# A General Theory of Fragment Linking in Molecular Design: Why Fragment Linking Rarely Succeeds and How to Improve Outcomes

## Supplementary Information:

**Haoyu S. Yu,[a], Kalyan Modugula,[b] Osamu Ichihara,[c] Kimberly Kramschuster,[d] Simon Keng,[d] Carolyn Carrie Mcquaw,[d] Robert Abel,[a] and Lingle Wang*[a]**

a) *Schrodinger, Inc., 120 West 45th Street, New York, New York 10036, United States*
b) *D. E. Shaw India Private Ltd., Plot No. 573, Jubilee Hills, Telangana, 500096, India*
c) *Schrodinger, Inc. 17F Marunouchi Trust Tower North, Chiyoda-ku, Tokyo 100-0005 Japan*
d) *Schrodinger, Inc., 101 SW Main Street, Suite 1300, Portland, Oregon, 97204, United States*

**Section A: Binding Mode Free Energy Prediction**

In order to calculate the binding free energy difference between two binding poses (pose A and pose B) of the same ligand, we annihilate the ligand in pose A and grow back the ligand in pose B in our FEP workflow. The binding poses are restrained at the two end points in our lambda schedule.

**Section B: Linker Enumeration Free Energy Prediction**

In order to calculate the free energy difference between ligands that have same fragments A and B but different linkers, we applied the free energy perturbation protocol such that the fragments A and B on the two ends of the ligand are mapped, the linker which is in the middle part of the ligand is unmapped. In the FEP simulation, the linker is annihilated in one ligand and grow back in the other ligand through the core-hopping lambda schedule.

**Section C: FEP Results of Relative Binding Affinity of RPA and LDH systems**

Table S1. Predicted Relative Binding Free Energies of RPA systems, unit kcal/mol

| Ligand1 | Ligand2 | FEP | FEP Error | CCC | CCC Error[a] |
|---|---|---|---|---|---|
| cmp38 | cmp11 | 0.63 | 0.16 | 0.8 | 0.57 |
| cmp38 | cmp7i | 2.78 | 0.14 | 2.54 | 0.65 |
| cmp38 | cmp7h | 0.76 | 0.30 | 1.19 | 0.77 |
| cmp38 | cmp7g | 4.34 | 0.29 | 3.98 | 0.77 |
| cmp7h | cmp7e | 1.03 | 0.04 | 1.07 | 0.55 |
| cmp7h | cmp7a | 2.32 | 0.14 | 2.71 | 0.77 |
| cmp7e | cmp7a | 1.60 | 0.13 | 1.64 | 0.55 |
| cmp7e | cmp7f | -1.20 | 0.24 | -1.2 | 0.45 |
| cmp7a | cmp7i | -1.60 | 0.22 | -1.36 | 0.65 |
| cmp7a | cmp7b | -0.72 | 0.16 | -0.89 | 0.57 |
| cmp7a | cmp7g | -0.27 | 0.13 | 0.09 | 0.77 |
| cmp11 | cmp7f | 0.08 | 0.10 | 0.25 | 0.57 |
| cmp7f | cmp7b | 1.78 | 0.36 | 1.95 | 0.57 |

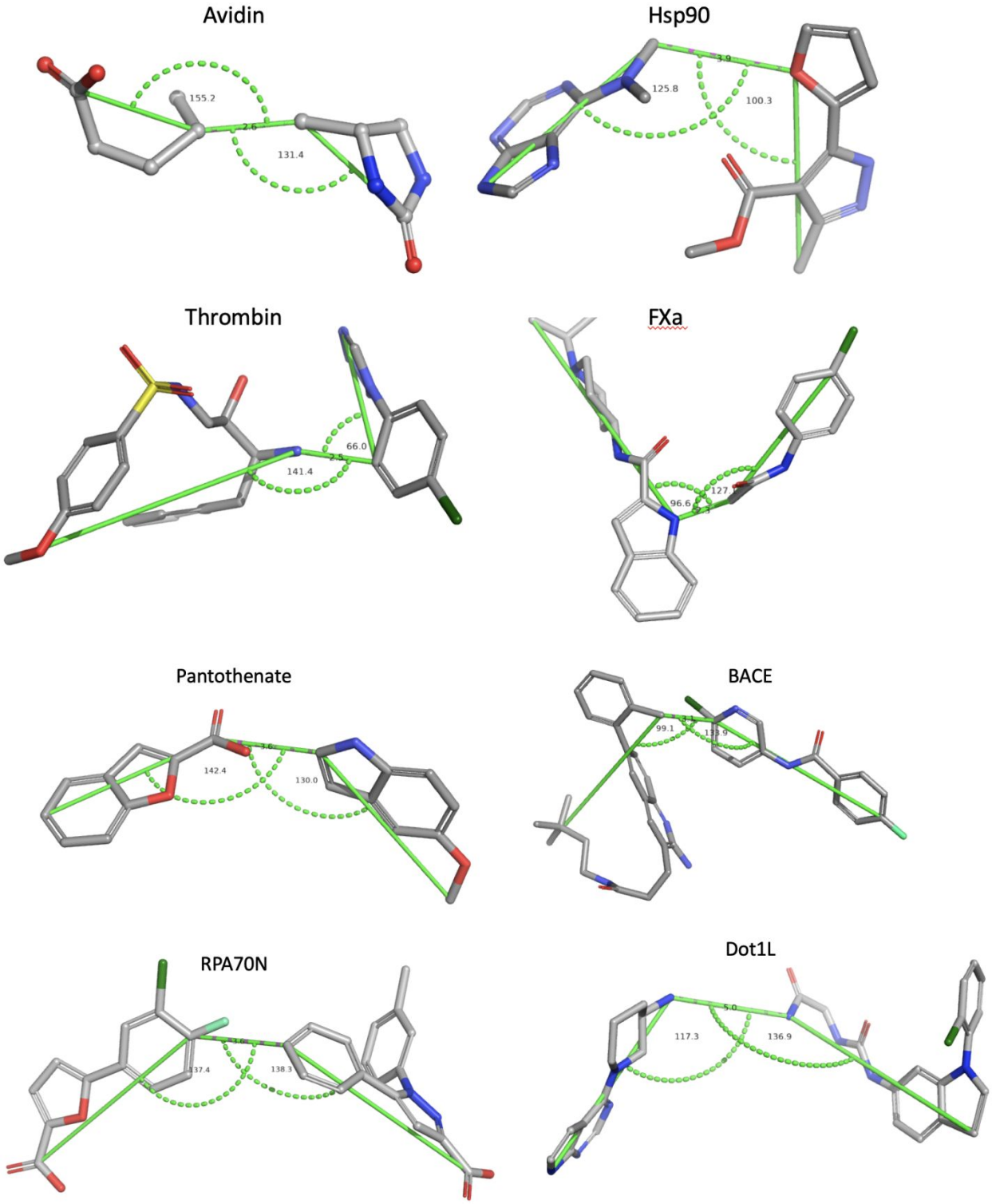
*a.* Cycle closure correction error, which is computed following the algorithm in reference 1,2.

Table S2. Predicted Relative Binding Free Energies of LDH systems, unit kcal/mol

| Ligand1 | Ligand2 | FEP | FEP Error | CCC | CCC Error[a] |
|---|---|---|---|---|---|
| lig11 | lig7 | -2.24 | 0.14 | -2.26 | 0.41 |
| lig11 | lig8 | -2.00 | 0.20 | -2.30 | 0.62 |
| lig11 | lig10 | -2.12 | 0.15 | -1.80 | 0.62 |
| lig9 | lig10 | 1.11 | 0.21 | 1.24 | 0.40 |
| lig9 | lig8 | 0.86 | 0.13 | 0.73 | 0.40 |
| lig7 | lig8 | -0.02 | 0.14 | -0.04 | 0.41 |
| lig8 | lig10 | 0.95 | 0.18 | 0.50 | 0.62 |

*a.* Cycle closure correction error, which is computed following the algorithm in reference 1,2.

Table S3. Angle and distance restraints of each fragment linking systems.



| Name | angle1[a] | angle2 | distance[b] | k(angle)[c] | k(distance)[d] |
|---|---|---|---|---|---|
| Avidin | 155.2 | 131.4 | 2.6 | 39 | 0.5 |
| Hsp90 | 125.8 | 100.3 | 3.9 | 39 | 0.5 |
| Thrombin | 141.4 | 66.0 | 2.5 | 39 | 0.5 |
| FXa | 96.6 | 127.1 | 2.3 | 39 | 0.5 |
| Pantothenate | 142.4 | 130.0 | 3.6 | 39 | 0.5 |
| BACE | 99.1 | 133.9 | 3.1 | 39 | 0.5 |
| RPA70N | 137.4 | 138.3 | 3.6 | 39 | 0.5 |
| Dot1L | 117.3 | 136.9 | 5.0 | 39 | 0.5 |

a. The unit of angle1 and angle2 is degree, which are shown in figure above.

b. The unit of distance is Å, which stand for distance between a and b atom in the figure above.

c. The unit of k(angle) is kcal/mol/rad$^2$

d. The unit of k(distance) is kcal/mol/A$^2$

**Section D: Binding Pose of ligands from RPA and LDH systems:**

Figure S4. RPA protein and ligand structures of compound 7a, 7b, 7e, 7f, 7g, 7h, 7i, 11, and 38.

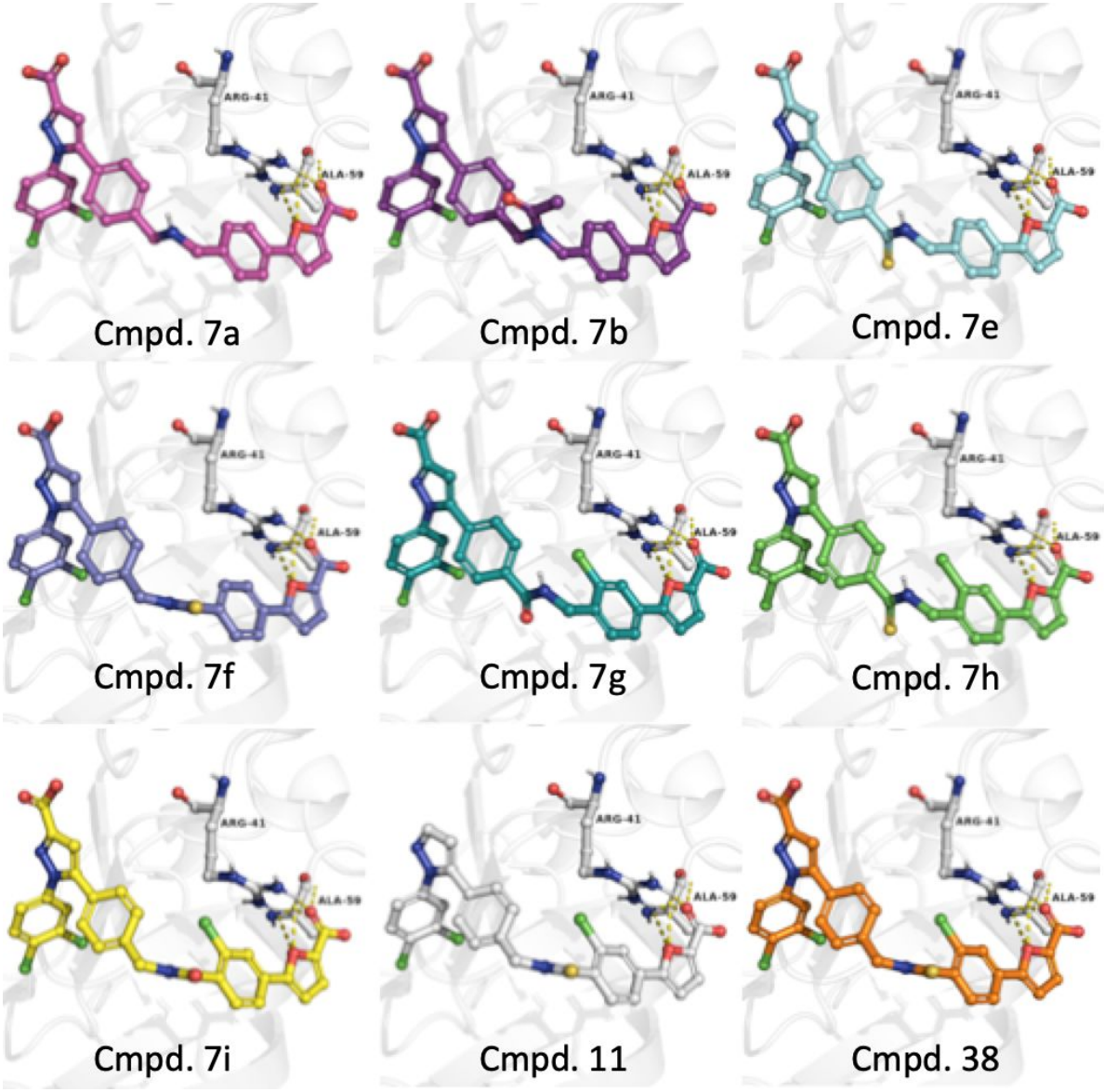


Figure S5. LDH protein and ligand structures of compound 7, 8, 9, 10, and 11.

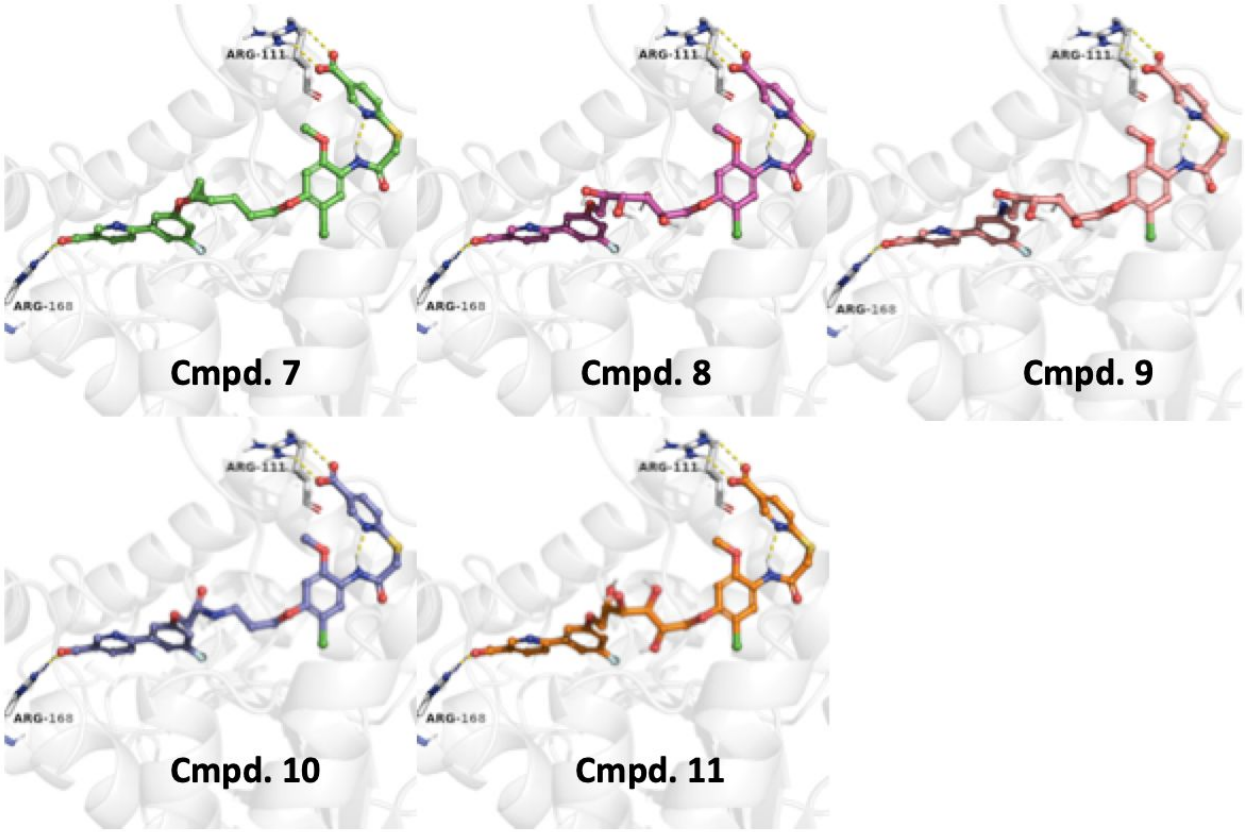
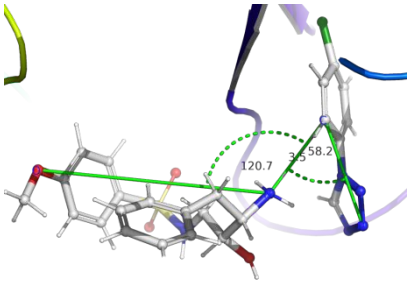
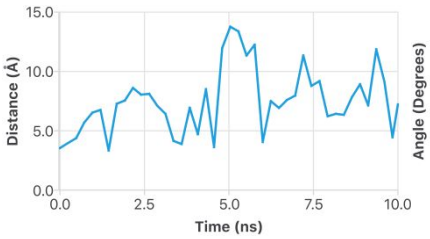


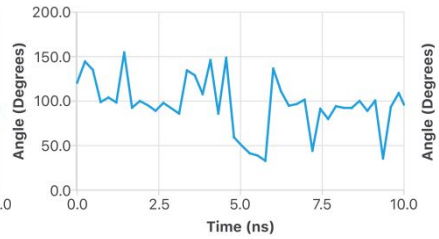
Figure S6. FEP complex trajectory time series plots of Thrombin system. Fragment A (methozybenzenesulfonamide) and fragment B (1-(4-Chlorophenyl)-1H-tetrazole) are shown in the figure S6-a. Two angles and one distance are used to estimate the flexibility of two fragments in the protein binding pocket. Distance plot and two angle plots are shown in figure S6-b, S6-c, S6-d respectively, where the distance is varying from 5.0 Å to 10.0 Å and two angles are varying from 50 to 150 degree and 40 to 60 degree. These plots showed that two ligands are relatively flexible in the protein binding pocket, which contribute to the relatively configurational entropy of the two fragments in the protein binding pocket opposing the binding of the linked molecule.



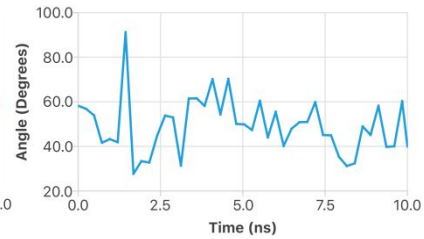

S6-a




S6-b

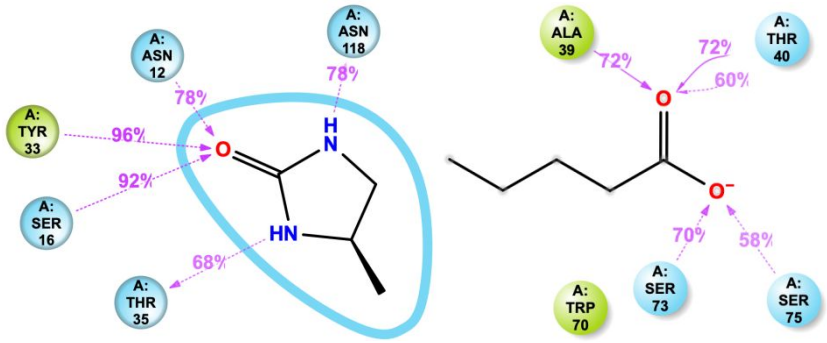



S6-c




S6-d

Figure S7. Protein Ligand Interaction for Avidin fragments 1,2-propylene urea and 1-pentanoate. The percentages are calculated on the basis of the number of frames that are forming strong intermolecular interaction with the residues in the binding pocket out of the total frames recorded in the molecular dynamics trajectory




**Section E: Input and output files of Fragment Linking FEP simulations**

The input and output files of the fragment linking FEP simulations are available for download from the following shared link:

https://drive.google.com/drive/folders/11-z9UMgANj03kclbYcRjoK_6NuwYwnlC

## REFERENCES

(1) L. Wang,, T. Lin., R. Abel. Cycle Closure Estimation of Relative Binding Affinities and Errors.

(2) L. Wang, Y. Deng, J. L. Knight, Y. Wu, B. Kim, W. Sherman, J. C. Shelley, T. Lin and R. Abel, *Journal of Chemical Theory and Computation*, 2013, **9**, 1282-1293.