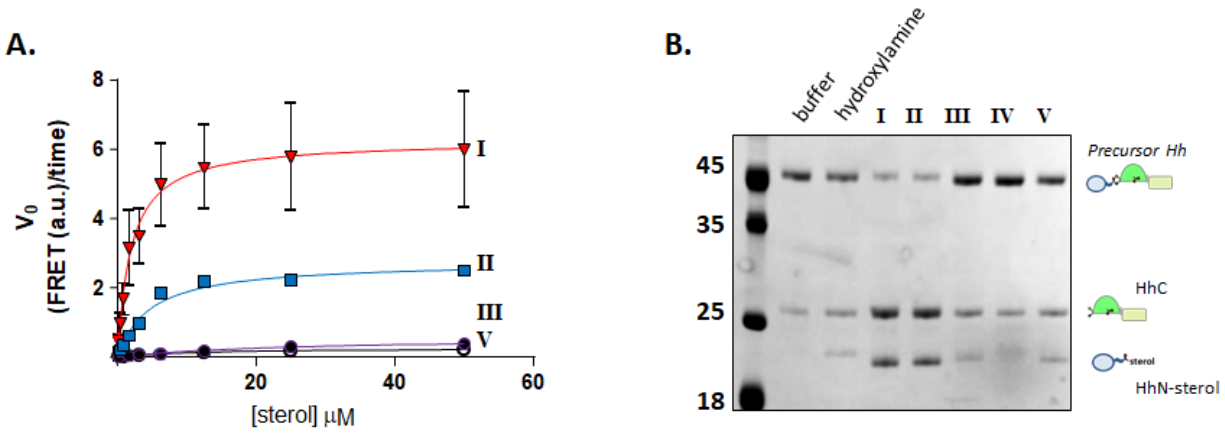
SUPPORTING TABLE 1

Substrate	$[M+H]^+$	Assignment	Experimental MW	Theoretical MW	Error (Da)
no sterol	20009.780	**HhN**	20008.77	20008.54	0.23
I	20365.017	**HhN-I**	20364.01	20364.41	-0.40
II	20371.066	**HhN-II**	20370.06	20366.43	3.63
III	20435.101	**HhN-III**	20434.09	20434.93	-0.84
IV	not observed	**HhN-IV**	NA	20434.93	NA
V	20369.962	**HhN-V**	20368.95	20366.40	2.55

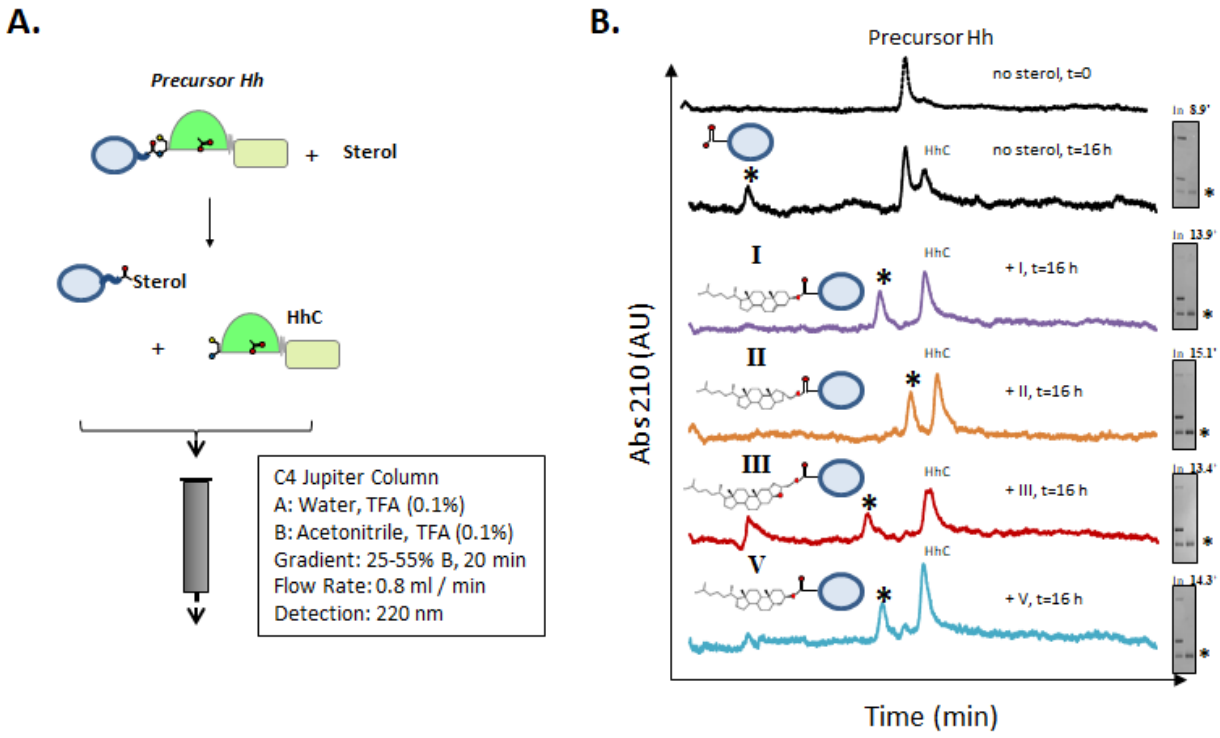
Table 1. MALDI-TOF mass spectrometry identifies hedgehog signaling protein (HhN) sterylated with cholesterol and A-ring analogs.

SUPPORTING FIGURE 1



A) Michaelis–Menten plot. Initial velocity for reaction of C-H-Y plotted as function of increasing concentration of sterols. Solid lines represent the expected kinetic behavior using the following K_M values: **I**, 1.9 µM; **II**, 4.7 µM; **III**, 12 µM; **V**, 23 µM. **B)** Gel analysis of substrate activity with chimeric hedgehog precursor, SHhN–DHhC. Processing of precursor plus/minus indicated sterols (50 µM). MW: SHhN–DHhC, 46 kDa; SHhN, 20 kDa; DHhC, 26 kDa.

SUPPORTING FIGURE 2



A) Schematic for the preparation of HhN–sterol product mixtures for analysis by RP-HPLC. **B)** RP-HPLC elution profiles of product mixtures following incubation of hedgehog precursor, SHhN-DHhC, without and with added sterol. Extended retention time observed for SHhN in the lower 4 traces is consistent with sterylation. *Right,* Fractions were collected at the times indicated by (*), evaporated to dryness, brought up in SDS-PAGE load buffer and separated by SDS-PAGE; the first lane in each of these gel images shows product mixture before RP-HPLC.

SUPPORTING METHODS

Chemicals: Reagents for synthesis were obtained through VWR. We purchased NMR solvents from Cambridge Isotopes and HPLC solvents from Pharmco-Aaper. Other reagents obtained commercially included: BIS–TRIS propane, NaCl, ethylenediaminetetraacetic acid (EDTA), imidazole, glycerol, sodium phosphate, and Lennox Broth (LB) Mix (Fisher Scientific); cholesterol, glucose, Triton X-100 (Sigma); cholestanone was purchased from Steraloids; tris(2-carboxyethyl)phosphine (TCEP; Hampton Research); Fos-choline-12 (Anatrace).

Chemical Synthesis: *A*-nor (**II**) cholestanol was prepared in two steps. Cholestanone was first oxidized to the *A*-nor acid to give an inseparable 9:1 mixture of the α and β isomers, respectively, according to a literature procedure.[1] ^{1}H NMR (major isomer, $CDCl_3$, 600 MHz): 2.921 (1H, dt, 2.8,9.0 Hz), 0.903 (3H, d 6.5 Hz), 0.866 (3H, d 6.7 Hz), 0.861 (3H, d 6.6 Hz), 0.708 (3H, s), 0.657 (3H, s). Reduction with $LiAlH_4$ gave the corresponding 2-α *A*-nor alcohol.[2] The alcohol (**II**) was also formed as an inseparable 9:1 mixture. A pure sample was obtained by Steglich esterification of the *A*-nor acid with (R)-1-phenylethanol to give the 1-phenethyl ester. ^{1}H NMR ($CDCl_3$, 600 MHz): 7.344 ppm (2H, d 1.4 Hz), 7.337 ppm (2H, s), 7.280 ppm (1H, m), 5.865 ppm (1H, q 6.6 Hz), 2.911 (1H, ddt 2.9, 10.9, 9.0 Hz), 1.522 (3H, d 6.6 Hz), 0.898 (3H, d 6.6 Hz), 0.864 (3H, d 6.6 Hz), 0.860 (3H, d 6.6 Hz), 0.694 (3H, s), 0.649 (3H, s). Purification of this ester by repeated preparative TLC (Hex/EtOAc 39:1), and $LiAlH_4$ reduction gave **II**. ^{1}H NMR ($CDCl_3$, 600 MHz): 3.503 (1H, AB, dd, 6.8,10.3 Hz), 3.483 (1H, AB, dd, 7.3,10.3 Hz), 2.256 (1H, m), 1.952 (1H, dt, 3.4,12.6 Hz), 1.804 (1H, ddt, 6.0,13.4,9.4 Hz), 0.903 (3H, d 6.6 Hz), 0.865 (3H, d 6.6 Hz), 0.860 (3H, d 6.6 Hz), 0.714 (3H, s), 0.655 (3H, s).

Protein Expression and Purification: The FRET reporter protein, **C-H-Y**, was expressed with a C-terminal His_6 tag in *Escherichia coli* LMG194 cells, as described before [3, 4]. Cultures were grown in Luria broth (100 µg/ml chloramphenicol) at 37°C until OD_{600} reached 0.5 at which point the temperature was reduced to 16°C and L-arabinose was added to (0.2 %). After orbital shaking for 16 h, cells were pelleted and lysed using chilled phosphate buffer ($2.0x10^{-1}$ M, pH 7.4) containing Triton X-100 (0.5%), glycerol (10%), potassium chloride ($1.0x10^{-1}$ M), sodium chloride ($4.0x10^{-1}$ M), lysozyme (10 µg/ml), DNase (2 µg/ml) and imidazole ($1.0x10^{-2}$ M). Soluble C-H-Y protein was purified using NiNTA affinity chromatography following manufacturer's protocol (Thermo Scientific). Column elution used chilled phosphate buffer (pH 7.4) containing glycerol (10%), sodium chloride ($5.0x10^{-1}$ M), imidazole ($5.0x10^{-1}$ M). Protein containing fractions, ranging from 8 to 15 µM in C-H-Y, were pooled, mixed with TCEP ($5.0x10^{-3}$ M, final) and stored at -80°C. **SHhN-DHhC**, chimeric hedgehog precursor protein, was expressed from a pET-22b plasmid in BL21-DE3 cells, purified by NiNTA chromatography, and stored in elution buffer with added TCEP as described [5].

Kinetics: Steady state kinetic experiments with C-H-Y were performed in 96-well black NBS-plates (Corning) using a Biotek plate reader with Gen5 software [4]. Reactions were initiated by adding sterol from an ethanol stock ($2.5x10^{-3}$ M) to wells containing C-H-Y ($1.0x10^{-7}$ M) in BIS-TRIS propane buffer (pH

7.1, $1.0x10^{-1}$ M) containing sodium chloride ($1.0x10^{-1}$ M), EDTA ($5.0x10^{-3}$ M), TCEP ($5.0x10^{-3}$ M), and detergent Fos-Choline-12 ($1.5x10^{-3}$ M). Reaction progress was monitored at 30-60 second intervals by changes in FRET, calculated by exciting samples at 400 nm, then taking the ratio of fluorescence emission at 540 nm over the emission at 460 nm. Kinetic parameters, K_m and k_{max} were calculated separately. K_m was derived from Michaelis Menten graphs of initial rate of FRET loss (V_o) plotted as a function of increasing sterol concentration, using the equation $V_O/V_{max}=[sterol]/([sterol]+K_m)$. Residual errors were minimized in Excel using the Solver function. Values for the first order rate of sterolysis, k_{max}, were determined by fitting the full kinetic trace from reactions at the highest concentration of sterol to the equation, $FRET=A*e^{-kmax*t}+C$

End point assays of autoprocessing using SHhN-DHhC, where the N-terminal domain (SHhN) is the human sonic hedgehog signaling domain and the C-terminal domain comes from *Drosophila* Hh protein, were resolved by reducing SDS-PAGE. DHhC is identical to the autoprocessing domain in C-H-Y used above. Extent of reaction was visualized with BioRad EZ plate reader after separation by SDS-PAGE and staining with Coomassie blue dye.

Identification of HhN-sterol conjugates by MALDI-TOF: In a 1:1 ratio, sterylated protein samples (~40 ng/µL) were mixed with 95% acetonitrile (HPLC grade, Fisher scientific), 0.5 % trifluoroacetic acid (LC-MS grade, Fisher Scientific) solution. Subsequently, 1 µL of sample solutions and 1 µL of saturated sinapinic acid (SA, Sigma-Aldrich) matrix solution were mixed well and applied to a plate spot. Mass spectra were acquired using a Bruker Daltonics AutoFlex Speed MALDI-TOF instrument under linear mode. The mass range was 17,400-30,000 m/z, laser power was 92%, and detector gain was 12.6 ×. MS data were processed with Bruker flexAnalysis 3.4 software.

Sterylation Assay by Reverse Phase HPLC: RP-HPLC analysis of the putative sterylated protein, SHhN-sterol, from sterolysis of SHhN-DHhC, was carried out as described by Bordeau et al [5]. Reactions included SHhN-DHhC protein ($1.0x10^{-6}$ M) in BIS-TRIS propane buffer ($1.0x10^{-1}$ M, pH 7.1) containing sodium chloride ($1.0x10^{-1}$ M), EDTA ($5.0x10^{-3}$ M), Fos-Choline-12 ($1.5x10^{-3}$ M), and TCEP ($5.0x10^{-3}$ M). Sterol was then added from an ethanol stock to final concentration $5.0x10^{-5}$ M. Following overnight incubation at 16°C, reaction mixtures (100 µL) were injected over a Phenomenex Jupiter 5 µm C4 300 Å column; flow rate of 0.8 mL/min. Sterylated HhN protein was eluted using a gradient of water/TFA to acetonitrile/TFA over 20 minutes.

Computational Modeling the HhC structure bound to cholesterol: The RCSB Protein DataBank (PDB) was mined for structures containing cholesterol using the advanced search ligand option using the precise SMILES description for cholesterol, which yielded 48 structures at the time of the search. These were examined manually to see whether cholesterol was binding internally within a monomer protein, which shortened the list to 7 structures (PDB IDs: 1LRI, 1N83, 1ZHY, 3GKI, 3N9Y, 4BOE, 4BQU). Of these, the 1.45 Å resolution structure of the cryptogein-cholesterol complex (PDB ID: 1LRI [6]) was chosen as its size was similar to the hedgehog cholesterol binding C-terminal domain (CTD), and its sequence could be aligned to obtain 15 % identity and 31 % similarity with the SRR sequence. This pairwise sequence alignment was performed using the program EMBOSS Needle [7] with the following parameter choices:

'gapopen' set to 10, 'gapextend' set to 10, 'endweight' set to false, 'endopen' set to 1, and 'endextend' set to 1. This alignment and the structure for the cryptogein-cholesterol complex was used as input for the program MODELLER[21] to generate a structure for the cholesterol-bound SRR. The refinement options for MODELLER were reduced using 'a.max var iterations' set to 1 and 'a.md level' set to 'refine.fast' so that the cholesterol binding site did not collapse during optimization in the absence of cholesterol. The best model chosen using the DOPE [8] and GA341 [9] energy functions was used for further modeling with cholesterol inserted appropriately into its binding pocket.

A continuous model of the HINT domain (PDB ID: 1AT0, [6]) and the SRR was generated manually by orienting the overlapping region present in both structures (residues 397-402) using VMD [10], and then minimizing the structure using 1000 steps of Steepest Descent (SD) minimization to a tolerance level of 0.001 kcal/mol with the CHARMM36 protein [11] and cholesterol [12] force fields and the CHARMM program [13]. This generated a structure where the SRR was connected to the HINT domain, but the O3 atom of the bound cholesterol was about 17 Å away from the first residue of the HINT domain (Cys258). To obtain a structure more suitable for catalysis of sterylation, the cholesterol was brought closer to Cys258 using an optimization protocol using restrained minimization and higher temperature Langevin molecular dynamics (MD) simulations.

This optimization protocol used mass-weighted harmonic restraints with a force constant 1.0 kcal/(mol. $Å^2$) to maintain the structure of HINT domain non-hydrogen atoms (for residues 258-396). An internal Root Mean Square Deviation (RMSD) restraint with a force constant 2000 kcal/(mol.$Å^2$) was used to maintain the internal structure of the SRR (for residues 410-471), and additional harmonic distance restraints, each with a force constant 1000 kcal/(mol. $Å^2$), were used to maintain contacts between cholesterol and the SRR. These were between Val403 CG2 and C18 (3.3 Å), Leu409 CD1 and C4 (3.7 Å), Arg418 NH1 and O3 (2.8 Å), Arg418 NH2 and C3 (4.0 Å), Trp426 CH2 and C19 (3.7 Å), Val439 CG1 and C17 (3.8 Å), Val439 CG2 and C22 (3.8 Å), Ser441 CB and C22 (3.9 Å), and Ile449 CG2 and C11 (4.1 Å). The movement of the cholesterol towards Cys258 was enforced using a harmonic distance restraint between the Cys258 SG and the cholesterol O3 atoms with a force constant of 1000 kcal/(mol.$Å^2$) whose minimum was gradually reduced in 0.1 A° decrements from 17 Å to 5 Å. At each decrement, the structure was optimized using 1000 steps of SD minimization to a tolerance level of 0.001 kcal/mol, 1000 timesteps (1 fs) of Langevin MD with a friction coefficient of 60 ps^{-1} and a temperature of 500 K, followed by another 1000 steps of SD minimization to a tolerance level of 0.001 kcal/mol.

This optimization still left the cholesterol blocked from the Cys258 residue by protein residues, suggesting that the relative orientation of the HINT and the SRR domains was still not optimal. It seemed that an internal change in a loop region towards the C-terminal end of the HINT domain (residues 400-410) could be used to gently change this relative orientation. The structure of this loop region was altered using two restrained MD simulations bringing the Ile404 C-atom up to 10 Å from the Pro416 C-atom, and then bringing the Glu388 C-atom up to 5 Å away from the Val403 C-atom, in 0.5 Å decrements, with the same minimization and Langevin dynamics protocol as above at each decrement. This was done in the absence of a restraint between the cholesterol and Cys258, and reduced the direct presence of protein residues in the path between the cholesterol O3 atom and the Cys258 SG atom, but increased the distance between those two atoms back to 17 Å. A second optimization, similar to the first

described above, was performed to move the cholesterol towards Cys258 to a distance of 5 Å between the Cys258 SG and the cholesterol O3 atoms in 0.5 Å decrements.

To optimize the manually adjusted loop conformation, a very high temperature (900 K) Langevin dynamics MD simulation for 10000 steps was performed with the residues 391-410 left unrestrained, except for distance restraints to maintain existing contacts with cholesterol. This was followed by five lower temperature (300 K) Langevin dynamics MD simulations, each for 10000 steps, interspersed by 1000 steps of SD minimization, all with no internal structural restraints on the SRR domain. In these simulations at both temperatures, mass-weighted harmonic restraints were still imposed on the HINT domain residues 258-397, on the cholesterol-SRR contacts, and on the distance between the cholesterol O3 atom and Cys258 SG atom to be 4 Å from each other. Finally, five additional 300 K Langevin dynamics MD simulations, each for 10000 steps, interspersed by 1000 steps of SD minimization, were performed with the same restraints, except with the cholesterol O3 and Cys258 SG distance restraint minimum set to 3 Å. These simulations allowed the internal structure of the SRR domain to relax in its reoriented position, resulting in a non-hydrogen atom RMSD of 3.3 Å from the original SRR homology model. To revert to a structure close to the original internal SRR structure containing a larger proportion of helical secondary structure, a series of 1000 step Langevin MD simulations were performed, with interspersed 1000 steps of SD minimization, with the HINT domain restraints, cholesterol-SRR contact restraints, and the cholesterol O3 to Cys258 SG distance restraint maintained, and the RMSD restraints on the internal structure of the SRR domain gradually reducing in 0.3 Å decrements from 3.3 Å to 0.0 Å. The final structure was minimized for 1000 SD steps with restraints mentioned above, and then for 1000 steps without any restraints to obtain an initial HhC model with bound cholesterol.

This model placed *Drosophila melanogaster* SRR residues Y450 and W451, which are highly conserved and possibly have mechanistic roles, more than 10 Å from the bound cholesterol. For the final model, these residues were brought closer to the cholesterol using consecutive 300 K Langevin dynamics MD simulations, each for 1000 steps, interspersed by 1000 steps of SD minimization. These simulations were performed using the previously mentioned mass-weighted harmonic restraints on the HINT domain residues 258-397 and on the distance between the cholesterol O3 atom and Cys258 SG atom to be 4 Å. In addition, the C14-CD2 atomic distance between cholesterol and the W451 residue and the C7-CG atomic distance between cholesterol and the Y452 residue in SRR were gradually reduced in 1 Å decrements to a final distance of 3 Å.

Computational Docking of sterols into the HhC homology model: The initial geometries for the sterol analogs were obtained from the PubChem Database [14], and quantum mechanical (QM) geometry optimization was carried out using Gaussian 09 [13] in the gas phase using the B3LYP functional [15 16] and the def2SVP basis set [17]. Autodock4 tools [18] were used for generating Pdbqt files from these QM-optimized sterol structures and the HhC protein model without its bound cholesterol.

Autodock Vina [19] was used for sterol docking into HhC. For both protein and energy-reduced sterol files, polar hydrogens were maintained while nonpolar hydrogens were merged. Partial charges on the protein model were added using Kollman Charges, and the same was done for the ligand file using Gasteiger-Marsili Charges. The protein model was set to be rigid, while the ligand's number of rotatable

bonds was specific to each sterol structure. The grid dimensions for docking were designed to encapsulate the entire protein. From multiple binding poses predicted in each docking run, the pose with the lowest binding energy score was picked for further analysis.

REFERENCES

1. P. Balakrishnan and S. C. Bhattacharyya, *Indian J. Chem., Sect B*, 1986, 25B, 1050-1051.
2. H. M. Hellman and R. A. Jerussi, *Tetrahedron*, 1964, 20, 741-745.
3. T. S. Owen, X. J. Xie, B. Laraway, G. Ngoje, C. Wang and B. P. Callahan, *Chembiochem*, 2015, 16, 55-58.
4. T. S. Owen, G. Ngoje, T. J. Lageman, B. M. Bordeau, M. Belfort and B. P. Callahan, *Anal. Biochem.*, 2015, 488, 1-5.
5. B. M. Bordeau, D. A. Ciulla and B. P. Callahan, *ChemMedChem*, 2016, 11, 1983-1986.
6. M. B. Lascombe, M. Ponchet, P. Venard, M. L. Milat, J. P. Blein and T. Prange, *Acta Crystallogr. D Biol. Crystallogr.*, 2002, 58, 1442-1447.
7. P. Rice, I. Longden and A. Bleasby, *Trends Genet.*, 2000, 16, 276-277.
8. M. Y. Shen and A. Sali, *Protein Sci.*, 2006, 15, 2507-2524.
9. A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, 234, 779-815.
10. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, 14, 33-38, 27-38.
11. R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. Mackerell, Jr., *J. Chem. Theory Comput.*, 2012, 8, 3257-3273.
12. J. B. Lim, B. Rogaski and J. B. Klauda, *J. Phys. Chem. B*, 2012, 116, 203-210.
13. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. A. , J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and F. D. J., Gaussian, Inc., , Wallingford CT, , 2009.
14. S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang and S. H. Bryant, *Nuc. Acids Res.*, 2016, 44, D1202-1213.
15. A. D. Becke, *J. Chem. Phys.*, 1993, 98, 5648-5652.
16. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.,* 1994, 98 11623-11627.
17. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, 7, 3297-3305.
18. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comp. Chem.*, 2009, 30, 2785-2791.
19. O. Trott and A. J. Olson, *J. Comp. Chem.*, 2010, 31, 455-461.

20. J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkins-Diehr, *Computing in Science & Engineering*, 2014, 16, 62-74.

21. B. Webb and A. Sali, *Curr. Protoc. Bioinformatics*, 2014, 47, 5 6 1-5 6 32.