
1efy	3up2	3lhs	1imx	1ppi	2clk	3bla	2qry	3mct
3cx9	3nyd	2wos	1j17	1fiv	4asj	1r9l	2z7g	4kn2
1m5w	1pr5	4fxp	1dgm	3k8d	3tg5	4dbm	3o5n	4jff
4l51	3cdb	1qaw	1xn	3zq9	3m94	1rql	1nu3	3n2e
4bps	1m7y	4hmq	1ll4	3d3x	4guj	4n6h	1q9d	4j86
1alw	3n3g	1yuc	4ksp	2h6b	2isc	1h8s	3eks	2f2h
1fch	3iog	3fh7	1dzk	2bjm	4b9w	4g3g	3tt4	2flh
1nf8	3t6r	1mh5	2oei	1c5c	2r0y	3ejq	1qca	1q1g
4lwc	3c2r	2vr4	2oz6	1juy	4glw	3o64	1dl7	1srg
2d0k	4eo4	3d0e	1dqx	2vmd	4c0r	2br6	3ri1	1pvn
2r75	1e2k	3vw2	4j93	2v2v	3m37	1ebg	4b3d	2eg8
4i1r	3iaw	2x4z	3s5y	3g19	3pm1	4b8y	4c73	3gds
1qin	1nq7	2xcg	1pot	4ih7	4g81	4fys	4f20	3pck
1h81	1fo0	3ikc	4hev	3h15	2vw2	2v88	4hj2	2pqb
3dow	2wl4	3ary	3dba	3lpl	1fkw	3imy	1a4k	4kzu
2xda	4mdn	3s9y	2wyn	3axk	1n1t	2ewb	2ovz	3fuc
2iyf	4ca5	4do4	1fq8	4kww	1tkc	1ind	3k5x	3ip9
1gar	4hws	1qyg	4loo	3aox	3a6t	3uyr	3ocz	3b50
1qbu	2qbw	3u3u	3p8h	2iuz	1m1b	1sqc	3udd	4hnc
4ear	3uj9	3tzm	2bmk	1gny	1a4r	3dix	4f3k	1qy2

4b9h	3ut5	3u90	4gki	4kp8	2psj	1lrh	4hf4	3nmq
4gr8	1sr7	4dq2	1sps	4jyv	4h4b	3kmm	3k5i	2hzl
2ceq	4kmd	3v5t	1kll	4btb	4in9	1utj	4jym	4ijq
2jfz	4dld	2y7k	1h6h	1c12	3o56	3zh8	3pwd	1ikt
3kyq	3sk2	3lpg	1fwv	4gg7	1j01	1qon	4k55	3lbz
3tgs	3b3x	2ig0	3m8t	4g4p	3ad8	3ryw	2vvp	1jak
4k5p	4bds	2qtr	2r9b	4k77	1hlk	1ogx	3aje	4dcv
2evl	3acl	3omg	3rf5	1xk5	1lox	1grp	8a3h	3jvr
4ayr	3rbq	1txr	4aq6	3ozs	3o84	2aj8	4io8	3hp9
3r5m	4ly1	4eoh	3tcg	2q6f	2pnc	3ny3	2yb0	2gvv
3pn1	4g68	4duh	3uf9	4e70	4mlu	4jfw	1cbr	1zhy
4gqr	1f28	3jwr	3sjf	3ldp	2oax	1wc6	4jfm	1m48
2usn	1h1h	1jr1	4k3r	4mfe	2haw	4io6	4fcm	
3zi8	2we3	4i7m	1g74	3p8p	3ms9	4l4v	1r6n	
2o4h	2x09	3mkn	4f1q	1lyb	4jia	4gfd	4j7e	
4k3n	1evh	2pem	3o9p	2ww0	1xk9	3uda	3kze	
4k3h	2ri9	2oz5	1hv5	4bg6	3t0x	3oaf	3w0l	
1rbo	1bq4	4g8o	3atp	1kuk	3rw9	2vk2	2o7e	