

## REPORT: BADOCK QUESTIONS

**Author:** Joaquim Aguirre Plans  
**E-mail:** [guim.aguirre@gmail.com](mailto:guim.aguirre@gmail.com)

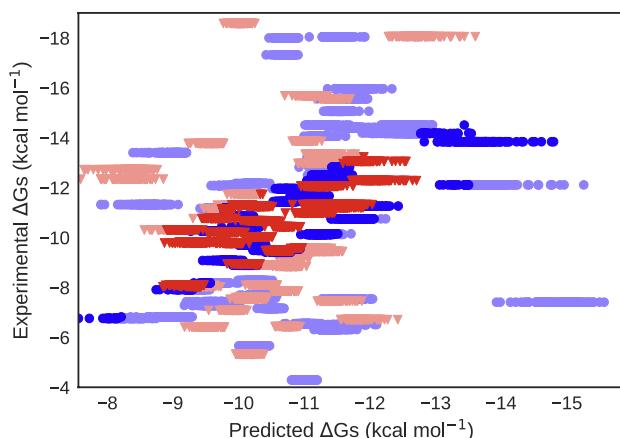
### **1. Reviewer 1 - Question 6**

**Q6 - It would be good to show as a plot the predicted vs. experimental dG values that provide the best correlations of Table 1.**

**QUIM: Select the best correlations from table 1 and plot. This also implies the selection from table 3 (see question Q3 of reviewer 3) and add the % with lower RMSE than 1.4 (and 2.8)**

**BACKUP/CHECK: Joan**

First, I have plotted the experimental vs. predicted energy values for the ES3DC potential. The strongly colored points are the ones corresponding to a difference of predicted vs. experimental energy smaller or equal than 1.4 Kcal/mol.



*Figure R1-Q6-1: Scatter plot between experimental and predicted  $\Delta G$  using the ES3DC potential. Predictions are made using the test sets of 1000 random ten-fold cross validation models with the ES3DC averaged scores of all docking poses in the AB2 dataset. The strongly colored points correspond to a difference of predicted vs. experimental energy smaller or equal than 1.4 Kcal/mol. The blue dots correspond to rigid cases of AB2 and the red triangles to flexible cases.*

On the other hand, we want to know what happens when we select the pairs of values “docking scores - experimental energies” that provide the best correlations. For this purpose, I have plotted the docking scores vs. the experimental energies for ES3DC, with its corresponding regression line. To select the values that provide the best regression, I have plotted two parallel lines of the regression line, at a distance of 1.4 Kcal/mol. The distance was decided to follow the trend of the other questions. Then, I have removed the points that were out of the area comprised by the two parallel lines.

Using the new pairs of values, I have calculated the predicted energies, and plotted the resulting predicted vs. experimental energies. The strongly colored points are the ones corresponding to a difference of predicted minus experimental energy smaller or equal than 1.4 Kcal/mol.

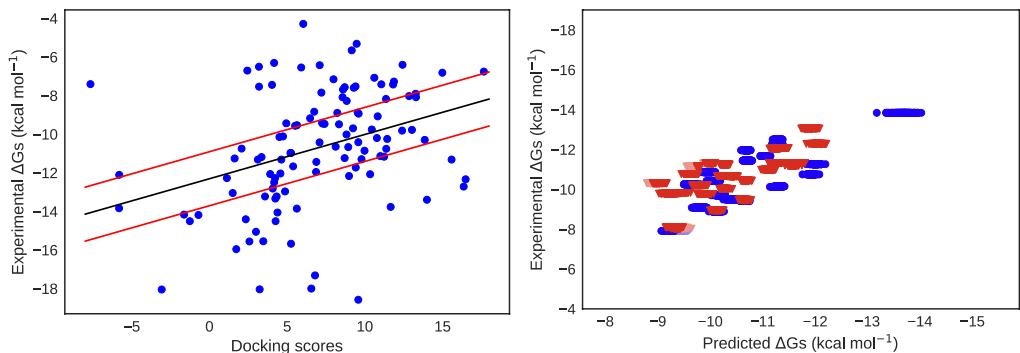


Figure R1-Q6-2 (left): Scatter plot between experimental  $\Delta G$  and the docking scores. The black line indicates the regression between the two distribution of values. The red lines indicate the limit of values which are in a distance equal or smaller than 1.4 Kcal/mol.

Figure R1-Q6-3 (right): Scatter plot between experimental and predicted  $\Delta G$  using the experimental  $\Delta G$  values which were in a distance equal or smaller than 1.4 Kcal/mol of the regression line. Predictions are made using the test sets of 1000 random ten-fold cross validation models with the ES3DC averaged scores of all docking poses in the AB2 dataset. The strongly colored points correspond to a difference of predicted vs. experimental energy smaller or equal than 1.4 Kcal/mol. The blue dots correspond to rigid cases of AB2 and the red triangles to flexible cases.

We can compare this plot with the initial one. The plot obtained from the removed points contains a higher proportion of points with difference  $\leq 1.4$  Kcal/mol (95.64%, 4782 / 5000) in the total analysis, in comparison with the initial (41.94%, 5033 / 12000).

I have done the same using 2.8 Kcal/mol. There is a proportion of 58.50% (5256 / 12000) of the points with difference  $\leq 1.4$  Kcal/mol in the total analysis.

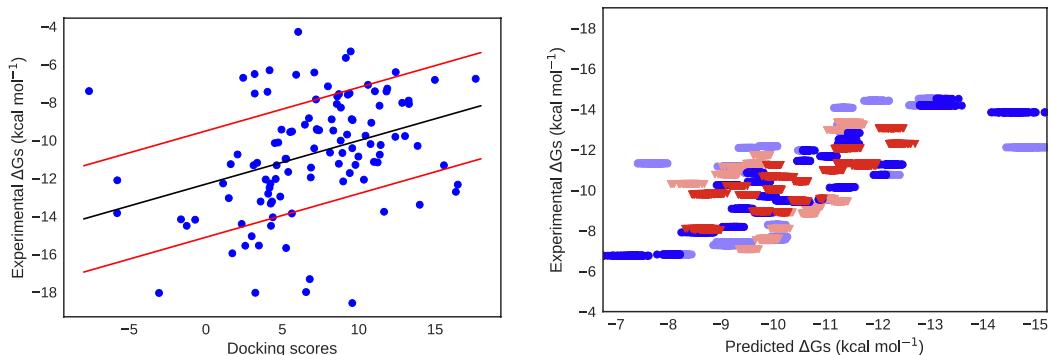
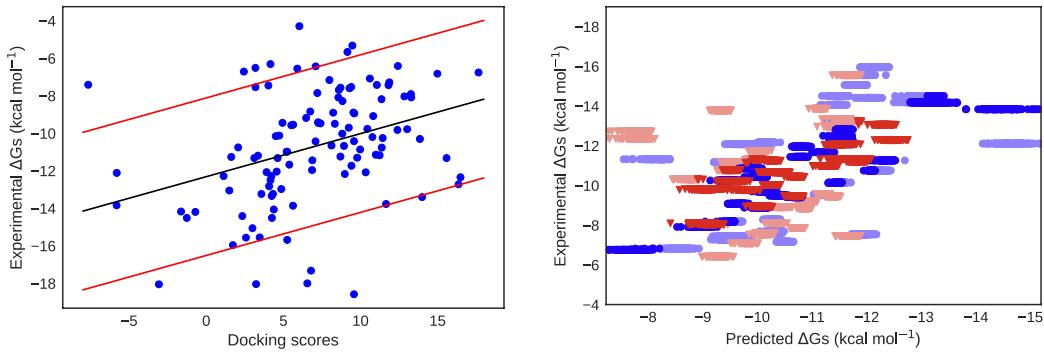


Figure R1-Q6-4 (left): Scatter plot between experimental  $\Delta G$  and the docking scores. The black line indicates the regression between the two distribution of values. The red lines indicate the limit of values which are in a distance equal or smaller than 2.8 Kcal/mol.

Figure R1-Q6-5 (right): Scatter plot between experimental and predicted  $\Delta G$  using the experimental  $\Delta G$  values which were in a distance equal or smaller than 2.8 Kcal/mol of the regression line. Predictions are made using the test sets of 1000 random ten-fold cross validation models with the ES3DC averaged scores of all docking poses in the AB2 dataset. The strongly colored points correspond to a difference of predicted vs. experimental energy smaller or equal than 1.4 Kcal/mol. The blue dots correspond to rigid cases of AB2 and the red triangles to flexible cases.

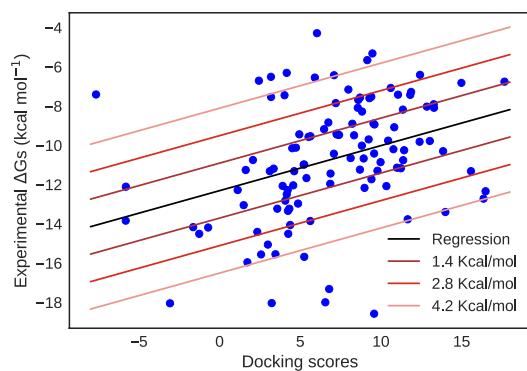
I have done the same using 4.2 Kcal/mol. There is a proportion of 50.18% (5018 / 12000) of the points with difference  $\leq 1.4$  Kcal/mol in the total analysis.



*Figure R1-Q6-6 (left): Scatter plot between experimental  $\Delta G$  and the docking scores. The black line indicates the regression between the two distribution of values. The red lines indicate the limit of values which are in a distance equal or smaller than 4.2 Kcal/mol.*

*Figure R1-Q6-7 (right): Scatter plot between experimental and predicted  $\Delta G$  using the experimental  $\Delta G$  values which were in a distance equal or smaller than 4.2 Kcal/mol of the regression line. Predictions are made using the test sets of 1000 random ten-fold cross validation models with the ES3DC averaged scores of all docking poses in the AB2 dataset. The strongly colored points correspond to a difference of predicted vs. experimental energy smaller or equal than 1.4 Kcal/mol. The blue dots correspond to rigid cases of AB2 and the red triangles to flexible cases.*

I have also done a combination of all the plots of experimental dG vs. docking scores:



*Figure R1-Q6-2-4-6: Scatter plot between experimental  $\Delta G$  and the docking scores. The black line indicates the regression between the two distribution of values. The red lines indicate the limit of values which are in a distance equal or smaller than 1.4, 2.8 and 4.2 Kcal/mol*

All the plots have been done using a modification of the script "affinity\_analysis.py", called "**affinity\_analysis\_R1Q6.py**". The instructions of how to run the script are in the README of the folder "BADock\_src\_quim".

## **2. Reviewer 2 - Question 2**

### **Q2. Page 3. Figure 1.**

The percentage of each decoy class would be of interest to show. It's an indication of the quality of docking and also the intrinsic weight of the total contribution for each decoy class.

**QUIM:** Check the number of cases for each class of the boxplots in Figure 1 and supplementary Figures S2, S3, S4, add them in the figure or legend to figures.

**BACKUP/CHECK:** Joan

I have counted the number of decoys for each class and the total number of decoys, and stored them in a Python dictionary. The result when printing the dictionary is:

```
{'I': 30725, 'P': 152224, 'total': 338094, 'E': 154895, 'N': 250}
```

And calculating the percentages, the results are:

- Total number of decoys for all proteins: 338094
- Decoys for **N** (near-native): 250 Percentage: **0.07 %**
- Decoys for **I** (face-face): 30725 Percentage: **9.09 %**
- Decoys for **P** (face-back): 152224 Percentage: **45.02 %**
- Decoys for **E** (back-back): 154895 Percentage: **45.81 %**

These percentages are the same for all the potentials, because the decoys are always the same.

The percentages have been added in the figure “BP\_ENE+Graph.pptx”.

The analysis has been done using a modification of the script “analyze\_scores.py” called “**analyze\_scores\_R2.py**”. The instructions of how to run the script are in the README of the folder “BADock\_src\_quim”.

### 3. Reviewer 3 - Question 3

**Q3 -- about experiments. reporting correlation coefficients is fine. But having the fraction of cases withing say 1.4 and 2.8 Kcal/mol is more interesting for practitioners: this give the accuracy on individual complexes. this is a good practice introduced in recent papers by Janin, Bonvin et al, Cazals et al; I think it should be adopted for all affinity predictions.**

**QUIM: Check individual complexes with predictions accurate (error < 1.4), medium (error <2.8) and get the fractions for Table S4 (we will added as a new table in supplementary)**

In Marillet et al. (<http://onlinelibrary.wiley.com/doi/10.1002/prot.24946/full>), they introduce the calculation of a statistic called **prediction ratio** ( $p_{\delta}^{error}$ ). The prediction ratio is defined as the percentage of cases such that the difference between the experimental and predicted free energies is equal or smaller than a specified amount  $\delta$ . They set  $\delta$  to 1.4, 2.8, and 4.2 Kcal/mol because it corresponds to cases “whose Kd is approximately within 1, 2, and 3 orders of magnitude respectively”.

To calculate the different values of prediction ratio, I have used a modification of the file “affinity\_analysis.py”, called “**affinity\_analysis\_R3.py**”. The script obtains the regression model from a machine learning process, and produces predicted values using the model and the docking scores. The values are stored into Pickle files. For comparing our scoring functions with other scoring functions, I have used the script “**CCharPPI\_affinityparse\_R3.py**”. Then, a script called “**percent\_analysis\_R3.py**” is used to obtain the table of prediction ratios.

The instructions of how to run the scripts are in the README of the folder “BADock\_src\_quim”.

Following the format of table S4, I proposed to create a supplementary table of the same format, but containing the values of prediction ratio for  $\delta$  1.4, 2.8, 4.2 and more than 4.2 Kcal/mol. The percentages are non-cumulative. It also contains the absolute number of predictions for each prediction ratio in the column *abs*.

*Supplementary Table R3-Q3-1. Calculation of the prediction ratio for the scores of the docking poses*

	$\delta$ (Kcal/mol)	AB2				AB2 Rigid				AB2 Flexible			
		Native		All decoys		Native		All decoys		Native		All decoys	
		%	abs	%	abs	%	abs	%	abs	%	abs	%	abs
FiberDock	$\delta \leq 1.4$	38.30	36	39.36	37	29.17	14	29.17	14	47.83	22	50.00	23
	$1.4 < \delta \leq 2.8$	29.79	28	31.91	30	27.08	13	35.42	17	32.61	15	28.26	13
	$2.8 < \delta \leq 4.2$	17.02	16	14.89	14	31.25	15	20.83	10	2.17	1	8.70	4
	$\delta > 4.2$	14.89	14	13.83	13	12.50	6	14.58	7	17.39	8	13.04	6
aVdW	$\delta \leq 1.4$	41.49	39	36.17	34	35.42	17	33.33	16	47.83	22	39.13	18
	$1.4 < \delta \leq 2.8$	25.53	24	28.72	27	25.00	12	22.92	11	26.09	12	34.78	16
	$2.8 < \delta \leq 4.2$	20.21	19	20.21	19	27.08	13	27.08	13	13.04	6	13.04	6
	$\delta > 4.2$	12.77	12	14.89	14	12.50	6	16.67	8	13.04	6	13.04	6
rVdW	$\delta \leq 1.4$	38.30	36	35.11	33	33.33	16	31.25	15	43.48	20	39.13	18
	$1.4 < \delta \leq 2.8$	20.21	19	29.79	28	12.50	6	22.92	11	28.26	13	36.96	17
	$2.8 < \delta \leq 4.2$	23.40	22	18.09	17	33.33	16	25.00	12	13.04	6	10.87	5
	$\delta > 4.2$	18.09	17	17.02	16	20.83	10	20.83	10	15.22	7	13.04	6
aElec	$\delta \leq 1.4$	37.23	35	41.49	39	33.33	16	33.33	16	41.30	19	50.00	23
	$1.4 < \delta \leq 2.8$	24.47	23	28.72	27	18.75	9	29.17	14	30.43	14	28.26	13
	$2.8 < \delta \leq 4.2$	19.15	18	13.83	13	25.00	12	22.92	11	13.04	6	4.35	2
	$\delta > 4.2$	19.15	18	15.96	15	22.92	11	14.58	7	15.22	7	17.39	8
rElec	$\delta \leq 1.4$	37.23	35	39.36	37	35.42	17	29.17	14	39.13	18	50.00	23
	$1.4 < \delta \leq 2.8$	23.40	22	29.79	28	14.58	7	31.25	15	32.61	15	28.26	13
	$2.8 < \delta \leq 4.2$	21.28	20	14.89	14	29.17	14	25.00	12	13.04	6	4.35	2
	$\delta > 4.2$	18.09	17	15.96	15	20.83	10	14.58	7	15.22	7	17.39	8

laElec	$\delta \leq 1.4$	37.23	35	42.55	40	33.33	16	35.42	17	41.30	19	50.00	23
	$1.4 < \delta \leq 2.8$	23.40	22	26.60	25	18.75	9	27.08	13	28.26	13	26.09	12
	$2.8 < \delta \leq 4.2$	21.28	20	15.96	15	27.08	13	22.92	11	15.22	7	8.70	4
	$\delta > 4.2$	18.09	17	14.89	14	20.83	10	14.58	7	15.22	7	15.22	7
lrElec	$\delta \leq 1.4$	38.30	36	41.49	39	33.33	16	33.33	16	43.48	20	50.00	23
	$1.4 < \delta \leq 2.8$	18.09	17	28.72	27	12.50	6	29.17	14	23.91	11	28.26	13
	$2.8 < \delta \leq 4.2$	25.53	24	14.89	14	33.33	16	22.92	11	17.39	8	6.52	3
	$\delta > 4.2$	18.09	17	14.89	14	20.83	10	14.58	7	15.22	7	15.22	7
HB	$\delta \leq 1.4$	37.23	35	38.30	36	31.25	15	33.33	16	43.48	20	43.48	20
	$1.4 < \delta \leq 2.8$	29.79	28	19.15	18	29.17	14	12.50	6	30.43	14	26.09	12
	$2.8 < \delta \leq 4.2$	15.96	15	25.53	24	16.67	8	35.42	17	15.22	7	15.22	7
	$\delta > 4.2$	17.02	16	17.02	16	22.92	11	18.75	9	10.87	5	15.22	7
EPAIR	$\delta \leq 1.4$	32.98	31	41.49	39	29.17	14	35.42	17	36.96	17	47.83	22
	$1.4 < \delta \leq 2.8$	29.79	28	26.60	25	20.83	10	27.08	13	39.13	18	26.09	12
	$2.8 < \delta \leq 4.2$	19.15	18	15.96	15	29.17	14	16.67	8	8.70	4	15.22	7
	$\delta > 4.2$	18.09	17	15.96	15	20.83	10	20.83	10	15.22	7	10.87	5
ES3DC	$\delta \leq 1.4$	38.30	36	46.81	44	33.33	16	39.58	19	43.48	20	54.35	25
	$1.4 < \delta \leq 2.8$	27.66	26	26.60	25	22.92	11	31.25	15	32.61	15	21.74	10
	$2.8 < \delta \leq 4.2$	19.15	18	9.57	9	27.08	13	10.42	5	10.87	5	8.70	4
	$\delta > 4.2$	14.89	14	17.02	16	16.67	8	18.75	9	13.04	6	15.22	7
E3D	$\delta \leq 1.4$	36.17	34	38.30	36	33.33	16	31.25	15	39.13	18	45.65	21
	$1.4 < \delta \leq 2.8$	30.85	29	20.21	19	25.00	12	16.67	8	36.96	17	23.91	11
	$2.8 < \delta \leq 4.2$	20.21	19	24.47	23	29.17	14	31.25	15	10.87	5	17.39	8
	$\delta > 4.2$	12.77	12	17.02	16	12.50	6	20.83	10	13.04	6	13.04	6
ZRANK	$\delta \leq 1.4$	38.30	36	-	-	35.42	17	-	-	41.30	19	-	-
	$1.4 < \delta \leq 2.8$	25.53	24	-	-	16.67	8	-	-	34.78	16	-	-
	$2.8 < \delta \leq 4.2$	22.34	21	-	-	33.33	16	-	-	10.87	5	-	-
	$\delta > 4.2$	13.83	13	-	-	14.58	7	-	-	13.04	6	-	-
ZRANK2	$\delta \leq 1.4$	38.30	36	-	-	35.42	17	-	-	41.30	19	-	-
	$1.4 < \delta \leq 2.8$	22.34	21	-	-	16.67	8	-	-	28.26	13	-	-
	$2.8 < \delta \leq 4.2$	24.47	23	-	-	31.25	15	-	-	17.39	8	-	-
	$\delta > 4.2$	14.89	14	-	-	16.67	8	-	-	13.04	6	-	-
RosettaDock	$\delta \leq 1.4$	37.23	35	-	-	33.33	16	-	-	41.30	19	-	-
	$1.4 < \delta \leq 2.8$	29.79	28	-	-	27.08	13	-	-	32.61	15	-	-
	$2.8 < \delta \leq 4.2$	17.02	16	-	-	20.83	10	-	-	13.04	6	-	-
	$\delta > 4.2$	15.96	15	-	-	18.75	9	-	-	13.04	6	-	-
PyDock	$\delta \leq 1.4$	38.30	36	-	-	29.17	14	-	-	47.83	22	-	-
	$1.4 < \delta \leq 2.8$	22.34	21	-	-	22.92	11	-	-	21.74	10	-	-
	$2.8 < \delta \leq 4.2$	25.53	24	-	-	35.42	17	-	-	15.22	7	-	-
	$\delta > 4.2$	13.83	13	-	-	12.50	6	-	-	15.22	7	-	-
PISA	$\delta \leq 1.4$	35.11	33	-	-	25.00	12	-	-	45.65	21	-	-
	$1.4 < \delta \leq 2.8$	30.85	29	-	-	33.33	16	-	-	28.26	13	-	-
	$2.8 < \delta \leq 4.2$	21.28	20	-	-	33.33	16	-	-	8.70	4	-	-
	$\delta > 4.2$	12.77	12	-	-	8.33	4	-	-	17.39	8	-	-
PIE	$\delta \leq 1.4$	41.49	39	-	-	33.33	16	-	-	50.00	23	-	-
	$1.4 < \delta \leq 2.8$	23.40	22	-	-	22.92	11	-	-	23.91	11	-	-
	$2.8 < \delta \leq 4.2$	19.15	18	-	-	29.17	14	-	-	8.70	4	-	-
	$\delta > 4.2$	15.96	15	-	-	14.58	7	-	-	17.39	8	-	-
SIPPER	$\delta \leq 1.4$	41.49	39	-	-	31.25	15	-	-	52.17	24	-	-
	$1.4 < \delta \leq 2.8$	22.34	21	-	-	25.00	12	-	-	19.57	9	-	-
	$2.8 < \delta \leq 4.2$	20.21	19	-	-	25.00	12	-	-	15.22	7	-	-
	$\delta > 4.2$	15.96	15	-	-	18.75	9	-	-	13.04	6	-	-

For testing purposes, I have done the same table but now, only with two rows: larger and smaller than 2.8 Kcal/mol:

*Supplementary Table R3-Q3-2. Calculation of the prediction ratio for the scores of the docking poses*

	$\delta$ (KCal/mol)	AB2				AB2 Rigid				AB2 Flexible			
		Native		All decoys		Native		All decoys		Native		All decoys	
		%	abs	%	abs	%	abs	%	abs	%	abs	%	abs
FiberDock	$\delta \leq 2.8$	68.09	64	71.28	67	56.25	27	64.58	31	80.43	37	78.26	36
	$\delta > 2.8$	31.91	30	28.72	27	43.75	21	35.42	17	19.57	9	21.74	10
aVdW	$\delta \leq 2.8$	67.02	63	64.89	61	60.42	29	56.25	27	73.91	34	73.91	34
	$\delta > 2.8$	32.98	31	35.11	33	39.58	19	43.75	21	26.09	12	26.09	12
rVdW	$\delta \leq 2.8$	58.51	55	64.89	61	45.83	22	54.17	26	71.74	33	76.09	35
	$\delta > 2.8$	41.49	39	35.11	33	54.17	26	45.83	22	28.26	13	23.91	11
aElec	$\delta \leq 2.8$	61.70	58	70.21	66	52.08	25	62.50	30	71.74	33	78.26	36
	$\delta > 2.8$	38.30	36	29.79	28	47.92	23	37.50	18	28.26	13	21.74	10
rElec	$\delta \leq 2.8$	60.64	57	69.15	65	50.00	24	60.42	29	71.74	33	78.26	36
	$\delta > 2.8$	39.36	37	30.85	29	50.00	24	39.58	19	28.26	13	21.74	10
laElec	$\delta \leq 2.8$	60.64	57	69.15	65	52.08	25	62.50	30	69.57	32	76.09	35
	$\delta > 2.8$	39.36	37	30.85	29	47.92	23	37.50	18	30.43	14	23.91	11
lrElec	$\delta \leq 2.8$	56.38	53	70.21	66	45.83	22	62.50	30	67.39	31	78.26	36
	$\delta > 2.8$	43.62	41	29.79	28	54.17	26	37.50	18	32.61	15	21.74	10
HB	$\delta \leq 2.8$	67.02	63	57.45	54	60.42	29	45.83	22	73.91	34	69.57	32
	$\delta > 2.8$	32.98	31	42.55	40	39.58	19	54.17	26	26.09	12	30.43	14
EPAIR	$\delta \leq 2.8$	62.77	59	68.09	64	50.00	24	62.50	30	76.09	35	73.91	34
	$\delta > 2.8$	37.23	35	31.91	30	50.00	24	37.50	18	23.91	11	26.09	12
ES3DC	$\delta \leq 2.8$	65.96	62	73.40	69	56.25	27	70.83	34	76.09	35	76.09	35
	$\delta > 2.8$	34.04	32	26.60	25	43.75	21	29.17	14	23.91	11	23.91	11
E3D	$\delta \leq 2.8$	67.02	63	58.51	55	58.33	28	47.92	23	76.09	35	69.57	32
	$\delta > 2.8$	32.98	31	41.49	39	41.67	20	52.08	25	23.91	11	30.43	14
ZRANK	$\delta \leq 2.8$	63.83	60	-	-	52.08	25	-	-	76.09	35	-	-
	$\delta > 2.8$	36.17	34	-	-	47.92	23	-	-	23.91	11	-	-
ZRANK2	$\delta \leq 2.8$	60.64	57	-	-	52.08	25	-	-	69.57	32	-	-
	$\delta > 2.8$	39.36	37	-	-	47.92	23	-	-	30.43	14	-	-
ROSSETADOCK	$\delta \leq 2.8$	67.02	63	-	-	60.42	29	-	-	73.91	34	-	-
	$\delta > 2.8$	32.98	31	-	-	39.58	19	-	-	26.09	12	-	-
PyDock	$\delta \leq 2.8$	60.64	57	-	-	52.08	25	-	-	69.57	32	-	-
	$\delta > 2.8$	39.36	37	-	-	47.92	23	-	-	30.43	14	-	-
PISA	$\delta \leq 2.8$	65.96	62	-	-	58.33	28	-	-	73.91	34	-	-
	$\delta > 2.8$	34.04	32	-	-	41.67	20	-	-	26.09	12	-	-
PIE	$\delta \leq 2.8$	64.89	61	-	-	56.25	27	-	-	73.91	34	-	-
	$\delta > 2.8$	35.11	33	-	-	43.75	21	-	-	26.09	12	-	-
SIPPER	$\delta \leq 2.8$	63.83	60	-	-	56.25	27	-	-	71.74	33	-	-
	$\delta > 2.8$	36.17	34	-	-	43.75	21	-	-	28.26	13	-	-

#### 4. Reviewer 3 - Question 11

Q11 \* having correlations is important, but not sufficient I think. From the supplemental, one sees that the RMSE are comparable. therefore, you should report for each comparison (e.g. ES3DC / Native versus ES3DC / Poses) (i) the p-value of a two-sample test to compare the two series of estimates, and (ii) an effect size. For example, in using the Wilcoxon Mann Whitney test, you could provide the Hodge-Lehman estimate of the population shift.

This is important since it might be the case that estimates on Native and Poses are indistinguishable... in this range of accuracy (RMSE in the range 2.5 - 3 Kcal/mol).

**JOAN (helped by Quim): calculate mann-Whitney/Hodge-Lehmann shift between Native-FaceFace etc. for ES3DC,**

I have calculated a Mann-Whitney test for the distribution of predicted values using only the Native poses, and the distribution of predicted values using all the poses. I have done it for the predicted values  $\leq 1.4$  Kcal/mol,  $\leq 2.8$  Kcal/mol,  $\leq 4.2$  Kcal/mol and  $> 4.2$  Kcal/mol, and for all our scoring functions.

To perform this analysis, I have used the same script as the previous question, “**percent\_analysis\_R3.py**”. The instructions of how to run the scripts are in the README of the folder “**BADock\_src\_quim**”.

$\delta$ (KCal/mol)	AB2		AB2 Rigid		AB2 Flexible		
	p-value	estimate	p-value	estimate	p-value	estimate	
FiberDock	$\delta \leq 1.4$	4.39E-01	0.1659	6.35E-01	0.1601	6.94E-01	0.1724
	$1.4 < \delta \leq 2.8$	8.47E-01	0.0871	5.09E-01	0.2186	4.13E-01	-0.5728
	$2.8 < \delta \leq 4.2$	8.22E-01	-0.1336	9.78E-01	-0.0315	8.00E-01	0.8794
	$\delta > 4.2$	6.16E-01	-0.1624	7.31E-01	0.1061	2.82E-01	-0.6036
aVdW	$\delta \leq 1.4$	6.09E-01	-0.0838	2.92E-01	0.2189	1.25E-01	-0.3454
	$1.4 < \delta \leq 2.8$	3.73E-01	0.2039	1.51E-01	0.6568	8.02E-01	-0.0565
	$2.8 < \delta \leq 4.2$	7.73E-01	0.0948	4.18E-01	0.2071	6.99E-01	-0.3848
	$\delta > 4.2$	1.06E-01	0.4227	6.66E-04	0.7996	4.85E-01	-0.2077
rVdW	$\delta \leq 1.4$	1.86E-01	0.0390	3.46E-04	0.1242	1.46E-01	-0.1532
	$1.4 < \delta \leq 2.8$	8.17E-01	0.0230	2.45E-01	0.0902	5.20E-01	-0.0999
	$2.8 < \delta \leq 4.2$	2.25E-02	0.1101	1.23E-01	0.0386	1.04E-01	0.2128
	$\delta > 4.2$	1.61E-04	0.2586	1.11E-04	0.3294	3.39E-01	0.0475
aElec	$\delta \leq 1.4$	4.57E-01	0.2068	2.14E-01	0.3377	8.99E-01	-0.0239
	$1.4 < \delta \leq 2.8$	8.29E-01	-0.1739	9.25E-01	0.0554	8.23E-01	-0.3118
	$2.8 < \delta \leq 4.2$	5.59E-01	-0.1150	9.75E-01	0.0062	1.81E-01	-0.3289
	$\delta > 4.2$	1.04E-02	0.2832	5.41E-04	0.5736	7.72E-01	0.0245
rElec	$\delta \leq 1.4$	2.05E-01	0.1963	2.64E-01	0.2343	4.03E-01	0.1730
	$1.4 < \delta \leq 2.8$	5.10E-01	0.0727	3.24E-01	0.4334	8.14E-01	-0.4879
	$2.8 < \delta \leq 4.2$	3.34E-01	-0.1741	8.17E-01	-0.0538	6.03E-02	-0.3167
	$\delta > 4.2$	1.36E-02	0.1742	5.79E-04	0.5405	8.60E-01	0.0091
laElec	$\delta \leq 1.4$	5.52E-01	0.1653	2.63E-01	0.2844	8.79E-01	-0.0289
	$1.4 < \delta \leq 2.8$	8.71E-01	-0.1023	8.41E-01	0.0097	6.73E-01	-0.1023
	$2.8 < \delta \leq 4.2$	3.39E-01	-0.1330	9.08E-01	-0.0481	5.97E-02	-0.4256
	$\delta > 4.2$	2.91E-02	0.2877	6.52E-04	0.5897	8.98E-01	-0.0196
IrElec	$\delta \leq 1.4$	7.75E-02	0.3277	1.59E-02	0.4292	8.16E-01	0.2153
	$1.4 < \delta \leq 2.8$	8.47E-01	-0.2850	1.00E+00	-0.0593	7.71E-01	-0.3682
	$2.8 < \delta \leq 4.2$	5.33E-01	-0.0958	6.39E-01	0.0565	1.38E-02	-0.2579
	$\delta > 4.2$	3.89E-03	0.3528	6.52E-04	0.7255	5.13E-01	0.1532
HB	$\delta \leq 1.4$	3.54E-01	0.1274	7.37E-01	-0.1160	1.26E-01	0.2520
	$1.4 < \delta \leq 2.8$	8.50E-02	0.2249	5.92E-01	0.1875	5.37E-02	0.2705
	$2.8 < \delta \leq 4.2$	1.06E-02	0.4828	2.32E-01	0.1992	2.89E-02	0.6348
	$\delta > 4.2$	4.78E-02	-0.2825	8.05E-02	-0.3906	7.55E-01	-0.0501
EPAIR	$\delta \leq 1.4$	7.51E-01	0.0867	4.93E-01	0.2796	9.44E-01	-0.0117

	1.4 < $\delta \leq$ 2.8	3.00E-01	-0.2108	3.13E-01	-0.4230	7.23E-01	-0.0404
	2.8 < $\delta \leq$ 4.2	1.66E-01	-0.3304	8.68E-01	0.0740	7.27E-02	-0.7724
	$\delta > 4.2$	5.02E-01	0.1877	6.31E-01	-0.1127	5.05E-03	0.5558
ES3DC	$\delta \leq 1.4$	6.41E-01	-0.1171	9.61E-01	0.0270	4.90E-01	-0.2321
	1.4 < $\delta \leq$ 2.8	6.20E-01	-0.1792	6.10E-01	-0.2443	5.67E-01	-0.2393
	2.8 < $\delta \leq$ 4.2	9.40E-01	0.1014	1.73E-01	1.0079	2.86E-01	-1.1376
	$\delta > 4.2$	6.08E-01	0.1281	3.21E-01	0.2410	8.36E-01	-0.3147
E3D	$\delta \leq 1.4$	7.84E-01	0.0445	5.72E-01	0.1402	8.35E-01	-0.0613
	1.4 < $\delta \leq$ 2.8	8.35E-01	-0.0332	5.71E-01	0.5519	4.04E-01	-0.1367
	2.8 < $\delta \leq$ 4.2	6.47E-02	0.2540	1.37E-02	0.2875	7.24E-01	-0.1469
	$\delta > 4.2$	7.67E-01	-0.0669	8.75E-01	0.0303	3.94E-01	-0.4570

I have calculated the same but using only two rows:

	$\delta$ (KCal/mol)	AB2		AB2 Rigid		AB2 Flexible	
		p-value	estimate	p-value	estimate	p-value	estimate
FiberDock	$\delta \leq 2.8$	5.12E-01	0.1332	3.60E-01	0.2562	8.65E-01	-0.0498
	$\delta > 2.8$	5.17E-01	-0.1542	7.72E-01	0.0883	2.11E-01	-0.4965
aVdW	$\delta \leq 2.8$	8.26E-01	0.0281	6.12E-02	0.3426	1.69E-01	-0.2385
	$\delta > 2.8$	2.23E-01	0.2080	6.35E-03	0.4513	2.19E-01	-0.3907
rVdW	$\delta \leq 2.8$	2.52E-01	0.0389	5.47E-04	0.1104	1.44E-01	-0.1065
	$\delta > 2.8$	1.31E-05	0.1520	8.74E-05	0.1548	5.43E-02	0.1101
aElec	$\delta \leq 2.8$	6.76E-01	0.0442	3.44E-01	0.2801	7.56E-01	-0.0322
	$\delta > 2.8$	1.19E-01	0.1536	4.41E-02	0.4028	9.24E-01	-0.1112
rElec	$\delta \leq 2.8$	1.49E-01	0.1523	1.19E-01	0.3280	5.88E-01	0.0728
	$\delta > 2.8$	3.02E-01	0.0696	1.30E-01	0.4417	7.28E-01	-0.0318
laElec	$\delta \leq 2.8$	6.74E-01	0.0411	2.16E-01	0.2829	6.16E-01	-0.0513
	$\delta > 2.8$	3.36E-01	0.1133	5.94E-02	0.4039	1.91E-01	-0.1521
lrElec	$\delta \leq 2.8$	2.45E-01	0.2179	1.04E-01	0.2776	1.00E+00	0.0709
	$\delta > 2.8$	7.89E-02	0.1902	1.08E-02	0.5740	5.86E-01	-0.0887
HB	$\delta \leq 2.8$	5.38E-02	0.1964	7.83E-01	0.0963	1.30E-02	0.2629
	$\delta > 2.8$	8.12E-01	0.0478	3.64E-01	-0.1108	1.57E-01	0.3375
EPAIR	$\delta \leq 2.8$	6.80E-01	-0.0517	8.43E-01	-0.0842	7.34E-01	-0.0393
	$\delta > 2.8$	6.52E-01	-0.0763	7.53E-01	-0.0446	9.76E-01	-0.0148
ES3DC	$\delta \leq 2.8$	5.32E-01	-0.1241	8.80E-01	-0.0434	4.07E-01	-0.2059
	$\delta > 2.8$	4.48E-01	0.1888	7.13E-02	0.6606	3.65E-01	-0.5838
E3D	$\delta \leq 2.8$	9.10E-01	0.0132	3.62E-01	0.2726	4.73E-01	-0.0969
	$\delta > 2.8$	2.25E-01	0.1445	2.49E-02	0.2090	2.91E-01	-0.3836

## **5. Reviewer 2 - Question 1**

### **Q1. Page 2, 2.2 Docking, refinement and scoring**

Some parameters and details of docking should be mentioned, such as the number of poses retained from PatchDock and FiberDock, the sampling interval for angle, etc.

**MANU:** Explain the default in method, add the configuration files of a dummy example in supplementary material

**BACKUP/CHECK:** Quim and Joan

**Joan:**

We amended the text to make such details clear:

*"PatchDock (Schneidman-Duhovny, et al., 2005) was used (default parameters) to obtain the docking poses (or decoys), which were all kept and ranked according to a geometric shape complementary score (rigid docking)"*

**Quim:** I have found a PatchDock execution in the server of BADock. It uses the following parameters file:

patchdock\_params.txt:

```
receptorPdb /sbi/users/webservices/BADock/dg_57c708e48de16/1AY7_r_u.pdb
ligandPdb /sbi/users/webservices/BADock/dg_57c708e48de16/1AY7_l_u.pdb
protLib /soft/bio/structure/PatchDock/chem.lib
log-file /sbi/users/webservices/BADock/dg_57c708e48de16/patchdock_log.txt
log-level 2
receptorSeg 10.0 20.0 1.5 1 0 1 0
ligandSeg 10.0 20.0 1.5 1 0 1 0
scoreParams 0.3 -5.0 0.5 0.0 0.0 1500 -8 -4 0 1 0
desolvationParams 500.0 1.0
clusterParams 0.1 4 2.0 8.0
baseParams 4.0 13.0 2
matchingParams 1.5 1.5 0.4 0.5 0.9
matchAlgorithm 1
receptorGrid 0.5 6.0 6.0
ligandGrid 0.5 6.0 6.0
receptorMs 10.0 1.8
ligandMs 10.0 1.8
```

I have also found the FiberDock execution parameters:

fiberdock\_params.txt:

```
receptorPDBFileName /sbi/users/manuel/Databases/AffinityBenchmark1/1E4K_r_u.pdb.HB
ligandPDBFileName /sbi/users/manuel/Databases/AffinityBenchmark1/1E4K_l_u.pdb.HB
transFileName /sbi/users/manuel/Affinity_docking/ba_jobs_u/1E4K/patchdock_transform.clean.txt
rotamerLibFile /soft/bio/structure/FiberDock1.1/lib/bbdep02.May.sortlib
protLib /soft/bio/structure/FiberDock1.1/lib/chem.lib
energiesOutFileName /sbi/users/manuel/Affinity_docking/ba_jobs_u/1E4K/fiberdock_energies
printRefinedComplexes 0
onlyEnergyCalculation 0
receptorOnlyClashesMovable 1
ligandOnlyClashesMovable 1
extraRotamers 0
lpMethod 2
exclusionsFile /soft/bio/structure/FiberDock1.1/lib/exclusions
rigidBodyMinimizationCycles 10
charmmParamFile /soft/bio/structure/FiberDock1.1/lib/par_all27_prot_na.prm
definitionsFile /soft/bio/structure/FiberDock1.1/lib/top_all27_prot_na.rtf
```

```

attrVdWWeight 1.5
repVdWWeight 0.8
ACEWeight 1.6
attrEIWeight 0.21
repEIWeight 0.21
l_attrEIWeight 0.0
l_repEIWeight 0.69
HBWeight 1.2
pipiWeight 1.0
catpiWeight 0.7
aliphWeight 2.5
insidenessWeight 0.7
confProbWeight 0.0
radiiScaling 0.80
performBackboneRef 1
Cycles 10
RecAnalyzedFile /sbi/users/manuel/Databases/AffinityBenchmark1/1E4K_r_u.pdb.CA
ReceptorNMs /sbi/users/manuel/Affinity_docking/ba_jobs_u/1E4K/1E4K_r_u.nma
Lambda 0.05
LigAnalyzedFile /sbi/users/manuel/Databases/AffinityBenchmark1/1E4K_l_u.pdb.CA
LigandNMs /sbi/users/manuel/Affinity_docking/ba_jobs_u/1E4K/1E4K_l_u.nma
log-file /sbi/users/manuel/Affinity_docking/ba_jobs_u/1E4K/fiberdock_log.txt
log-level 2

```

## **6. Reviewer 3 - Question 12**

-- section 3, subsection 3.3

**Q12 \* Fig 3. tell us with which method the density estimate was carried out. due to smoothing, having the points themselves would probably be more informative.**

**MANU: add method from matplotlib with information on how it was smoothed the graph etc.**

**New plot for suppl. material**

**BACKUP: Quim and Joan**

The graph was plotted using the Seaborn function “**seaborn.kdeplot**”. It fits and plots a univariate or bivariate kernel density estimate.

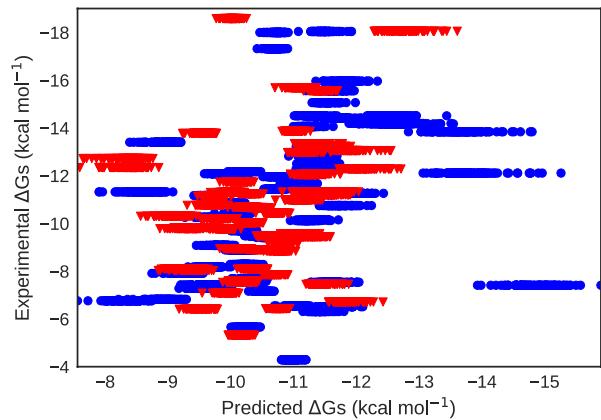
```

sns.kdeplot(np.array(R_predicted_G), np.array(R_experimental_G), shade_lowest=False,
cmap="Blues")
sns.kdeplot(np.array(F_predicted_G), np.array(F_experimental_G), shade_lowest=False,
cmap="Reds")
plt.savefig(self.img_path + '/BA_ES3DC_Docking.eps', bbox_inches='tight', dpi=350)

```

Information about the function: <http://seaborn.pydata.org/generated/seaborn.kdeplot.html>

The corresponding **scatter plot**:



## 7. Corrections

- Pages 6, 7, 8 of supplementary material: "*Least square's fitting curve is shown (slope and y-coordinate interception are in supplementary [table S8](#) for the sake of comparison).*" Would it be table S1?