How Many Conformations Need to Be Sampled to Obtain Converged QM/MM Energies? the Curse of Exponential Averaging

Ryde, Ulf

Published in:
Journal of Chemical Theory and Computation

DOI:
10.1021/acs.jctc.7b00826

2017

Document Version:
Early version, also known as pre-print

How many conformations need to be sampled to obtain converged QM/MM energies? The curse of exponential averaging

Ulf Ryde

Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se, Tel: +46 – 46 2224502, Fax: +46 – 46 2228648

2017-09-13

Abstract

Combined quantum mechanical and molecular mechanical (QM/MM) calculations is a popular approach to study enzymatic reactions. They are often based on a set of minimised structures obtained on snapshots from a molecular dynamics simulation to include some dynamics of the enzyme. It has been much discussed how the individual energies should be combined to obtain a final estimate of the energy, but the current consensus seems to be to use an exponential average. Then, the question is how many snapshots are needed to reach a reliable estimate of the energy. In this paper, I show that the question can be easily be answered if it is assumed that the energies follow a Gaussian distribution. Then, the outcome can be simulated based on a single parameter, σ, the standard deviation of the QM/MM energies from the various snapshots, and the number of required snapshots can be estimated once the desired accuracy and confidence of the result has been specified. Results for various parameters are presented and it is show that much more snapshots are required than is normally assumed. The number can be reduced by employing a cumulant approximation to second order. It is shown that most convergence criteria work poorly, owing to the very bad conditioning of the exponential average when σ is large (more than ~7 kJ/mol), because the energies that contribute most to the exponential average have a very low probability. On the other hand, σ serves as an excellent convergence criterion.

Key Words: QM/MM, exponential averaging, cumulant approximation, enzyme reaction.

Introduction

Combined quantum mechanical and molecular mechanics (QM/MM) is a popular method to study (among other things) enzyme reaction mechanisms.[1,2] It has the advantage of providing a detailed atomistic account of the surrounding protein in the calculations. On the other hand, it is sensitive to the conformation of the surrounding protein and solvent – the addition or removal of a single hydrogen bond far from the active site affects the energy by ~20 kJ/mol in a way that most likely is irrelevant for the reaction. A complicated molecule like a protein has in principle an infinite number of possible conformations and it is very hard to ensure that the surrounding enzyme remain in the same local minima throughout a reaction sequence. This problem can be alleviated by considering free energies, which are obtained by averaging over all relevant conformations.[1–3] However, for a high-level QM method, this is very expensive.

Therefore, most QM/MM studies are performed using minimised structures. To reduce the local-minima problem, various strategies have been used, e.g. keeping most of the surroundings fixed or running the reaction forth and back several times.[2] However, the most common approach is to repeat the calculations for a number starting structures, typically snapshots from a molecular dynamics (MD) simulation, selected either by random [4–12] or by their similarity to the transition state.[13,14]

Unfortunately, such an approach leads to several new problems. First, it must be decided how the calculated energies (or free energies, with entropies from a normal-mode analysis of harmonic frequencies) for the various snapshots (e.g. reaction or activation energies, E_i) should be combined to a final estimate. Some authors have used an arithmetic average[6,13,15–24]

$$\Delta E_{\mathrm{AA}} = \frac{1}{n}\sum_{i=1}^{n} E_i \qquad (1),$$

which gives a proper averaging of the snapshots at the MM level (if they are picked at random from the simulation). Others have instead used an exponential average[8–13,18–20,23,25–30]

$$\Delta E_{\mathrm{EA}} = -RT\,\ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\frac{-E_i}{RT}}\right) \qquad (2),$$

where n is the number of energies, T is the temperature and R is the gas constant. This can be seen as an attempt to reweight the snapshots with the QM/MM energy function (although for activation barriers, it may seem more natural to reweight with the total QM/MM energy, not the activation energy). Other methods have also been suggested, e.g. taking the minimum activation barrier.[31] Recently, Cooper and Kästner compared the results of various approaches and they showed that the exponential averages gave the most accurate results compared to explicitly calculated free energies.[13] Still, it should be remembered that all approaches are only approximations, avoiding the much more demanding free-energy calculations.

The next question to answer is how many snapshots are needed to obtain a converged value of ΔE_{EA}. Most studies have used only a few (3–10) snapshots,[30] but a few studies used 20–65 energies;[13,25,27,29,30] Table 1 shows a compilation of 24 studies. In a recent paper, Li et al. tried to answer this question by considering the convergence of the results, using 20 snapshots.[30] They also tried to design proper convergence criteria to decide whether the enough snapshots are considered. They concluded that ~20 snapshots are needed for convergence. Unfortunately, their results are strongly misleading, because they have failed to recognize how badly conditioned the exponential average is.

Similar convergence questions have recently been discussed in the related area of employing QM/MM methods to postprocess free-energy perturbation calculations performed at the MM level.[32–34] In particular, Boresch and Woodcock employed statistical probability

distributions to study the convergence.[35] In the present paper, I use a similar approach to answer how many QM/MM energies are needed to obtain a reliable estimate of ΔE_{EA} and how it can be known if the sum is converged. I compare four convergence criteria and illustrate the problem of the exponential average.

Theory and Methods

Theory

The question about the minimum value of n needed to obtain a reliable estimate of the exponential average in Eqn. 2 can easily be answered by assuming a certain distribution of the energies (E_i). Throughout this paper, I will assume that they follow a Gaussian distribution:

$$G(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

The validity of this assumption will be discussed below. A Gaussian distribution is characterised by two parameters, the expectation value (mean; μ) and the standard deviation (σ).

If the energies E_i follow a Gaussian distribution, the sum in Eqn. 2 can be turned into the integral:[35]

$$\Delta E = -RT \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{x}{RT}} \mathrm{d}x\right) \quad (4)$$

This integral has an analytical solution,

$$\Delta E = \mu - \frac{\sigma^2}{2RT} \quad (5)$$

Throughout this article, I will assume that T = 300 K and use kJ/mol as the energy unit.

Convergence criteria

I have tested four criteria previously used to decide whether the exponential sum in Eqn. 2 is converged or not. The first is Kish's effective sampling size, [36]

$$Q = \frac{(\sum_{i=1}^{n} w_i)^2}{\sum_{i=1}^{n} w_i^2} \quad (6),$$

where the w_i is the weight of each term in Eqn. 2, i.e.

$$w_i = \frac{e^{-\frac{E_i}{RT}}}{\sum_{i=1}^{n} e^{-\frac{E_i}{RT}}} \quad (7).$$

The second is the maximum value of these weights, w_{max}, [37,38] which is obtained for the smallest (most negative) energy E_{min}. The third is the weighting entropy:[39]

$$S_w = \sum_{i=1}^{n} w_i \ln w_i$$

Finally, the fourth is the disproportionation effect suggested by Li et al.,[30]

$$DE = \frac{|\Delta E_{\mathrm{EA}}(n) - \Delta E_{\mathrm{EA}}(n-1_{\mathrm{min}})|}{\Delta E_{\mathrm{EA}}(n)} \quad (8),$$

where $\Delta E_{\mathrm{EA}}(n)$ is the exponential average from Eqn. 2, based on all n energies, whereas $\Delta E_{\mathrm{EA}}(n–1_{\mathrm{min}})$ is the same estimate, based on all energies except E_{min}.

Simulation

With a Gaussian approximation, the summation in Eqn. 2 can be studied by a numerical simulation. A simple program was written that generates a certain number of Gaussian-distributed energies (by the Box–Muller transform[40]) and calculates the exponential average in Eqn. 2 and the four convergence criteria. This is repeated 1000 times and it is noted how many times this average was within a certain limit (θ) from the analytical results (Eqn. 5). The program automatically finds the minimum number of snapshots (within 0.1%) needed to fulfil these criteria. The Fortran code can be provided by the author upon request.

Result and Discussion

In this paper, I investigate how many QM/MM energies are needed to give a converged and correct result for the exponential sum in Eqn. 2. This can be easily done if it assumed that the energies follow a Gaussian distribution. Then, the correct answer is analytically known (Eqn. 5) and a simple simulation can be used to determine how many energies are needed to reach a certain accuracy (θ) and confidence. The Gaussian distribution depends on only a single parameter, the standard deviation, σ (because the mean value only translocates the distribution and therefore is arbitrary in this discussion).

The number of energies

To start with, I assume that we want to obtain an energy ΔE_{EA} that is within θ = 4 kJ/mol of the correct answer with 95% confidence, which seems to be a reasonable limit for accurate calculations. In Figure 1, the results of the simulations are shown, viz. the number of required QM/MM energies as a function of the standard deviation, σ. It can be seen that for a small standard deviation, only few energies are needed. For example, with a σ of 1 kJ/mol, a single value is enough, because the spread of the energies is so small (compared to the desired accuracy of 4 kJ/mol). However, when σ gets larger, the number increases exponentially (note the logarithmic scale in Figure 1).

Let us now re-evaluate the recent study by Li et al..[30] They studied fluoroacetate dehalogenase with two substrates, fluoroacetate and chloroacetate, and two different sizes of the QM system. From the energies in their supplementary Table 1, σ can be calculated: 10, 13, 14 and 17 kJ/mol, respectively. According to Figure 1 and Table 2, ~4200 energies are needed to get an answer that is converged to 4 kJ/mol with 95% confidence for the first case (chloroacetate with the small QM system), whereas 500 000, 4 000 000 and ~10^9 samples are needed for the other three cases. This is in sharp contrast to the conclusions in the original paper, which suggested that ~20 energies were enough. According to the simulations, 20 samples would give errors of 8, 16, 20 and 33 kJ/mol on average in the four cases. How can we reach such different conclusions? The answer is that Li et al. have failed to realise how badly conditioned the sum in Eqn. 2 is.

The integral version of the exponential sum in Eqn. 4, consists of two terms, the Gaussian

function ($G(E)$ in Eqn. 3) and the Boltzmann function ($B(E) = e^{-\frac{E}{RT}}$). They are plotted together with their product ($P(E)$) in Figure 2. The desired result is the (logarithm of the) area below the product curve, which is also shown (right axis). It can be seen in Figure 2a that G shows the well-known bell-shape, whereas B is monotonously growing as E decreases. For small values of σ, G decreases faster than B and therefore, the product has also nearly a bell shape, which is only slightly skewed towards lower values. For σ = 1, it attains its maximum at a slightly negative value (–0.4 kJ/mol) and both terms attain small values around the range where the product is significant (G = 0.4 and B = 1.2). This means that all values that significantly contribute to the product have a large probability to be observed. ΔE_{EA} shows a rapidly decreasing trend as E decreases, but the variation is small and it is converged within 4 kJ/mol already at E = –0.2 kJ/mol, near the peak of the Gaussian distribution.

However, for σ = 10 kJ/mol, the situation is different as can be seen in Figure 2b. In this figure, the curves span a much larger range of energies, so that they need to be shown on a logarithmic scale. B is still monotonically increasing, giving a straight line on the logarithmic scale. Likewise, G still shows a bell-shaped curve, but with a much larger spread (owing to the larger σ), so that a much larger range of energies is important. This means that G decreases slower and it does not dominate B until E < –74 kJ/mol. In particular, the product attains its maximum around E = –40 kJ/mol. The probability to observe this or lower energies is only $8 \cdot 10^{-5}$, i.e. you need to sample ~12000 points before such energies are observed. Moreover, many even more unlikely energies need to be sampled before you have found all energies that contribute significantly to the integral (P > 1 for –74 < E < –6, with probabilities down to $1 \cdot 10^{-14}$). Fortunately, the logarithm in Eqn. 4 makes ΔE_{EA} more stable, as is also shown in Figure 2b (right axis): ΔE_{EA} is converged within 4 kJ/mol already around E = –30 kJ/mol, in accordance with the simulated result that 4200 energies are needed. However, with only 20 energies, the observed energies are typically in the range –19 < E < 19 kJ/mol, giving a major underestimate of ΔE_{EA}. Figure 2b also explains why Li et al. reached an incorrect conclusion regarding the number of points needed. They suggested that you should sample until ΔE_{EA} estimated from the last five points does not change by more than 4 kJ/mol. However, five points is not much if the probability to find the important points is $<10^{-3}$.

Instead a numerical simulation is needed, as the one in Figure 1 and Table 2, which show how many points are needed for a certain accuracy and confidence. However, we have seen that if σ is large, the number of points rapidly becomes prohibitively large. The compilation of previous QM/MM studies in Table 2 show that such large σ values are not rare: σ ranges from 0.6 to 97 kJ/mol and 73% of the studies have at least one series of energies with σ > 10 kJ/mol. What can be done in these cases? Is this the end of QM/MM minimisation studies?

Fortunately, not necessarily. If the user is willing to accept an approximation, the situation improves significantly: If it is assumed that the distribution indeed is Gaussian, ΔE can be estimated from the analytical solution in Eqn. 3, instead of the sum in Eqn. 2. This is the cumulant approximation to the second order.[41–43] It requires only estimates of μ and σ, which are much more well-behaving, because they are based on arithmetic averages. Figure 3 and Table 2 shows the required number of energies as a function of σ with the cumulant approximation. It can be seen that for low σ, the difference gain is minimal, but for σ > 6 kJ/mol, appreciably lower values of n are required with the cumulant approximation. For example, for the four sets in the study of Li et al. (with σ = 10–17 kJ/mol), 230–1700 energies are needed for an accuracy of 4 kJ/mol with 95% confidence. However, this is still much more than what is normally used.

This may indicate that we have aimed at a too high accuracy. Therefore, the corresponding results for accuracies of 10 and 20 kJ/mol are shown in Figure 4 and Table 2 (still with a confidence of 95%). It can be seen that in the first case, $92–2 \cdot 10^6$ snapshots are

needed with a full exponential average if the energies show a standard deviation (σ) of 10–17 kJ/mol (as in the study of Li et al.). This is still too large. On the other hand, if the cumulant approximation is employed, it is enough with 35–266 snapshots, which starts to be acceptable. With an accuracy of 20 kJ/mol, the corresponding numbers are 7–7400 for the exponential average and 10–66 with the cumulant approximation. This is probably the accuracy we can hope for in a standard QM/MM study, unless the spread of the individual energies is low.

As mentioned in the introduction, some authors have suggested the use of an arithmetic average instead. However, as can be seen from Eqn. 5, this will not reproduce the exponential average if σ is large (remember that μ is the arithmetic average). In fact, the arithmetic average will deviate from the exponential average by more than 4, 10 and 20 kJ/mol when σ is larger than 4.5, 7.1 and 10.0 kJ/mol, respectively.

The situation is similar for the minimum value: It does not coincide with or converge towards the exponential average. However, the minimum value is approximately proportional to σ, whereas the exponential average depends on σ^2 (Eqn. 5). Therefore, the minimum value is too negative for small σ, whereas it is too large for larger σ. The cross-over and the detailed accuracy depends on the number of samples. With 10 samples, the minimum value agrees with the exponential average within 4, 10 and 20 kJ/mol with 95% confidence when σ is less than 1, 8 and 11 kJ/mol, respectively, and the cross-over is around σ = 8 kJ/mol. With 1000 samples, the corresponding numbers are 1, 3 and 19 kJ/mol and the cross-over is at σ = 16 kJ/mol).

An important use of QM/MM studies on enzymes is to compare alternative reaction mechanisms, using the height of activation barrier as the discriminating criterion. This is an appreciably easier task, because it is not necessary to accurately estimate the actual barriers, only their relative size. Figure 5 shows the number of energies needed to reach the correct ranking in 95% of the simulations if μ of the two normal distributions differ by either Δμ = 10, 20 or 40 kJ/mol, and both are characterised by the same σ. Four different estimates of the activation energies are tested. The exponential average (red curves in Figure 5) shows a steep increase in the number of energies when σ is increased, as can be expected from the results in Figure 4. However, the number of energies does not rise to prohibitive large values until σ ≈ Δμ. For example, less than 11 energies are needed as long as σ < 30 kJ/mol if Δμ = 40 kJ/mol. Moreover, it can be seen that using the minimum value does not lead to any improvement in the convergence (yellow curves in Figure 6).

On the other hand, the convergence is strongly improved is the cumulant approximation is used (blue curves), but only for large values of σ. In fact, up to σ = Δμ, fewer energies are needed with exponential averaging than with the cumulant approximation. Instead, the most stable results are obtained with the arithmetic average (green curves in Figure 6). It always gives the lowest *n* among the four methods tested. Moreover, the increase is more modest as σ is increased. Thus, to discriminate between two mechanism with the activation barrier, the arithmetic average gives the most stable results, but the actual values of the barriers will be inaccurate. Of course, the detailed results may change if the two distributions are allowed to have different values of σ, but the arithmetic average will still have much better convergence properties than the other averages (and the situation can easily be simulated numerically).

Convergence criteria

Finally, I have also studied and compared four convergence criteria, used in previous investigations.[30,36–39] The results for three typical σ values (7, 10 and 13 kJ/mol) are shown in Figures 6a–c. It can be seen that all four convergence criteria give reasonable results, i.e. they give a meaningful variation for *n* values around which the Gaussian simulations indicate that

convergence is obtained. The reason for this is of course that they study essentially the same property, viz. how many energies that provide a significant contribution to the exponential sum and how sensitive this sum is to the largest term. If only one or a few values contribute, it is likely that the convergence is poor and that the result may change significantly if more energies are calculated. Among these three estimates, S_w shows the smallest variation: The range is only 0.1 throughout the range considered for all three cases and it may therefore be hard to design an accurate convergence value. DE has the disadvantage of being dependent on the value of ΔE_{EA}. Q shows the largest variation.

Table 3 shows the convergence value for each estimate for the four convergence criteria and the three σ values (using the convergence limits shown in Table 2). Unfortunately, it can be seen that all the convergence criteria depend somewhat on σ: Q increases with σ (from 7 to 13), whereas the other three estimates decrease. w_{max} shows the smallest variation, from 0.38 to 0.33, whereas for S_w, the variation is larger than the total range in the three studied cases, making it useless for general use. However, also for the other three criteria, the variation is quite large compared to the variation with n.

Furthermore, all criteria are independent on the accuracy threshold used. This is illustrated in Figure 6d and the last row in Table 3, in which the results are shown for σ = 18 kJ/mol with the desired accuracy of 20 kJ/mol. In that case, totally different values of the various criteria need to be used (last line in Table 3). Therefore, it must be concluded that the convergence limits vary with the system and with the desired accuracy in a way that needs to be tested for each system. Therefore, I instead recommend the use of σ as the convergence criterion. Using σ, the required number of energies can be directly read from Table 2 and the simulation can easily be redone if other accuracies or confidences are required.

Conclusions

In this paper, I have discussed how many QM/MM-minimised energy estimates are needed to obtain an accurate estimate of the exponential average (e.g. for the activation barrier or reaction energy). Assuming that the energies follow a Gaussian distribution, an unambiguous answer to this question can easily be obtained, as soon it is decided what accuracy and what confidence is required. Values needed to reach accuracies of 4, 10 and 20 kJ/mol with a confidence of 95% are shown in Table 2 as a function of a single parameter, the standard deviation of the individual energy values from different snapshots (σ).

For practical use, I suggest to first calculate three energy values, from which an estimate of σ is obtained. Then, Table 2 directly shows how many additional points are needed. Once these are obtained, a better estimate of σ is also gained. If σ is above ~10 kJ/mol, the user may prefer to employ the cumulant approximation, for which the required n is smaller (also in Table 2). However, if σ is too large, the required number can still become too large. Then, the user may need to use other methods to estimate the energy, preferably free-energy perturbation or related methods.[1–3] On the other hand, if the main interest is to compare alternative mechanisms, arithmetic averages give more stable results (but less accurate energies), as is shown in Figure 5.

Many different methods have been suggested to determine whether estimates using exponential averaging are reasonable. Unfortunately, many of them do not work at all when the sum is badly conditioned, i.e. when the probability to find the values that contribute most to the integral in Eqn. 4 becomes very low. Convergence criteria that judge how many energies significantly contribute to ΔE_{EA} (like w_{max}, Q and S_w) work better, but it is hard to suggest general convergence limits. Instead, I recommend the standard deviation of the energy distribution (σ) as the convergence criterion and it gives a direct estimate of how badly conditioned the problem is and how many energies are needed. It is also directly available

from any sample of energy estimates. It has previously been used to judge the convergence of free-energy perturbations, suggesting that σ should be less than 1–2 RT (2.5–5 kJ/mol).[44–46]

The estimates presented in this paper are based on the assumption that the QM/MM energy values follow a Gaussian distribution. Owing to the central limit theorem, this seems to be a reasonable assumption. Figure S1 in the supporting information shows histograms and normal plots of the eight data samples in Table 1 with 20 or more data points. They show a reasonably Gaussian distribution, possibly with a single outlier in one case (Figure S1a). Moreover, the estimates in Table 2 should not be considered as accurate values, but only as an indication of how many samples are needed and especially an indication whether there is any chance that a converged result can be obtained or not. In particular, it is unlikely that the distribution will be so favourable that a significantly smaller n is needed, and if a user want to make such a claim, he needs to provide strongly convincing evidence that this is actually the case. Likewise, if a user wants to employ the cumulant approximation, it becomes more important to provide evidence that the distribution actually is Gaussian.

The main take-home message from this paper is that QM/MM energy estimates from minimised MD snapshots with an exponential average is much more problematic than what has previously been assumed. The use of less than 20 snapshots will work only if the energy distribution is narrow or if a rather low accuracy is accepted. This is owing to the curse of the exponential averaging, illustrated in Figure 2b. Fortunately, the accuracy of the approximation used can easily be estimated from one single parameter, σ. It should be noted that the same method can be used and the same conclusions apply to any property that is obtained by exponential averaging, which is the proper approach when conformational sampling is performed with another energy function than the one used for the property calculation, e.g. when trying to calculate binding free energies at the QM level, based on MD simulations at the MM level.[32–35]

Supporting information

Histograms and normal plots of the eight data samples in Table 1 with 20 or more data points. The Supporting Information is available free of charge on the ACS Publications website.

Acknowledgements

This investigation has been supported by grants from the Swedish research council (project 2014-5540), from Knut and Alice Wallenberg Foundation (KAW 2013.0022) and from COST through Action CM1305 (ECOSTBio).

References

(1) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chemie - Int. Ed.* **2009**, *48* (7), 1198–1229.

(2) Ryde, U. QM / MM Calculations on Proteins. *Methods Enzymol.* **2016**, *577*, 119–158.

(3) Hu, H.; Yang, W. Development and Application of Ab Initio QM/MM Methods for Mechanistic Simulation of Reactions in Solution and in Enzymes. *J. Mol. Struct. THEOCHEM* **2009**, *898* (1–3), 17–30.

(4) Dieterich, J. M.; Werner, H. J.; Mata, R. A.; Metz, S.; Thiel, W. Reductive Half-Reaction of Aldehyde Oxidoreductase toward Acetaldehyde: Ab Initio and Free Energy Quantum Mechanical/molecular Mechanical Calculations. *J. Chem. Phys.* **2010**, *132* (3).

(5) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational Enzymology. *Chem. Commun.* **2010**, *46* (14), 2333–2512.

(6) Hu, L.; Söderhjelm, P.; Ryde, U. Accurate Reaction Energies in Proteins Obtained by Combining QM/MM and Large QM Calculations. *J. Chem. Theory Comput.* **2013**, *9* (1), 640–649.

(7) Oláh, J.; Mulholland, A. J.; Harvey, J. N. Understanding the Determinants of Selectivity in Drug Metabolism through Modeling of Dextromethorphan Oxidation by Cytochrome P450. *Proc. Natl. Acad. Sci.* **2011**, *108* (15), 6050–6055.

(8) Li, Y.; Zhang, R.; Du, L.; Zhang, Q.; Wang, W. Insight into the Catalytic Mechanism of Meta-Cleavage Product Hydrolase BphD: A Quantum Mechanics/molecular Mechanics Study. *RSC Adv.* **2015**, *15*, 66591–66597.

(9) Lonsdale, R.; Reetz, M. T. Reduction of A,ß-Unsaturated Ketones by Old Yellow Enzymes: Mechanistic Insights from Quantum Mechanics/molecular Mechanics Calculations. *J Am Chem Soc* **2015**, *137*, 14733–14742.

(10) Abad, E.; Zenn, R. K.; Kästner, J. Reaction Mechanism of Monoamine Oxidase from QM/MM Calculations. *J Phys Chem B* **2013**, *117*, 14238–14246.

(11) Sanchez-Martinez, M.; Marcos, E.; Tauler, R.; Field, M. J.; Crehuet, R. Conformational Compression and Barrier Height Heterogeneity in the N-Acetylglutamate Kinase. *J Phys Chem B* **2013**, *117*, 14261–14272.

(12) Rommel, J. B.; Kästner, J. The Fragmentation-Recombination Mechanism of the Enzyme Glutamate Mutase Studied by QM/MM Simulations. *J Am Chem Soc* **2011**, *133*, 10195–10203.

(13) Cooper, A. M.; Kästner, J. Averaging Techniques for Reaction Barriers in QM/MM Simulations. *ChemPhysChem* **2014**, *15* (15), 3264–3269.

(14) Li, Y.; Zhang, R.; Du, L.; Zhang, Q.; Wang, W. Catalytic Mechanism of C–F Bond Cleavage: Insights from QM/MM Analysis of Fluoroacetate Dehalogenase. *Catal. Sci. Technol.* **2016**, *6*, 73–80.

(15) Cohen, S.; Kozuch, S.; Hazan, C.; Shaik, S. Does Substrate Oxidation Determine the Regioselectivity of Cyclohexene and Propene Oxidation by Cytochrome P450? *J. Am. Chem. Soc.* **2006**, *128* (34), 11028–11029.

(16) Hu, P.; Zhang, Y. Catalytic Mechanism and Product Specificity of the Histone Lysine Methyltransferase SET7/9: An Ab Initio QM/MM-FE Study with Multiple Initial Structures. *J. Am. Chem. Soc.* **2006**, *128* (4), 1272–1278.

(17) Mata, R. A.; Werner, H. J.; Thiel, S.; Thiel, W. Toward Accurate Barriers for Enzymatic Reactions: QM/MM Case Study on P-Hydroxybenzoate Hydroxylase. *J. Chem. Phys.* **2008**, *128* (2), 1–8.

(18) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Compound I Reactivity Defines Alkene

Oxidation Selectivity in Cytochrome p450cam. *J. Phys. Chem. B* **2010**, *114* (2), 1156–1162.
(19) Kamp, M. W. Van Der; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl- CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* **2010**, *114*, 11303–11314.
(20) Kästner, J.; Sherwood, P. The Ribosome Catalyses Peptide Bond Formation by Providing High Ionic Strength. *Mol. Phys.* **2011**, *108*, 293–306.
(21) Metz, S.; Thiel, W. QM/MM Studies of Xanthine Oxidase Variations of Cofactor Substrate and Active-Site Glu802. *J. Phys. Chem. B* **2010**, *114*, 1506–1517.
(22) Lonsdale, R.; Hoyle, S.; Grey, D. T.; Ridder, L.; Mulholland, A. J. Determinants of Reactivity and Selectivity in Soluble Epoxide Hydrolase from Quantum Mechanics/molecular Mechanics Modeling. *Biochemistry* **2012**, *51* (8), 1774–1786.
(23) Christov, C. Z.; Lodola, A.; Karabencheva-Christova, T. G.; Wan, S.; Coveney, P. V.; Mulholland, A. J. Conformational Effects on the pro-S Hydrogen Abstraction Reaction in Cyclooxygenase-1: An Integrated QM/MM and MD Study. *Biophys. J.* **2013**, *104* (5), L5–L7.
(24) Sokkar, P.; Boulanger, E.; Thiel, W.; Sanchez-Garcia, E. Hybrid Quantum Mechanics/molecular Mechanics/coarse Grained Modeling: A Triple-Resolution Approach for Biomolecular Systems. *J. Chem. Theory Comput.* **2015**, *11* (4), 1809–1818.
(25) Logunov, I.; Schulten, K. Quantum Chemistry: Molecular Dynamics Study of the Dark-Adaptation Process in Bacteriorhodopsin. *J Am Chem Soc* **1996**, *118*, 9727–9735.
(26) Lonsdale, R.; T., H. K.; Zurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J. Quantum Mechanics/molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9. *J Am Chem Soc* **2013**, *135*, 8001–8015.
(27) Saura, P.; Suardíaz, R.; Masgrau, L.; Lluch, J. M.; González-Lafont, A. Unraveling How Enzymes Can Use Bulky Residues to Drive Site-Selective C–H Activation: The Case of Mammalian Lipoxygenases Catalyzing Arachidonic Acid Oxidation. *ACS Catal.* **2014**, *4*, 4351–4363.
(28) Li, Y.; Shi, X.; Zhang, Q.; Hu, J.; Chen, J.; Wang, W. Computational Evidence for the Detoxifying Mechanism of Epsilon Class Glutathione Transferase toward the Insecticide DDT. *Environ. Sci. Technol.* **2014**, *48*, 5008–5016.
(29) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.Ramos, M. J.; Fernandes, P. A. Enzymatic Flexibility and Reaction Rate: A QM/MM Study of HIV-1 Protease. *ACS Catal.* **2015**, *5*, 5617–5626.
(30) Li, Y.; Zhang, R.; Du, L.; Zhang, Q.; Wang, W. How Many Conformations of Enzymes Should Be Sampled for DFT / MM Calculations ? A Case Study of Fluoroacetate Dehalogenase. *Int. J. Mol. Sci.* **2016**, *17*, 1372 (8 pages).
(31) Jiang, J.; Lu, J.; Lu, D.; Liang, Z.; Li, L.; Ouyang, S.; Kong, X.; Jiang, H.; Shen, B.; Luo, C. Investigation of the Acetylation Mechanism by GCN5 Histone Acetyltransferase. *PLoS One* **2012**, *7* (5), e36660.
(32) Ryde, U.; Söderhjelm, P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016**, *116*, 5520–5566.
(33) Olsson, M. A.; Söderhjelm, P.; Ryde, U. Converging Ligand-Binding Free Energies Obtained with Free-Energy Perturbations at the Quantum Mechanical Level. *J. Comput. Chem.* **2016**, *37*, 1589–1600.
(34) Olsson, M. A.; Ryde, U. Comparison of QM/MM Methods To Obtain Ligand-Binding Free Energies. *J. Chem. Theory Comput.* **2017**, *13*, 2245–2253.
(35) Boresch, S.; Woodcock, H. L. Convergence of Single-Step Free Energy Perturbation.

Mol. Phys. **2017**, *DOI: /10.1*, 14 pages.
(36) Kish, L. *Survey Sampling*; Wiley: New York, 1965.
(37) Rod, T. H.; Ryde, U. Accurate QM/MM Free Energy Calculations of Enzyme Reactions: Methylation by Catechol O-Methyltransferase. *J. Chem. Theory Comput.* **2005**, *1* (6), 1240–1251.
(38) Mikulskis, P.; Genheden, S.; Ryde, U. A Large-Scale Test of Free-Energy Simulation Estimates of Protein-Ligand Binding Affinities. *J. Chem. Inf. Model.* **2014**, *54* (10), 2794–2806.
(39) Wang, M.; Li, P.; Jia, X.; Liu, W.; Shao, Y.; Hu, W.; Jun, Z.; Brooks, B. R.; Mei, Y. An Efficient Strategy for the Calculation of Solvation Free Energies in Water and Chloroform at the Quantum Mechanical/Molecular Mechanical Level. *J. Chem. Inf. Model.* **2017**, *in press*.
(40) Box, G. E. P.; Muller, M. E. A Note on the Generation of Random Normal Deviates. *Ann. Math. Stat.* **1968**, *29* (2), 610–611.
(41) Zwanzig, R. W. *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, 2001.
(42) Hummer, G. Fast-Growth Thermodynamic Integration: Error and Efficiency Analysis. *J. Chem. Phys.* **2001**, *114* (17), 7330–7337.
(43) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction. *J. Chem. Theory Comput.* **2006**, *2* (2), 452–461.
(44) Wood, R. H.; Miihlbauer, W. C. F. Systematic Errors In Free Energy Perturbation Calculations Due to a Finlte Sample of Configuration Space: Sample-Slze Hysteresis. *J. Phys. Chem.* **1991**, *95*, 6670–6675.
(45) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114* (32), 10235–10253.
(46) Dellago, C.; Hummer, G. Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy* **2014**, *16* (1), 41–61.
(47) Cohen, S.; Kozuch, S.; Hazan, C.; Shaik, S. Does Substrate Oxidation Determine the Regioselectivity of Cyclohexene and Propene Oxidation by Cytochrome P450? *J. Am. Chem. Soc.* **2006**, *128* (34), 11028–11029.
(48) Johannes, K.; Paul, S. The Ribosome Catalyses Peptide Bond Formation by Providing High Ionic Strength. *Mol. Phys.* **2010**, *108*, 293–306.

Table 1. Method of averaging (arithmetic average, AA, or exponential average, EA), number of energies in the average (n) and the standard deviation among the energies (σ in kJ/mol) in a number of QM/MM studies of enzyme reactions. The list is not exhaustive, but includes the articles listed in references [13] and [30].

Group	Year	Ref.	Method	n	σ
Schulten	1996	[25]	EA	25	44–49
Shaik	2006	[47]	AA	3	6.8
Zhang	2006	[16]	AA	11	5.0–8.5
Thiel	2008	[17]	AA	10	5.4–12
Mulholland	2010	[18]	AA, EA	7–10	6.6–26
Mulholland	2010	[19]	AA, EA	5	3.3–6.3
Kästner	2010	[48]	AA, EA	10	72–97
Thiel	2010	[21]	AA	2	0.6–5.3
Kästner	2011	[12]	EA	6	8.2–53
Mulholland	2011	[7]		3	2.1–3.1
Mulholland	2012	[22]	AA	10	4.8–13
Field	2013	[11]	EA	5	4.7–18
Kästner	2013	[10]	EA	4	40–63
Mulholland	2013	[23]	AA, EA	5	18
Mulholland	2013	[26]	EA	2–10	5.3–16
Ryde	2013	[6]	AA	10	4.4–9.2
Kästner	2014	[13]	AA, EA	65	16
Lluch	2014	[27]	EA	9–11	11–24
Wang	2014	[28]	EA	7	36
Lonsdale	2015	[9]	EA	3	1.5–36
Ramos	2015	[29]	EA	39	33
Thiel	2015	[24]	AA	4–10	4.6–24
Wang	2015	[8]	EA	3	12–21
Wang	2016	[30]	EA	20	10–17

Table 2. Number of energies needed to obtain the correct ΔE_{EA} within a certain threshold (θ = 4, 10, or 20 kJ/mol) with a confidence of 95% as a function of σ (in kJ/mol).

σ	Exponential average			Cumulant approximation		
	θ = 4	10	20	4	10	20
1	1	1	1	2	2	2
2	2	1	1	2	2	2
3	3	1	1	5	2	2
4	8	1	1	10	3	2
5	20	2	1	18	4	2
6	42	4	1	35	6	3
7	126	8	2	55	11	4
8	372	15	3	101	15	5
9	1560	38	4	147	24	7
10	4180	92	7	228	35	10
11	19900	251	12	318	50	13
12	89300	800	28	448	68	18
13	511000	2660	67	593	102	24
14	3960000	11800	178	809	127	32
15	28400000	57200	535	1040	167	42
16		317000	1780	1330	214	53
17		2190000	7410	1680	266	66
18			33800	2290	340	83
19			182000	2650	407	106
20			1250000	3170	522	127
25				7420	1320	292
30				15760	2590	627
35				30100	4580	1120
40				51000	7670	2070

Table 3. Quality criteria at the minimum required number of energy values (according to Table 2) in simulations using four different values of σ (kJ/mol) and two values for the accuracy threshold (θ, kJ/mol).

σ	θ	w_{max}	Q	S_w	DE
7	4	0.38±0.01	6.7±0.1	0.49±0.00	-0.29±0.01
10	4	0.35±0.01	9.7±0.2	0.36±0.00	-0.40±0.01
13	4	0.33±0.01	13.3±0.4	0.28±0.00	-0.49±0.01
18	20	0.60±0.01	3.4±0.1	0.14±0.00	-0.52±0.01

Figure 1. Number of energies needed to obtain the correct ΔE_{EA} within 4 kJ/mol with a confidence of 95% as a function of σ. Note the logarithmic scale of the *y* axis.

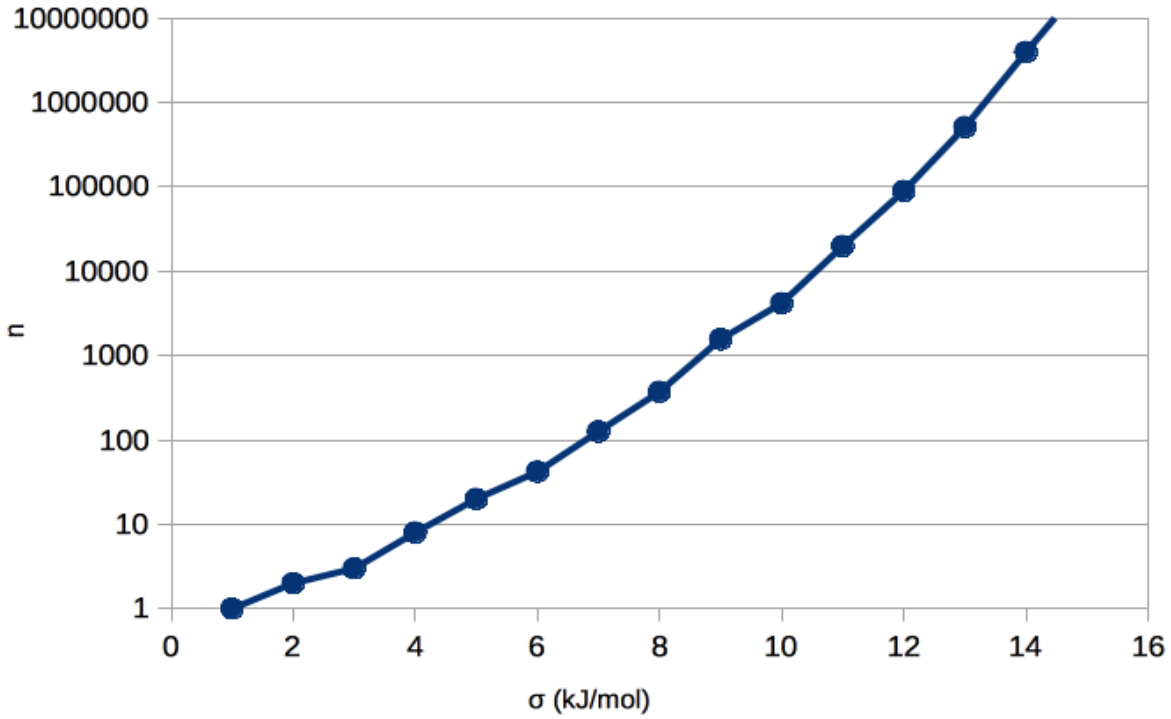
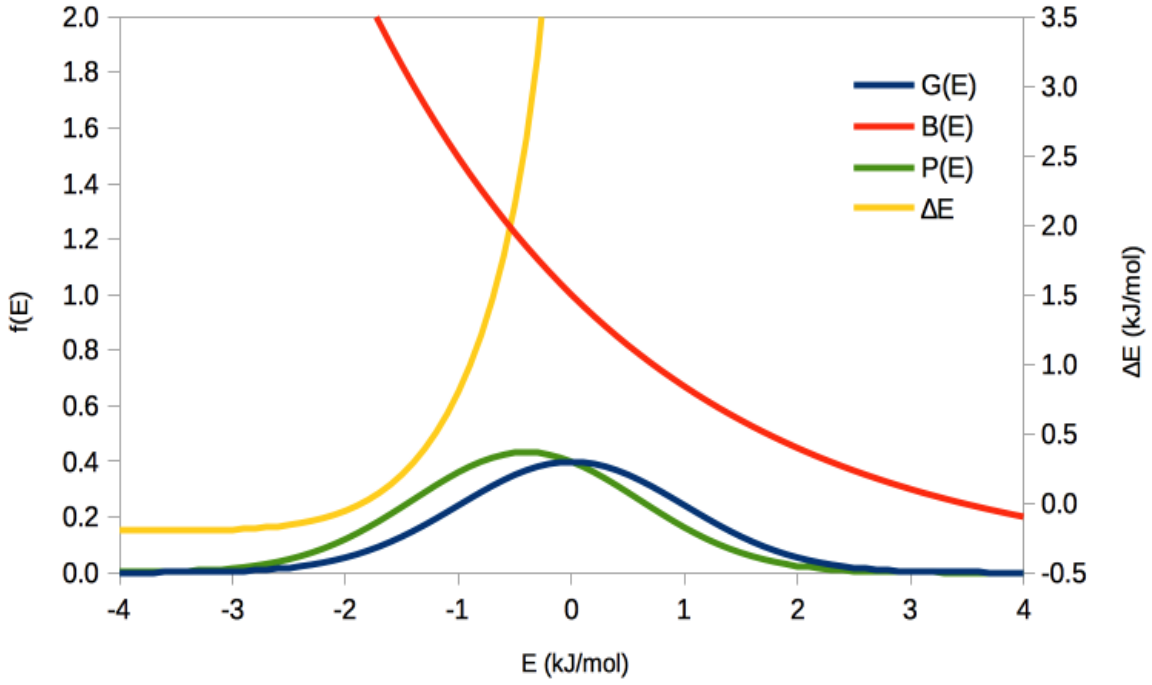
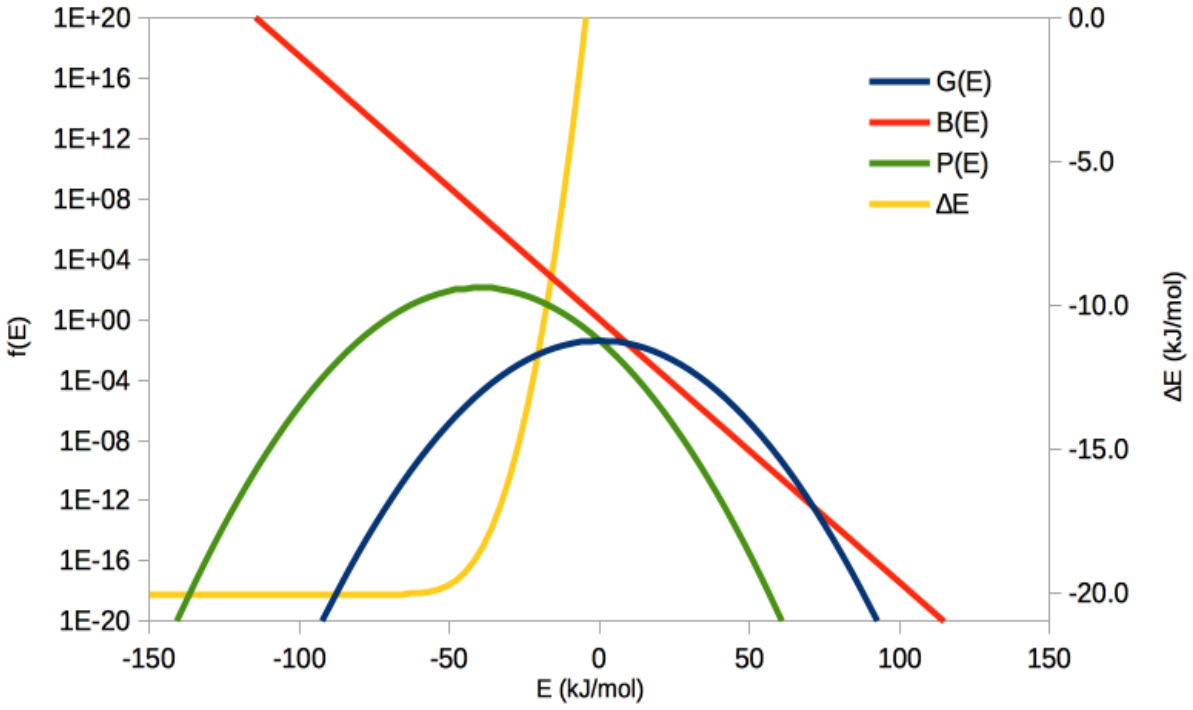


Figure 2. The two terms within the integral in Eqn. 4, viz. the Gaussian probability function, $G(E)$ and the Boltzmann distribution, $B(E)$, as well as their product, $P(E) = G(E) * B(E)$, plotted as a function of the energy for two different values of σ, (a) 1 and (b) 10 kJ/mol. Note the logarithmic scale in (b). Both figures also show ΔE_{EA} (right y axis), which is calculated from the integral of $P(E)$ for values between $-E$ and $+E$.



a



b

Figure 3. Number of energies needed to obtain the correct ΔE_{EA} within 4 kJ/mol with a confidence of 95% as a function of σ employing the cumulant approximation.

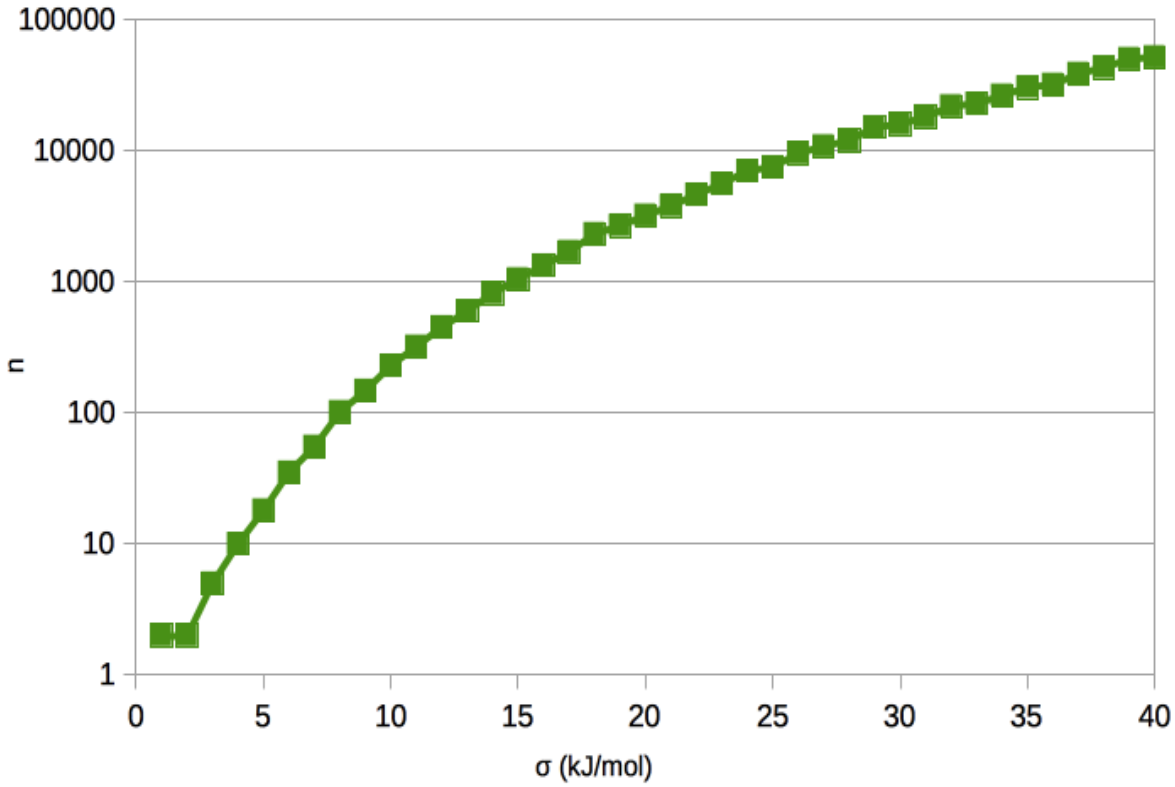


Figure 4. Number of energies needed to obtain the correct ΔE_{EA} within a certain accuracy (4, 10 or 20 kJ/mol) with a confidence of 95% as a function of σ using either the exponential average (EA) in Eqn. 2 or the cumulant approximation (CA) in Eqn. 3.

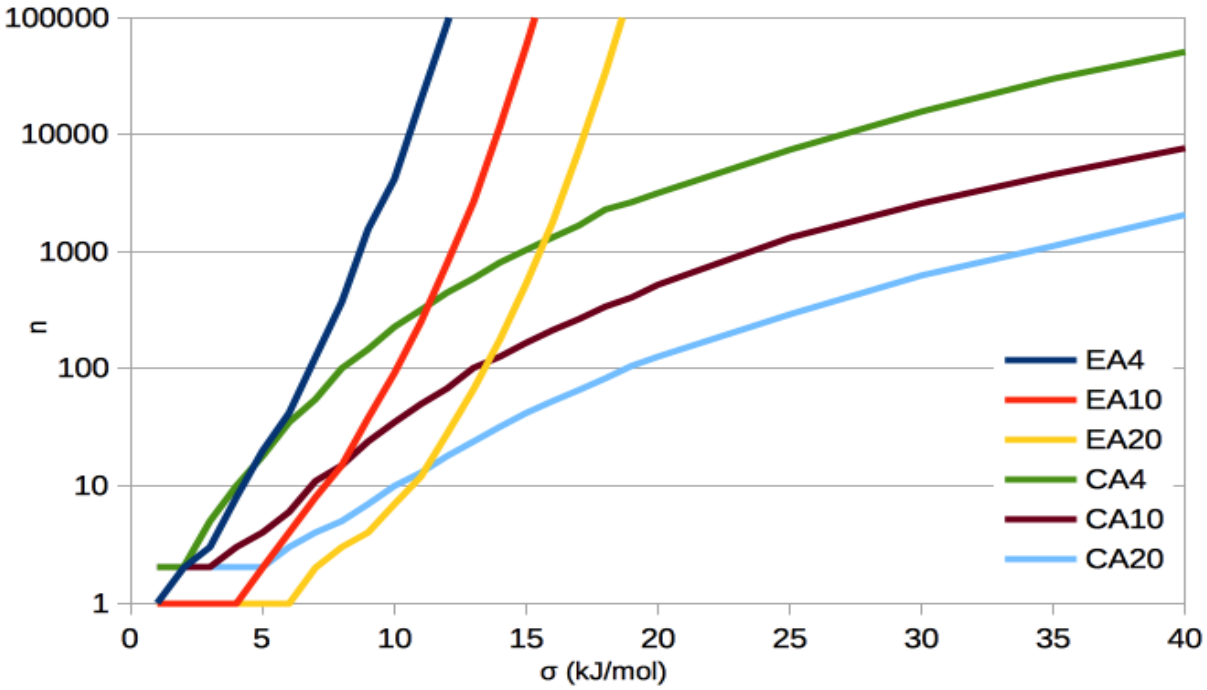


Figure 5. Number of energies needed to obtain the correct ranking of two Gaussian distributions $G_1(0,\sigma)$ and $G_2(\Delta\mu,\sigma)$ in 95% of the 1000 simulations. Four different energies were tested: exponential averages (E; red curves), cumulant approximation (C; blue), arithmetic averages (A; green) and the minimum value (M; yellow–brown) and three values of $\Delta\mu$, 10, 20 and 40 kJ/mol. Note the logarithmic scale of the *y* axis.

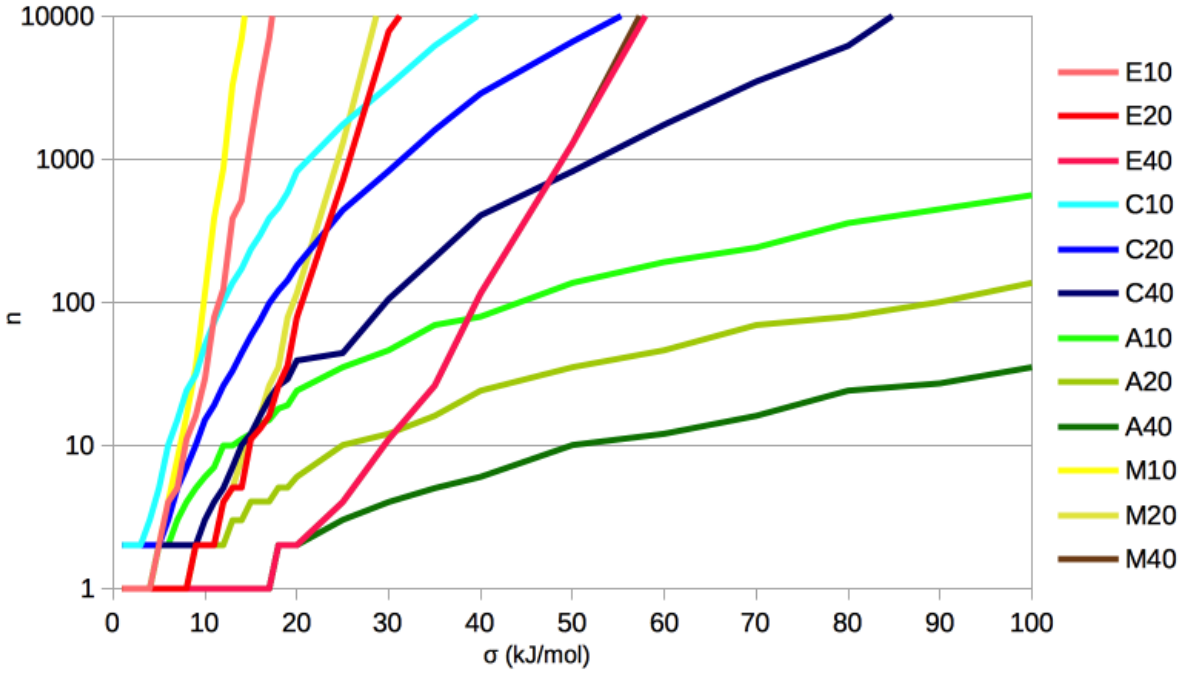
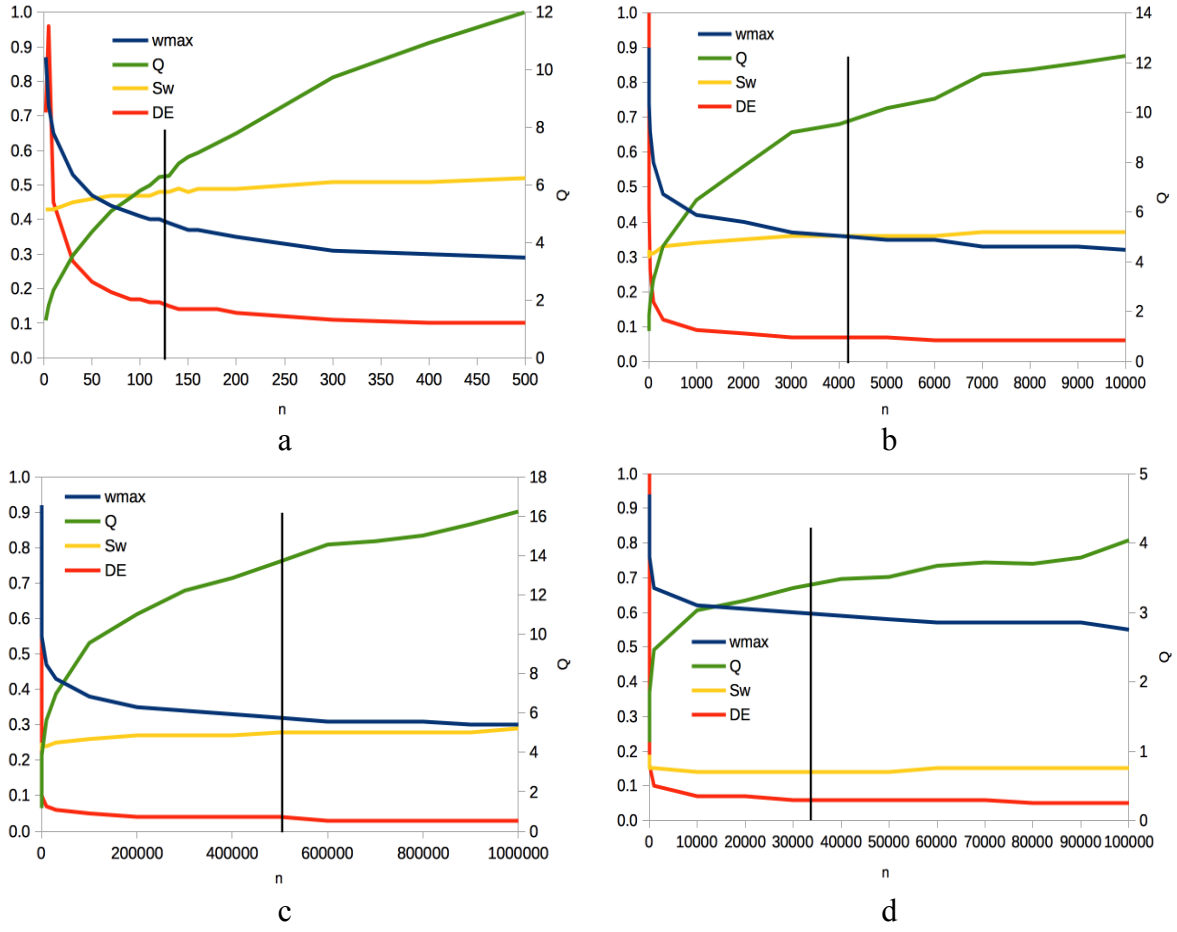


Figure 6. The dependence of four convergence criteria (w_{max}, S_w and DE on left axis and Q on left axis; for DE, the negative value is shown) on the number of energies included (n). Results are shown for (a) σ = 7, (b) 10, (c) 13 and (d) 18 kJ/mol. The black vertical bar indicates the number of energies giving a ΔE_{EA} converged to within 4 (a–c) or 20 (d) kJ/mol with a confidence of 95%.



TOC graphics