Electronic Supplementary Material (ESI) for Chemical Communications

Supporting information for: Envisioning an enzymatic Diels-Alder reaction by *in-situ* acid-base catalyzed diene generation

Mats Linder, Adam Johannes Johansson, Bianca Manta, Philip Olsson, and Tore Brinck*

Applied Physical Chemistry, KTH Royal Institute of Technology, Teknikringen 30, S-100 44 Stockholm

E-mail: tore@physchem.kth.se

1 Computational details

1.1 Additional references for the quantum chemical calculations.

This section provides additional methodological references for the basis sets and continuum models employed in the calculations.

- Program: All nonenzymatic optimizations were performed using Gaussian09.[1] Optimizations and large-basis single point calculations on the KSI cluster model (see below) were performed using Jaguar 7.6[2] for increased speed.
- Basis sets: Pople style 6-31+G(d) basis set for optimizations and frequency analysis, 6-311++G(2d,2p) for single point calculations.[3]
- Solvent model: The SMD-PCM model, as implemented in Gaussian09.[4]
- Graphics: All molecular graphics was created using USCF Chimera.[5]

1.2 KSI cluster model

The *Pseudomonas testosteroni* KSI crystal structure was obtained from the Protein Data Bank,[6] entry 1QJG.[7] The 137-atom cluster was generated by extracting the catalytically important amino acids Tyr14, Asp38, Tyr55 and Asp99, as well as neighboring residues constituting the cavity in the vicinity of the catalytic functionalities. The amino acids were truncated so that only the the atoms close to the cavity were maintained. This meant that most of the backbone structure was removed. To conserve the active site structure, the side chain carbons closest to the removed moieties, as well as the hydrogens replacing the dangling bonds, were held fixed throughout the optimizations. Figure S1 shows the empty cluster model, with frozen carbons and hydrogens colored magenta.

1.3 Molecular dynamics simulations

As with the cluster calculations, the PBD structure 1QJG was used as a basis for the molecular dynamics (MD) simulations. Starting structures for the enzyme-ligand complexes were generated using the GOLD 5.0 docking software.[8] We used the AMBER10[9] suite with the Amber force field FF03[10] and the pmemd reimplementation of the sander engine. Force field parameters for the non-protein atoms were generated from the general Amber force field (GAFF)[11] using the Anterchamber and Parmchk modules in AmberTools. Partial charges were computed using the embedded mopac scheme at the AM1-BCC level.[12] 3 neutralizing Na^+ions and a 8 Å solvent shell (truncated octahedron) of TIP3P[13] waters were added using the LEaP program.

Simulations were performed under periodic boundary conditions, using the pmemd module in Amber, with long-range electrostatics handled by Particle Mesh Ewald (PME) summation.[14] Each system was first minimized in two steps, where the protein was held fixed during the first (1000 iterations) and unconstrained during the second (5000 iterations). In both cases, the minimizations were done using a steepest decent algorithm during the first half followed by the conjugate gradient method. The systems were then heated to 300 K during 20 ps, with mild restraints on the protein atoms (10 kcal/mol,Å), followed by 2 ns unconstrained production at 300 K and 1 bar. 2

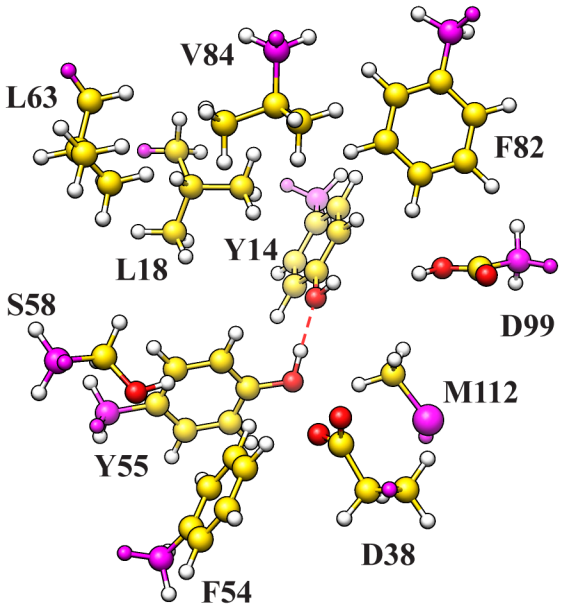


Figure S1: Ball-and-stick representation of the 137-atom active site model created from PDB entry 1QJG. Frozen atoms are colored magenta. The labels show from which amino acid each residue originates.

fs time steps were used, and consequently hydrogen movements were constrained by the SHAKE algorithm.[15]

Only the last half of each simulation was used for enzyme-ligand binding energy calculations, to ensure that the system had been sufficiently equilibrated. The ligand binding energies were calculated using the Linear Interaction Energy approach,[16,17] as seen in eq. 1.

$$\Delta G_{bind} = \alpha[\langle V_{l-s}^{vdw}\rangle_{bound} - \langle V_{l-s}^{vdw}\rangle_{free}] + \beta[\langle V_{l-s}^{elec}\rangle_{bound} - \langle V_{l-s}^{elec}\rangle_{free}] + \gamma. \quad (1)$$

The ensembles were represented by 300 and 500 snapshots from the enzyme-ligand and free ligand systems, respectively. α and β were set to 0.18 and 0.43, respectively.[17] A solvent accessible surface area (SASA) correction was chosen as the (otherwise empirical) γ correction.[18] We used the Generalized-Born surface area term from the Amber MMPBSA script, that is $\gamma = \Delta G_{SA}^{gb}$. While no rigorous fitting has been performed for the augmentation at this point, the good correlation with experimental values for the KSI substrate 5-Androstene-3,17-dione **5** served as a benchmark.

As noted in the main text, binding of the small molecule **2** was weak and nonspecific. We have attempted some rational mutagenesis to improve the association of both one and two substrates, however this work is still in its cradle will be the subject of future work.

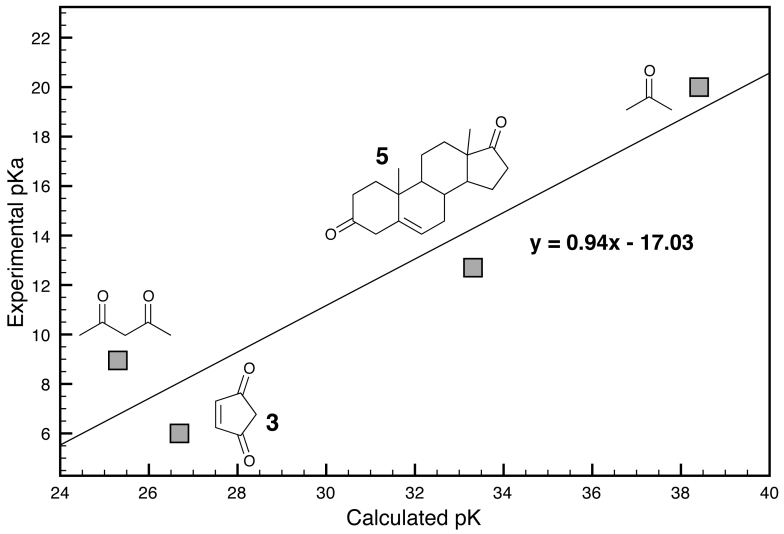
2 pK_a estimations

The estimated pK_a values in Scheme 1 of the main article were done by an SMD-PCM M06-2X/6-311++G(2d,2p)//6-31+G(d) evaluation of the reaction

$$\text{ketone} + H_2O \rightleftharpoons \text{enolate} + H_3O^+. \quad (2)$$

The resulting equilibrium constants were then fitted to the experimental pK_a of acetone (20), acetylacetone (9), cyclopentene-1,3-dione **3** (6)[19] and 5-Androstene-3,17-dione **5** (12.7).[20] Table 1 shows the regression results and following pK_a estimations as presented in Scheme 1 of the main text. Note that the sources for the given experimental pK_a values are diverse and of limited accuracy.

Table S1: pK_a regression and estimation[a]



Compound	**#**	**Calc.** pK (pK_a)[a]	**ReCalc.** pK_a [b]
Acetone		38.4 (20)	19.1
Acetylacetone		25.3 (9)	6.8
Cyclopenten-1,3-dione	**3**	26.7 (6)	8.1
5-Androstene-3,17-dione	**5**	33.3 (12.7)	14.0
Cyclopent-2-enone	**2**	39.0	19.7
Methylvinylketone	**4**	38.6	19.3
Anthrone	**1**	27.4	8.7

[a] The linear regression was based on the first four entries in the table, for which experimental data were available. Experimental pK_a :s are given in parentheses in the table. [b] Using the displayed regression formula, where x is the calculated pK value for eq. 2.

We note that the viable pro-dienes have relatively high pK_a values compared to *e.g.* anthrone. However, an α,β-unsaturated ketone is required to be able to reprotonate the enolic product and complete the catalytic cycle. This is not possible in any of the mentioned compounds except **2**, **3** and **4**. **4** may seem like an ideal choice, but its documented reactivity towards nucleophiles and tendency to base-catalyzed polymerization[19,21–23] made us focus on the less reactive **3**.

3 IRC trajectory and Michael pathway

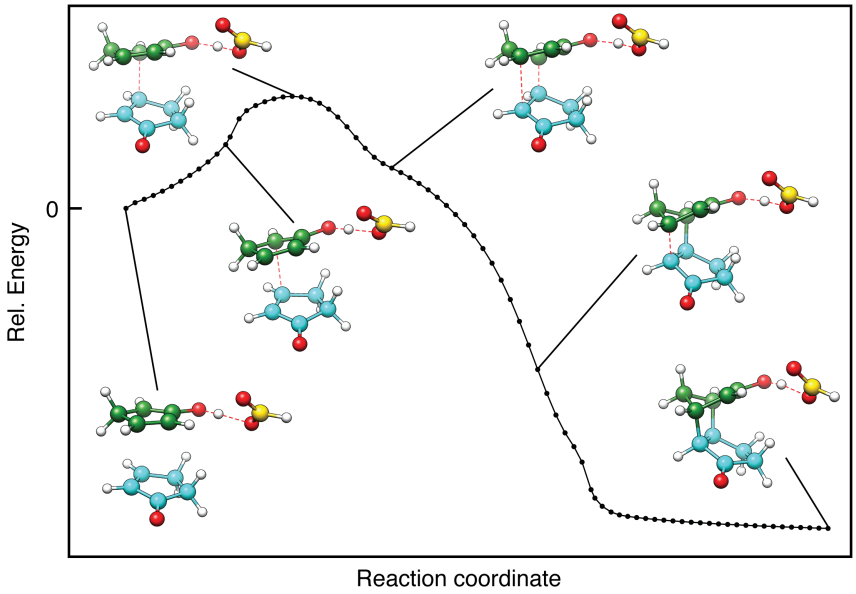


Figure S2: IRC trajectory for starting from **TS-DA** of **2+2'·AH**. There is a ridge-like region on the curve where the second bond begins to form (4^{th} depicted point from the left). Note also how the acidic proton is transfered from the diene to the formate and back again during the reaction.

We have previously seen that activated Diels-Alder reactants can prefer a stepwise pathway over the classic, concerted one, especially if the TS involves large charge transfers.[24]The optimized TS geometries were highly asynchronous, although the extent varied with the computational method (see below). It was therefore interesting to investigate i) the reaction coordinate following an optimized TS and ii) the energy profile of a tentative Michael pathway,[24] leading to either the Diels-Alder adduct *via* a second TS or a Michael adduct *via* reprotonation of the dienophile α-carbon.

An intrinsic reaction coordinate (IRC)[25] calculation was therefore performed at the MP2/6-31+G(d) level. We started from the **2+2'·AH TS-DA** and obtained the energy profile depicted in Figure S2. It confirms that at the MP2 level, the reaction is concerted.

Results for the Michael-type pathway of **6+2'·AH** is tabulated in the main text, and Figure S3 gives a graphical depiction as well as the results for the **2+2'·AH** system. We were not able to find **M-INT**, **TS-MDA** using the MP2 method. In comparing Figure S2 with Figure 1 in the main article, one can note that **TS-M** is much closer to **TS-DA** for the **2+2'·AH** system than when acrolein is the dienophile. This is possibly due to a larger amount of strain in **TS-DA** with **2**, which raises the barrier with respect to **TS-M**.

4 A comment on the choice of quantum chemical methods.

Several studies have recently emerged that show the inadequacy of employing the ubiquitous B3LYP[26] hybrid functional for accurate treatment of dispersion[27,28] and $\pi \rightarrow \sigma$ transformations.[29]

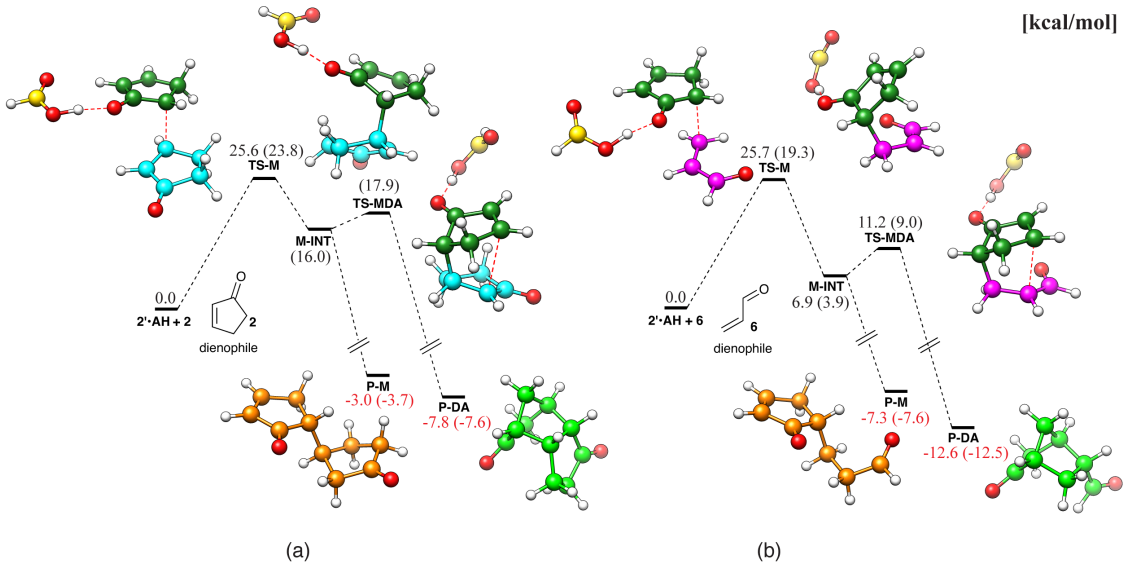


Figure S3: Michael type addition (**TS-M**) of **2'·AH** to (a) **2** and **6**, followed by the bifurcation from **M-INT** into the Michael or Diels-Alder product (**P-M** or **P-DA**). The latter reaction proceeds *via* the second barrier **TS-MDA**. SMD-PCM 6-311++G(2p,2d)/6-31+G(d) free energies are given at the SCS-MP2 level, with M06-2X values in parentheses. Energies in red are relative to the separated keto forms of the reactants. Note that **M-INT** and **TS-MDA** in (a) could not be found at the MP2 level; M06-2X geometries are shown instead.

Since these components are central in governing pericyclic reactions such as Diels-Alder, we wanted to use alternative methods in this study. One choice could be Grimme's DFT-D scheme,[27] which however does not specify a given choice of functional. Another problem with functionals such as B3LYP is that errors increase in size with larger basis sets, mainly due to a reduction in the cancellation of errors.[29–31] Since we have previously employed the M06-2X functional[32,33] and SCS-MP2[30,34,35] method on a related system,[24] obtaining good correlation with experimental data, we decided to use these methods once more.

The M06-2X functional is not yet extensively benchmarked, although the number of studies employing it is rapidly increasing.[36] However, B3LYP is still the abundant choice in geometry optimizations. This has recently been motivated by a thorough study by Simón and Goodman.[37] M06-2X has so far been used for geometry optimizations in only a limited number of cycloaddition studies.[38] In a forthcoming study, we compare B3LYP and M06-2X geometries and conclude that for species with a large extent of charge separation, the results diverged significantly.[39] Due to our previous results,[24] we decided to rely more on the M06-2X geometries in this case, and we follow the same reasoning for employing M06-2X here.

In the present study we also used MP2 for geometry optimizations of the small systems considered initially. This allows us to not only compare energies but also geometries between the different methods. As seen in Figure S4, the differences in **TS-DA** asynchronicity between B3LYP, M06-2X and MP2 are quite dramatic. The B3LYP geometries are virtually conjugate addition TSs, while the MP2 TSs are much more synchronous. The M06-2X geometries place somewhere in between. The extreme B3LYP TS geometries was a strong reason for discarding the functional in this study. One could argue that since uncorrected MP2 generally leads to underestimated cycloaddition bar-

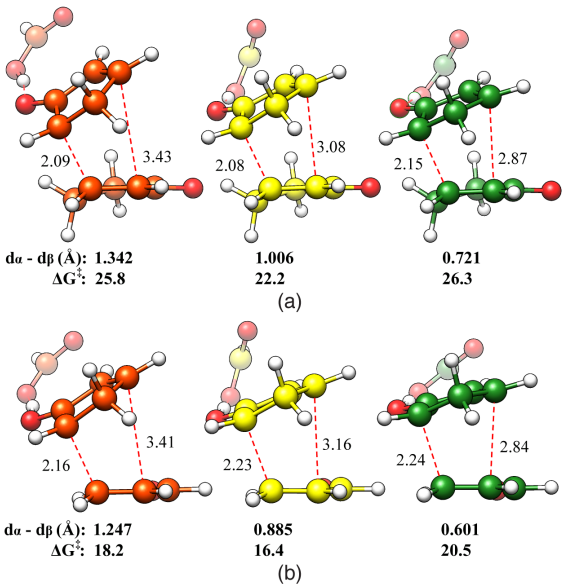


Figure S4: Comparison of the optimized **TS-DA** geometry of (a) **2'·AH+2** and (b) **2'·AH+6**, using B3LYP (orange), M06-2X (yellow) and MP2 (green). All structures are optimized at the gas-phase 6-31+G(d) level. The asynchroniciy distance is defined as $d_\alpha - d_\beta$,[40] where the greek letter refer to the dienophile carbon. For each geometry, $\Delta G^\ddagger$ is reported at the SMD-PCM–M062X/6-311++G(2d,2p) level.

riers,[30] the MP2 geometry might be "too" synchronous, but this does of course not warrant that M06-2X should yield a correct TS geometry. A cross-comparison of the TS free energies at the SMD-PCM–M06-2X/6-311++(2d,2p) level is included in Figure S4, illustrating that the high-level energy varies with up to 4 kcal/mol depending on optimization method.

In conclusion, we feel confident that the choice of methods in this study is reasonable, because i) B3LYP energies overestimate cycloaddition barriers, and ii) B3LYP gave extremely asynchronous TS geometries.

5 Molecular coordinates

`.xyz`-files of the relevant optimized M06-2X and MP2 geometries of the nonenzymatic reactions are provided as two separate `.zip` archives; `m062x.zip` and `mp2.zip`. The optimized KSI cluster geometries are given in the `ksi_model.zip` archive.

Notes and References

1 Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Jr.,; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Strat-

mann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; ; Fox, D. J. *Gaussian 09, Revision A.02*, 2009.

2 *Jaguar v. 7.6*, 2009, www.schrodinger.com.

3 Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728.

4 (a) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001; (b) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681; (c) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

5 (a) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. *J. Comput. Chem.* **2004**, *25*, 1605–1612; (b) *UCSF Chimera*, 2004, http://www.cgl.ucsf.edu/chimera, accessed 07-14, 2011.

6 Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

7 Cho, H.-S.; Ha, N.-C.; Choi, G.; Kim, H.-J.; Lee, D.; Oh, K. S.; Kim, K. S.; Lee, W.; Choi, K. Y.; Oh, B.-H. *J. Biol. Chem.* **1999**, *274*, 32863–32868.

8 (a) *GOLD 5.0*, 2010, http://www.ccdc.cam.ac.uk/-products/life_sciences/gold/, accessed 07-14, 2011; (b) Jones, G.; Willet, P.; Glen, R. *J. Mol. Biol.* **1995**, *245*, 43–53; (c) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748; (d) Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. *Proteins* **2002**, *49*, 457–471; (e) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609–623.

9 Case, D.; Darden, T.; III, T. C.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K.; B.Wang,; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.; Seetin, M.; Sagui, C.; Babin, V.; Kollman, P. *AMBER 10*, 2008.

10 Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

11 Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

12 (a) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132–146; (b) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

13 Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926–935.

14 (a) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10093; (b) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pederson, L. *J. Chem. Phys.* **1995**, *103*, 8577–8592.

15 Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

16 (a) Åqvist, J.; Medina, C.; Samuelsson, J. *Protein Eng.* **1994**, *7*, 385–391; (b) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free Energy Calculations and Ligand Binding. In *Protein Simulations*; Daggett, V., Ed.; Academic Press, 2003; Vol. 66, pp 123–158; (c) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. *Acc. Chem. Res.* **2002**, *35*, 358–365.

17 Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput.-Aided. Mol. Design* **1998**, *12*, 27–35.

18 Linder, M.; Johansson, A. J.; Olsson, T. S. G.; Liebeschuetz, J.; Brinck, T. *J. Chem. Inf. Model.* **2011**, *51*, 1906–1917.

19 DePuy, C. H.; Lyons, C. E. *J. Am. Chem. Soc.* **1960**, *82*, 631–633.

20 Pollack, R. M.; Zeng, B.; Mack, J. P. G.; Eldin, S. *J. Am. Chem. Soc.* **1989**, *111*, 6419–6423.

21 DePuy, C. H.; Zaweski, E. F. *J. Am. Chem. Soc.* **1957**, *79*, 3923–3924.

22 DePuy, C. H.; Zaweski, E. F. *J. Am. Chem. Soc.* **1959**, *81*, 4920–4924.

23 DePuy, C. H.; Wells, P. R. *J. Am. Chem. Soc.* **1960**, *82*, 2909–2913.

24 Linder, M.; Johansson, A. J.; Brinck, T. *Org. Lett.* **2012**, *14*, 118–121.

25 (a) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154–2161; (b) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523–5527.

26 (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100; (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377; (c) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

27 Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

28 Grimme, S.; Steinmetz, M.; Korth, M. *J. Org. Chem.* **2007**, *72*, 2118–2126.

29 Pieniazek, S.; Clemente, F.; Houk, K. *Angew. Chem. Int. Ed.* **2008**, *47*, 7746–7746.

30 Goumans, T. P. M.; Ehlers, A. W.; Lammertsma, K.; Würthwein, E.-U.; Grimme, S. *Chem. Eur. J.* **2004**, *10*, 6468–6475.

31 Grimme, S.; Muck-Lichtenfeld, C.; Wurthwein, E.-U.; Ehlers, A.; Goumans, T.; Lammertsma, K. *J. Phys. Chem. A* **2006**, *110*, 2583–2586.

32 (a) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103; (b) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382; (c) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

33 Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

34 Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

35 Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

36 Zhao, Y.; Truhlar, D. G. *Chem. Phys. Lett.* **2011**, *502*, 1 – 13.

37 Simón, L.; Goodman, J. M. *Org. Biomol. Chem.* **2011**, *9*, 689–700.

38 (a) Gu, J.; Wang, J.; Leszczynski, J.; Xie, Y.; III, H. F. S. *Chemical Physics Letters* **2008**, *459*, 164–166; (b) Garr, A. N.; Luo, D.; Brown, N.; Cramer, C. J.; Buszek, K. R.; VanderVelde, D. *Organic Letters* **2010**, *12*, 96–99.

39 Linder, M.; Brinck, T. *Step-wise Diels-Alder – more than just an Oddity?*, 2012, Submitted for publication.

40 Kong, S.; Evanseck, J. *J. Am. Chem. Soc.* **2000**, *122*, 10418–10427.