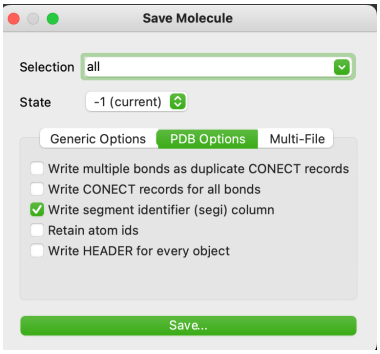
Preparation of crystal structure to run Hi-patch:
Software needed: PyMOL (v3.0)

1. Download structure from Protein Data Bank (https://www.rcsb.org/) in .pdb format (e.g., ABCD.pdb).
2. Open structure in PyMOL and strip any non-polypeptide objects and hydrogens. To do this, execute the following command: "remove hydrogen; remove solvent; remove org; remove inorg".
3. Next, create a new PDB file by clicking: File>Export Molecule…>(Selection: enabled; State -1(current))>Save…>Select .pdb format instead of .cif. (e.g., ABCD_clean.pdb).



4. Close PyMOL and re-open newly created file.
5. Using the command line, move current directory to where the file "areas.py" is located, by typing:
   >cd
   >cd Users/[input here the directory path].
   (for example: cd Users/hectorsm/hi-patch)
6. Run areas.py script by executing: "run areas.py". It will calculate the Solvent Accessible Surface Area (SASA) for each individual atom in the structure that is currently displayed in PyMOL. This command will typically freeze PyMOL until the calculation is finished. The elapsed time of calculation depends in the coarseness of the SASA calculation. This can be tuned by opening the script "areas.py" and changing the value of 'dot_density'. This value can go between 1-4, with 4 being the finest but also longest to calculate. Setting 'dot_density' to 3 yields highest quality to computational time in my experience. This is particularly important for proteins larger than ~300 residues.

```
import numpy as np

cmd.set('dot_solvent', 1)
cmd.set('dot_density', 3) # <-- Tune coarseness of SASA calc

num_atoms=cmd.count_atoms('all')
res=np.zeros(1)

for i in range(1, num_atoms+1):
        print(cmd.get_area('index %d' %i))
        res=np.append(res,cmd.get_area('index %d' %i))

np.savetxt('areas_results.txt',res)
```

7. Open Excel and open ABCD_clean.pdb as a text file. This will look like this:

	A	B	C	D	E	F	G	H
1	ATOM	1 N	GLU A	2	18.297 13.730 -3.131 1.00 0.00	N		
2	ATOM	2 CA	GLU A	2	18.993 14.932 -3.683 1.00 0.00	C		
3	ATOM	3 C	GLU A	2	19.498 15.737 -2.493 1.00 0.00	C		
4	ATOM	4 O	GLU A	2	18.811 15.848 -1.478 1.00 0.00	O		
5	ATOM	5 CB	GLU A	2	18.059 15.777 -4.551 1.00 0.00	C		
6	ATOM	6 CG	GLU A	2	18.684 17.014 -5.184 1.00 0.00	C		
7	ATOM	7 CD	GLU A	2	18.706 18.229 -4.270 1.00 0.00	C		
8	ATOM	8 OE1	GLU A	2	17.685 18.544 -3.615 1.00 0.00	O		
9	ATOM	9 OE2	GLU A	2	19.762 18.892 -4.189 1.00 0.00	O		
10	ATOM	10 N	MET A	3	20.691 16.307 -2.629 1.00 0.00	N		
11	ATOM	11 C	MET A	3	20.796 17.929 -0.719 1.00 0.00	C		

8. Separate into columns by using Data>Text to column. Delete columns A, E, J and K, so columns look like this:

	A	B	C	D	E	F	G	H	I
1	1	N	GLU	2	18.297	13.73	-3.131	N	
2	2	CA	GLU	2	18.993	14.932	-3.683	C	
3	3	C	GLU	2	19.498	15.737	-2.493	C	
4	4	O	GLU	2	18.811	15.848	-1.478	O	
5	5	CB	GLU	2	18.059	15.777	-4.551	C	
6	6	CG	GLU	2	18.684	17.014	-5.184	C	
7	7	CD	GLU	2	18.706	18.229	-4.27	C	
8	8	OE1	GLU	2	17.685	18.544	-3.615	O	
9	9	OE2	GLU	2	19.762	18.892	-4.189	O	
10	10	N	MET	3	20.691	16.307	-2.629	N	
11	11	C	MET	3	20.796	17.929	-0.719	C	

9.

10. Copy areas_results.txt into the last column I.

	A	B	C	D	E	F	G	H	I
1	1	N	GLU	2	18.297	13.73	-3.131	N	26.4822578
2	2	CA	GLU	2	18.993	14.932	-3.683	C	9.8896904
3	3	C	GLU	2	19.498	15.737	-2.493	C	0.75347936
4	4	O	GLU	2	18.811	15.848	-1.478	O	19.0069008
5	5	CB	GLU	2	18.059	15.777	-4.551	C	10.318119
6	6	CG	GLU	2	18.684	17.014	-5.184	C	61.3056145
7	7	CD	GLU	2	18.706	18.229	-4.27	C	39.7711067
8	8	OE1	GLU	2	17.685	18.544	-3.615	O	0.7203052
9	9	OE2	GLU	2	19.762	18.892	-4.189	O	11.6837816
10	10	N	MET	3	20.691	16.307	-2.629	N	2.84267211
11	11	C	MET	3	20.796	17.929	-0.719	C	4.07722855

11. Name tab as ABCD and save spreadsheet as “Proteins.xlsx”. The protein is ready to execute in Hi-patch.