

# Training a Scoring Function for the Alignment of Small Molecules

S. Ling Chan, Paul Labute

SUPPORTING INFORMATION to be PUBLISHED

## Section A

In the main manuscript, the scoring function was trained by considering the “top 25” scoring poses (see the first paragraph of the Results and Discussion section). If training was performed using the “top 100” scoring poses instead (but still excluding the validation sets), 5 of the 6 optimal parameters for the scoring function, as given in table 3, would remain the same. Only  $\alpha_a$  would change from  $0.5 \text{ \AA}^{-2}$  to  $0.35 \text{ \AA}^{-2}$ . As for results of the validation runs, for the 28 rigid body alignments of the thermolysin data set (table 3), the success rate would drop from all but 2 correct (93%) to all but 3 correct (89%). As for validation results on the CDK2, HIV, P38, ESR1, Trypsin, Rhinovirus systems (table 5), the percentage of correct results for the rigid body alignments would change from 40%, 85% 43%, 59%, 80%, 50% to 39%, 86%, 42%, 64%, 76%, 50%, while those for the flexible alignments would change from 22%, 16%, 30%, 41%, 61%, 50% to 21%, 15%, 32%, 41%, 61%, 50%. To summarize, the quality of the results seem little changed. But in any case, we believe that using “top 25” is more meaningful, as explained in the main manuscript.

## Section B

Tables a to h correspond to tables 2a to 2h in the main manuscript when all data, including those related to the test sets, were used.

**Table a.** Parameter Optimization for  $\alpha_a$  <sup>a</sup>

$\alpha_a$ ( $\text{\AA}^{-2}$ )	0.06	0.125	0.175	0.25	0.35	<b>0.5</b>	0.7
n25	3.02	4.04	4.38	4.62	4.74	<b>4.78</b>	4.71

<sup>a</sup> The scoring function at this stage is  $T_0$ , the volume overlap term. Refer to table 1 for details of  $T_0$ . This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $\alpha_a$ .

**Table b.** Parameter Optimization for  $w_{DA}$  <sup>a</sup>

$w_{DA}$	0	0.5	1	2	<b>4</b>	8	16	32
n25	4.78	5.22	5.54	5.90	<b>6.03</b>	5.78	5.32	4.90

<sup>a</sup> The scoring function at this stage is  $(T_0+T_1+T_2)$ . The new terms are  $T_1$  and  $T_2$ , which correspond to the attraction between hydrogen bond donor / acceptor atoms. Refer to table 1 for the details of these terms. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{DA}$ .

**Table c.** Parameter Optimization for  $w_{HH}$ , Using the MOE Definition for Hydrophobic Atoms <sup>a</sup>

$w_{HH}$	0	0.5	<b>1</b>	2	4	8	16
n25	6.03	6.26	<b>6.30</b>	6.18	5.84	5.39	4.99

<sup>a</sup> The scoring function at this stage is  $(T_0+T_1+T_2+T_3)$ . The new term is  $T_3$ . It corresponds to the attraction between hydrophobic atoms. Refer to table 1 for the details of this term. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{HH}$ .

**Table d.** Parameter Optimization for  $w_{HH}$ , where Hydrophobic Atoms Are Defined as Atoms That Are At Least Two Bonds Away from Any Hydrogen Bonding Atom <sup>a</sup>

$w_{HH}$	0	0.5	<b>1</b>	2	4	8
n25	6.03	6.24	<b>6.27</b>	6.13	5.77	5.25

<sup>a</sup> The scoring function at this stage is ( $T_0+T_1+T_2+T_3$ ). The new term is  $T_3$ . It corresponds to the attraction between hydrophobic atoms. Refer to table 1 for the details of this term. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{HH}$ .

**Table e.** Parameter Optimization for  $w_{RI}$ , Using the MOE Definition for Hydrophobic Atoms <sup>a</sup>

$w_{RI}$	0	-0.125	-0.25	-0.5	<b>-1</b>	-2	-4	-8
n25	6.30	6.32	6.34	6.38	<b>6.40</b>	6.31	5.87	4.54

<sup>a</sup> The scoring function at this stage is ( $T_0+T_1+T_2+T_3+T_4+T_5$ ). The new terms  $T_4$  and  $T_5$  correspond to the repulsion between hydrophobic atoms and hydrogen bonding atoms. Refer to table 1 for the details of these terms. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{RI}$ .

**Table f.** Parameter Optimization for  $w_{RI}$ , where Hydrophobic Atoms Are Defined as Atoms That Are At Least Two Bonds Away from Any Hydrogen Bonding Atom <sup>a</sup>

$w_{RI}$	0	-0.125	-0.25	<b>-0.5</b>	-1	-2	-4	-8
n25	6.27	6.28	6.29	<b>6.30</b>	6.29	6.19	5.85	5.17

<sup>a</sup> The scoring function at this stage is ( $T_0+T_1+T_2+T_3+T_4+T_5$ ). The new terms  $T_4$  and  $T_5$  correspond to the repulsion between hydrophobic atoms and hydrogen bonding atoms. Refer to table 1 for the details of these terms. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{RI}$ .

**Table g.** Parameter Optimization for  $w_{R2}$  <sup>a</sup>

$w_{R2}$	<b>0</b>	-0.125	-0.25	-0.5	-1
n25	<b>6.40</b>	6.39	6.39	6.39	6.37

<sup>a</sup> The scoring function at this stage is  $(T_0+T_1+T_2+T_3+T_4+T_5+T_6+T_7)$ . The new terms  $T_6$  and  $T_7$  correspond to the repulsion between hydrogen bond donor atoms that are not acceptors, and hydrogen bond acceptor atoms that are not donors. Refer to table 1 for the details of these terms. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{R2}$ .

**Table h.** Parameter Optimization for  $w_{PF}$  and  $\alpha_p$  <sup>a</sup>

For  $\alpha_p = 0.5 \text{ \AA}^{-2}$ :

$w_{PF}$	0	0.25	0.5	1	2	4	8
n25	6.40	6.42	6.44	6.46	6.42	6.30	5.95

For  $\alpha_p = 0.25 \text{ \AA}^{-2}$ :

$w_{PF}$	0	0.25	0.5	1	2	4	8
n25	6.40	6.45	6.48	6.53	6.55	6.45	6.14

For  $\alpha_p = 0.125 \text{ \AA}^{-2}$ :

$w_{PF}$	0	0.25	0.5	1	<b>2</b>	4	8
n25	6.40	6.48	6.55	6.61	<b>6.63</b>	6.52	6.19

For  $\alpha_p = 0.0625 \text{ \AA}^{-2}$ :

$w_{PF}$	0	0.25	0.5	1	2	4	8
n25	6.40	6.51	6.57	6.62	6.58	6.35	5.88

<sup>a</sup> The scoring function at this stage is  $(T_0+T_1+T_2+T_3+T_4+T_5+T_8+T_9)$ . The new terms  $T_8$  and  $T_9$  correspond to the attraction between hydrogen bond donor / acceptor projected features. Refer to table 1 for the details of these terms. This table gives the average number of good poses, n25, within the top 25 scoring poses for various values of  $w_{PF}$  and  $\alpha_p$ .