Quantum harmonic free energies for biomolecules and nanomaterials

Alec White
Quantum Simulation Technologies, Inc.
Chenghan Li
California Institute of Technologu
Garnet Kin-Lic Chan (gkc1000@gmail.com)
California Institute of Technology

Brief Communication

Keywords:

Posted Date: January 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1132772/v1

Quantum harmonic free energies for biomolecules and nanomaterials

Alec F. White[1,2], Chenghan Li[1] and Garnet Kin-Lic Chan[1]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, 91125, CA, United States of America.

[2]Present address: Quantum Simulation Technologies, Inc., Cambridge, MA 02139, United States of America.

Abstract

Obtaining the free energy of large molecules from quantum mechanical energy functions is a longstanding challenge. We describe a method that allows us to estimate, at the quantum mechanical level, the harmonic contributions to the thermodynamics of molecular systems of unprecedented size, with modest cost. Using this approach, we compute the vibrational thermodynamics of a series of diamond nanocrystals, and show that the error per atom decreases with system size in the limit of large systems. We further show that we can obtain the vibrational contributions to the binding free energies of prototypical protein-ligand complexes where the exact computation is too expensive to be practical. Our work raises the possibility of routine quantum mechanical estimates of thermodynamic quantities in complex systems.

The contributions to the free energy from atomic motion are critically important to the thermodynamics and kinetics of biological, chemical, and materials systems. Changes in such contributions govern processes ranging from the affinity of drug binding to structural phase transitions in crystals. When the internal energy is computed at the quantum mechanical level, a harmonic approximation is often the only feasible option to describe atomic motion. However, for large systems such as nanostructures and biomolecules, computing free energy contributions is expensive even within the harmonic approximation. For a system of N atoms, the Hessian matrix which describes the vibrations requires $O(3N)$ gradient calculations, or $O(3N)$ times the cost of computing the internal energy. This is clearly prohibitive when N

is large, making free-energy computation in large systems with quantum mechanical methods a major contemporary challenge [1].

There are many possible strategies to speed up harmonic vibrational analysis, including methods based on a partial Hessian[2, 3], iterative diagonalization [4], and Hessian-free methods that use molecular dynamics to approximate the the harmonic problem[5, 6]. Here we describe a different strategy where we estimate vibrational thermodynamic quantities directly without ever computing the full Hessian or taking advantage of any local structure.

The starting point is to express each harmonic thermodynamic quantity as a matrix function trace. Then, our technique contains three elements. First, we sample the matrix trace operation using random vectors and stochastic Lanczos quadrature [7]. Second, we compute the Hessian-vector product at the same cost as the gradient from the difference of gradients at displaced geometries, bypassing the Hessian construction entirely. Third, we ameliorate the stochastic error, especially for free energy differences, through a form of correlated sampling. Related stochastic methods have been used for anharmonic corrections to the harmonic free energy[8, 9], as well as in stochastic electronic structure[10], but to our knowledge this is the first time these ideas have been brought to bear on the harmonic thermodynamic quantities themselves. As we demonstrate, this allows us to compute at the quantum mechanical level and with modest cost, vibrational free energy contributions for nanocrystals with more than 600 atoms, and free energy differences in protein-ligand complexes with more than 3000 atoms.

Theory

We first express the harmonic thermochemical quantities as traces of matrix functions. In particular, we are interested in the zero-point energy (ZPE),

$$\mathrm{ZPE} = \sum_I \frac{\omega_I}{2} = \sum_I \frac{\sqrt{\omega_I^2}}{2} = \mathrm{Tr}\left[\frac{\sqrt{\mathbf{D}}}{2}\right] \tag{1}$$

the thermal contribution to the enthalpy,

$$H_{\mathrm{vib}} = \sum_I \omega_I \left(\frac{e^{-\beta\omega_I}}{1-e^{-\beta\omega_I}}\right) = \mathrm{Tr}\left[\frac{\sqrt{\mathbf{D}}\exp(-\beta\sqrt{\mathbf{D}})}{1-\exp(-\beta\sqrt{\mathbf{D}})}\right] \tag{2}$$

and the entropy,

$$\begin{aligned} S_{\mathrm{vib}}/k_B &= \sum_I \left[\beta\omega_I \frac{e^{-\beta\omega_I}}{1-e^{-\beta\omega_I}} - \ln\left(1-e^{-\beta\omega_I}\right)\right] \\ &= \mathrm{Tr}\left[\beta\frac{\sqrt{\mathbf{D}}\exp(-\beta\sqrt{\mathbf{D}})}{1-\exp(-\beta\sqrt{\mathbf{D}})} - \ln\left(1-\exp(-\beta\sqrt{\mathbf{D}})\right)\right] \end{aligned} \tag{3}$$

Here β is the inverse temperature, $\{\omega_I\}$ is the set of normal mode frequencies, and $\mathbf{D}$ is the mass-weighted Hessian matrix.

The above expressions have the form $\mathrm{Tr} f(\mathbf{D})$. We now employ a stochastic estimator of the trace. The simplest version writes $\mathrm{Tr} f(\mathbf{D}) \approx \frac{M}{n} \sum_{l=1}^{n} \mathbf{v}_l^{\mathrm{T}} f(\mathbf{D}) \mathbf{v}_l$ where $\mathbf{v}_l$ are a set of n random vectors with zero mean and unit covariance, and $M = 3N - 6$ is the dimension of $\mathbf{D}$. This direct stochastic evaluation requires a polynomial approximation of $f(\mathbf{D})$, which is typically carried out using a Chebyshev expansion [11]. A closely related idea, which we use in this work, is stochastic Lanczos quadrature[7]. In this technique, the polynomial approximation is generated for the scalar $\mathbf{v}_l^T f(\mathbf{D}) \mathbf{v}_l$ rather than globally for the function $f(\mathbf{D})$. We have found the stochastic Lanczos method to be slightly superior to the Chebyshev polynomial approach for the quantities in this work (see the supporting information for further comparison). Within the polynomial expansion, the main operation is the matrix-vector product $\mathbf{Dx}$, where $\mathbf{x}$ is in the stochastic Lanczos space. This can be computed from the difference of gradients, displaced by $\delta \mathbf{x}$ in mass-weighted coordinates, for small δ. Thus no Hessian is needed at all in this approach (further details are provided in the Methods section).

Within the above scheme, there are two sources of error. The first is from the order of the Lanczos quadrature, m. This vanishes when m is greater than or equal to the matrix dimension. The second is the sampling error, which decreases like $1/\sqrt{n}$ for n random vectors. For a quadrature order of m and n samples, the cost of the method is equal to $O(mn)$ gradient calculations. To reduce the statistical error, we use a form of correlated sampling. For the absolute thermodynamic quantities computed for the diamond nanocrystals we employ a high-level quantum mechanical approach as well as a cheaper low-level method (for example, a force-field, or semi-empirical quantum-mechanical approach) where the exact computation of the harmonic thermodynamic quantity X_{low} is possible. Then, the free energy contribution for the high-level method is obtained as

$$X_{\mathrm{high}} = X_{\mathrm{low}} + \Delta \tag{4}$$

where Δ is computed by applying the stochastic Lanczos quadrature to the difference of high-level and low-level methods. In the case of protein(P)-ligand(L) binding, we are interested in the difference between the holo (ligand-bound) state and the apo (ligand-free) state, i.e.

$$X_{\mathrm{bind}} = X_{\mathrm{P+L}} - X_{\mathrm{P}} - X_{\mathrm{L}} \tag{5}$$

where X_{bind} represents the ligand binding free energy, enthalpy, entropy, etc. In this case, we perform correlated sampling by using the same random vectors in the stochastic Lanczos treatment of the P+L, P, L systems (zeroing out elements for P and L respectively).

Results

As a first application, we take diamond nanocrystals (Figure 1a) as prototypical nanomaterial, and compute the free energies as a function of size. We employ the high-

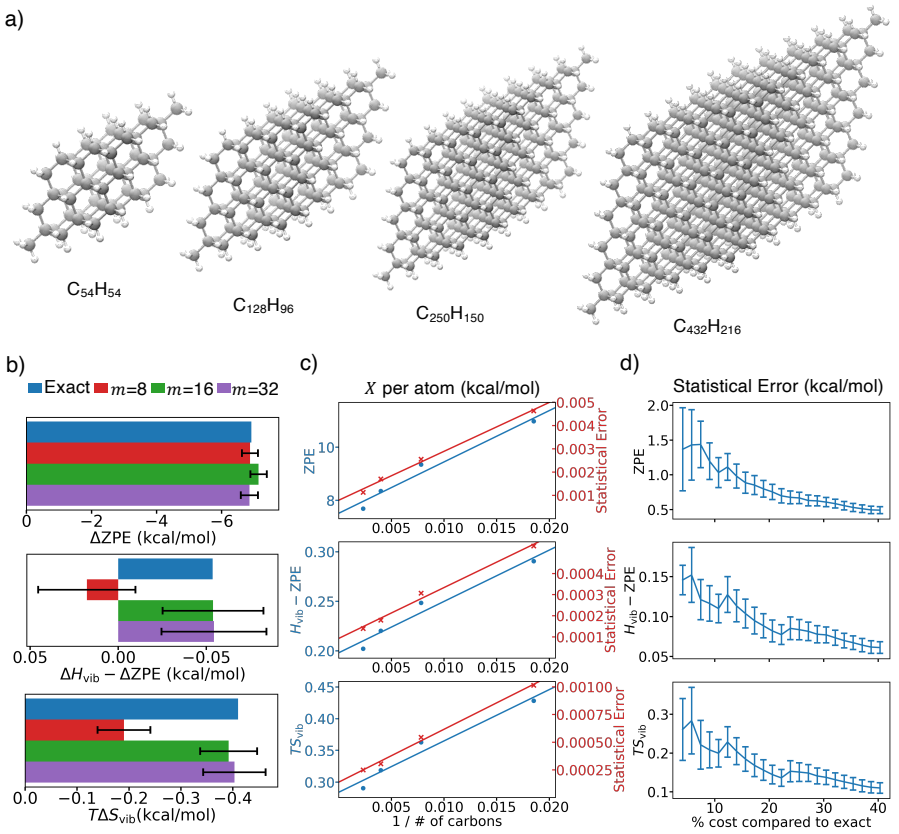

Fig. 1 **a)** Structures and chemical formulae of the diamond nanocrystals used in the calculations. **b)** Δ values estimated with different Lanczos quadrature orders (m) for the smallest system ($C_{54}H_{54}$) compared to the exact values. Error bars were estimated using the standard error of 50 samples. **c)** Per-atom quantity values and errors as a function of system size for a fixed number of samples (50 samples). The solid lines correspond to linear fits to inverse size. **d)** Statistical errors (for $m = 16$ quadrature) in absolute thermodynamic quantities of $C_{432}H_{216}$ as a function of % of computational cost of the exact calculation (error bars denote error of error).

and low-level correlated sampling approach described above, with Kohn-Sham density functional theory (DFT) with the PBE functional [12] as the high-level method and the semi-empirical extended tight-binding (XTB) method [13] as the low-level method. For the smallest system ($C_{54}H_{54}$), we can compute the Hessian explicitly to provide an exact reference. Figure 1b shows Δ quantities for the zero-point energy, the thermal enthalpy, and the entropy respectively, as a function of quadrature order m. The error bar indicates the statistical error for 50 samples. A stochastic quadrature level of $m = 8$ does not provide sufficient accuracy, so we choose $m = 16$ for further calculations. Figure 1c shows the value and stochastic error per atom (estimated as one standard deviation) for the ZPE, thermal enthalpy, and entropy respectively. We note the error decreases with the size of the system, which is evidence of "self-averaging" due to the large system size. Thus if one is interested in

per-atom quantities, as is often the case for thermodynamics, for example to locate phase transitions, our stochastic approach becomes increasingly more efficient in a large system. The difference between our largest simulation and the extrapolated thermodynamic limits for the per-atom ZPE, enthalpy, and entropy is only 0.2 kcal/mol, 0.004 kcal/mol, and 0.006 kcal/mol respectively; statistical errors with 50 samples are about 0.001 kcal/mol or less. In fact, with a *single* sample, one can estimate the per-atom quantities with a statistical error of less than 0.01 kcal/mol at a 120-fold speedup relative to the exact Hessian calculation.

If one is instead interested in the absolute values, the method can still be cheaper than the full computation of the Hessian. Figure 1d shows the stochastic error for the largest carbon system ($C_{432}H_{216}$) as a function of computational cost. At less than 20% of the exact calculation's computational effort, the error estimate is well within 1 kcal/mol. Less precise estimates can be obtained even more cheaply; a 20 times speedup is possible if 2 kcal/mol of error in the absolute quantities is tolerable.

Vibrational contributions to the free energy are also central to the study of biomolecules. For protein-ligand interactions in particular, where the aggregate binding is often only 5-10 kcal/mol, the vibrational contribution to binding free energies can be significant. Although the harmonic approximation is not necessarily a faithful approximation in these systems, the harmonic contributions nonetheless provide a first estimate of the thermal and entropic contributions. For these quantities, we use the fixed random vector correlated sampling approach described above, and employ the semi-empirical extended tight binding method as the quantum mechanical theory. We do not include explicit water molecules in the simulation: in principle, these could be included at additional cost, or the desolvation contribution to the free energy can be separately estimated by standard continuum methods [14]. We first study a system which is just small enough that exact results can be obtained at a high computational cost: a cutout ($\sim$ 1600 atoms) of the human tankyrase 2 (TNKS2) protein with a bound ligand shown in Figure 2a (see Methods for more information). In Figure 2b we show the thermal contributions to the binding enthalpy, entropy and free energy for stochastic Lanczos quadrature order $m = 16$. We do not separately plot the zero-point energy in this case because the change in the zero-point energy due to binding is small and not easily resolvable with a small number of random samples. For each of the above quantities, the error due to the Lanczos order is less than 1 kcal/mol. A significant physical quantity is the thermal contribution to the binding free energy ($\Delta G_{\text{vib}} - \Delta\text{ZPE}$). From Figure 2c we see that this can be estimated with a statistical error of less than 1 kcal/mol with a cost of roughly 10% of the exact Hessian calculation.

Finally, we apply our approach to HIV protease bound to a small molecule (JE-2147)[16] (Figure 3a). This system contains over 3000 atoms including hydrogens. The number of gradient calculations required to compute the full Hessian in this case ($\sim$ 10,000) is so large that the exact computation is expensive even at the level of semi-empirical quantum mechanics (and would be completely impractical at the level of full density functional theory). In Table 1 we show our stochastic estimates of the harmonic contributions to the thermodynamic binding quantities, using a quadrature order of $m = 20$. Interestingly, the thermal contributions ($\Delta H_{\text{vib}}-$ZPE and TS_{vib})

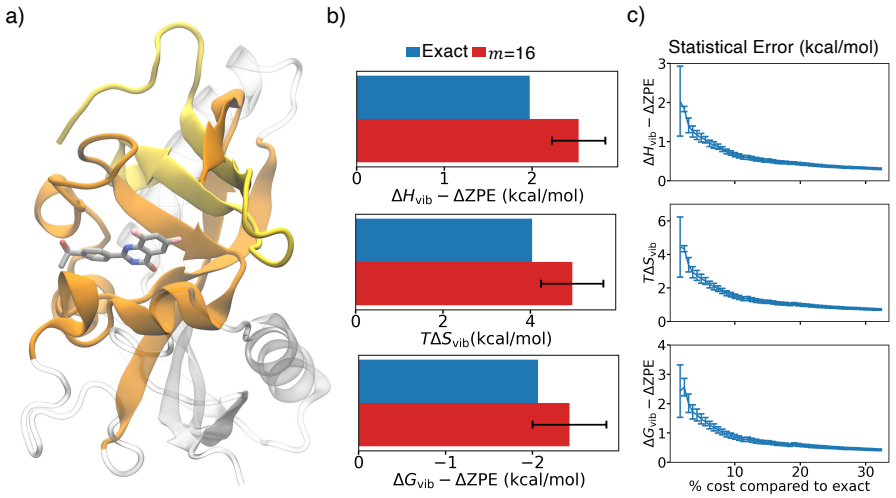

Fig. 2 **a)** The TNKS2 complex. The truncated part of the protein is shown as transparent, while the remaining two protein chains and the ligand are colored orange, yellow and grey, respectively. Image rendered by VMD[15]. **b)** The thermal enthalpy and entropy, and free energy of binding for the TNKS2 system for Lanczos order ($m = 16$). The error bars represent $\pm$ one standard error from 100 random samples. **c)** Statistical errors in binding free energy quantities as a function of % of cost of the exact calculation (error bars denote error of error).

Quantity	value	error
$\Delta H_{\mathrm{vib}} - \Delta\mathrm{ZPE}$	1.88	0.72
$T\Delta S_{\mathrm{vib}}$	1.69	1.79
$\Delta G_{\mathrm{vib}} - \Delta\mathrm{ZPE}$	0.19	1.13

Table 1 Contributions to the binding free energy from thermal enthalpy and entropy at 298.15 K for the JE-2147-HIV protease system. The error estimate in the final column is one standard error from 50 random samples (corresponding to 10% of exact cost). All values are given in kcal/mol.

to the free energy are both similar, small, and of opposite sign, meaning that the total thermal free energy contribution is almost zero. Nonetheless, estimating $\Delta G_{\mathrm{vib}} - \mathrm{ZPE}$ to an accuracy of 1 kcal/mol is clearly feasible within our scheme at roughly 10% of the estimated cost of the exact calculation (Figure 3b).

Discussion

We have presented results which demonstrate the feasibility of computing harmonic contributions to the free energy at the quantum mechanical level for systems of several thousand atoms. The cost is greatly reduced from that needed to compute the Hessian of the system. This is particularly true when one is interested in intensive (or "per-atom") quantities, where self-averaging behavior shows that in large systems, we may estimate the quantities at a cost comparable to that of a few energy evaluations. This holds promise in evaluating thermodynamic transitions in materials involving large unit cells, for example, those associated with alloys and disorder. In the case of free energy differences, the correlated sampling technique employed

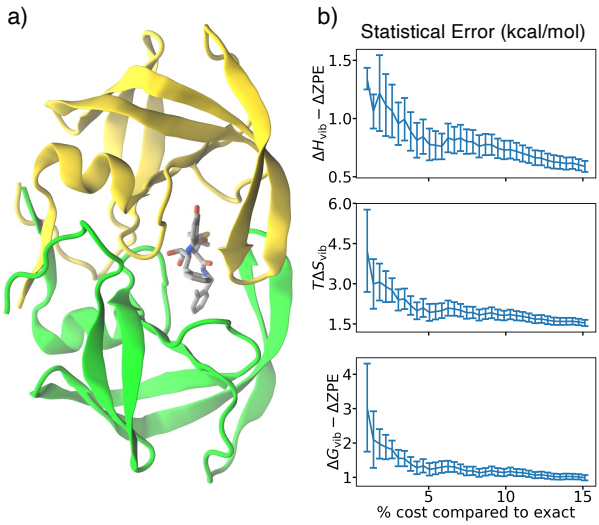

Fig. 3 **a)** An image of the JE-2147-HIV protease complex. The two protein chains and the ligand are shown in yellow, green, and grey respectively. Image rendered by VMD[15]. **b)** Statistical errors in binding free energy quantities as a function of % of cost of the exact calculation (error bars denote error of error).

here makes the evaluation of even small thermal free energy differences, as found in protein-ligand complexes, feasible at the level of 1 kcal/mol.

An additional advantage of the current approach is that the cost may be continually tuned. This is relevant to new applications, for example in the computational screening for therapeutics [1, 17], where less precise estimates are an acceptable tradeoff for speed. Although the harmonic approximation is not necessarily quantitative in biomolecular systems, the increased facility to obtain harmonic estimates also raises the possibility for new approaches to compute anharmonic contributions to free energies. In summary, the technique presented here suggests that the estimation of free energy effects at the quantum mechanical level for systems with hundreds or thousands of atoms need not be considered a future challenge [1], but one which can be begin to be addressed today.

Methods

Stochastic Lanczos quadrature

The stochastic Lanczos method is a numerical method which has been employed in different contexts (see e.g. Ref. [18] for an early application in quantum many-body systems). We follow the general mathematical formulation in Ref. [7]. The Lanczos iterations were performed starting from a vector randomly selected from a Rademacher distribution. The Lanczos iterations require the action of the mass-weighted Hessian matrix on this random vector. We compute this matrix-vector

product from finite difference gradient calculations:

$$\mathbf{Dv} = \mathbf{M}^{-1/2}\frac{\mathbf{g}(\delta\mathbf{v}) - \mathbf{g}(-\delta\mathbf{v})}{2\delta}. \tag{6}$$

Here, $\mathbf{D}$ as the mass-weighted hessian matrix, $\mathbf{g}$ is the gradient, $\mathbf{M}$ is the diagonal matrix of masses. The displacement is given by

$$\delta\mathbf{v} = \mathbf{M}^{-1/2}\mathbf{v} \tag{7}$$

where the factor of $\mathbf{M}^{-1/2}$ accounts for the mass-weighting. The value of δ is chosen based on the norm of the random vector so that the average displacement per atom is 0.0012 Å.

Additionally, we also implemented and tested a Chebyshev fitting method as an alternative to the stochastic Lanczos quadrature. We found that often a higher order Chebyshev fit was required making it a slightly more expensive alternative to Lanczos quadrature. However, there may situations where Chebyshev fitting is preferable. Results from the Chebyshev fitting method are presented in the supplementary information.

Calculations on diamond nanocrystals

Diamond nanocrystals were constructed by creating supercells of the bulk diamond unit cell and then capping with hydrogens. The resulting structure was optimized using the PBE functional and the def2-SV(P) basis set. All DFT calculations were performed with the ORCA program[19, 20]. The structure was also optimized using the second generation extended tight-binding (GFN2-xTB) method[21] as implemented in the Semiempirical Extended Tight-Binding (xTB) program package[22].

Calculations on protein-ligand systems

All calculations on protein-ligand systems used the second generation extended tight-binding (GFN2-xTB) method[21] as implemented in the Semiempirical Extended Tight-Binding (xTB) program package[22]. Generalized Born, solvent-accesible area (GBSA) solvation was used to mimic an aqueous environment for all calculations.

The truncated TNKS2 protein was constructed from the the ligands/protein-structure obtained from Ref. [23]. The entire protein was minimized using AmberTools, using the Generalized Born implicit, igb=5), the Amber 14 force field[24], and the general AMBER force field (GAFF)[25] for ligands, assigned using Antechamber from AmberTools[26]. Following minimization, truncation and capping of the terminals were carried out using PyMol [27]. Truncation was performed to remove all protein atoms beyond 3-4Å around the ligand. Truncated ends were capped using ACE/NME terminal patches. The ligand bound to the protein is one of the many inhibitors identified in Ref. [28] whose structure is available in the Protein Data Bank [29](PDB: JKN).

The structure of the JE-2147-HIV protease complex was obtained from PDB 1KZK. Hydrogens were added using UCSF Chimera[30] and the structure was

optimized first using the GFN-FF force field as implemented in the xTB program package[22] and finally with the GFN2-xTB method[21] ultimately used for harmonic vibrational analysis.

Supplementary information. Supplementary figure comparing stochastic Lanczos and Chebyshev methods.

Acknowledgments. We thank S. Murlidaran for help with protein preparation.

Declarations

- Funding. Work by Alec F. White was supported by the US Department of Energy, via grant no. DE-SC0018140. Work by Chenghan Li was supported by the US National Science Foundation via grant no. 1931328.
- GKC is a part owner of QSimulate, Inc.
- Availability of data and materials: Data is available from the authors upon reasonable request.
- Code availability: Code is available from the authors upon reasonable request.
- Authors' contributions: AFW and GKC conceived the project. AFW, CL carried out the work. All authors contributed to the writing of the paper.

References

[1] Grimme, S., Schreiner, P.R.: Computational chemistry: the fate of current methods and future challenges. Angewandte Chemie International Edition **57**(16), 4170–4176 (2018)

[2] Li, H., Jensen, J.H.: Partial Hessian vibrational analysis: The localization of the molecular vibrational energy and entropy. Theoretical Chemistry Accounts **107**(4), 211–219 (2002). https://doi.org/10.1007/s00214-001-0317-7

[3] Woodcock, H.L., Zheng, W., Ghysels, A., Shao, Y., Kong, J., Brooks, B.R.: Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. Journal of Chemical Physics **129**(21) (2008). https://doi.org/10.1063/1.3013558

[4] Filippone, F., Parrinello, M.: Vibrational analysis from linear response theory. Chemical Physics Letters **345**(1-2), 179–182 (2001). https://doi.org/10.1016/S0009-2614(01)00843-0

[5] Karplus, M., Kushick, J.N.: Method for Estimating the Configurational Entropy of Macromolecules. Macromolecules **14**(2), 325–332 (1981). https://doi.org/10.1021/ma50003a019

[6] Brooks, B.R., Janezic, D., Karplus, M.: Harmonic Analysis of Large Systems. I. Methodology. Journal of Computational Chemistry **16**(12), 1522–1542 (1995)

[7] Ubaru, S., Chen, J., Saad, Y.: Fast estimation of $\mathrm{tr}(f(A))$ via stochastic Lanczos quadrature. SIAM Journal on Matrix Analysis and Applications **38**(4), 1075–1099 (2017)

[8] Hellman, O., Steneteg, P., Abrikosov, I.A., Simak, S.I.: Temperature dependent effective potential method for accurate free energy calculations of solids. Physical Review B - Condensed Matter and Materials Physics **87**(10), 1–8 (2013) arXiv:1303.1145. https://doi.org/10.1103/PhysRevB.87.104111

[9] Errea, I., Calandra, M., Mauri, F.: Anharmonic free energies and phonon dispersions from the stochastic self-consistent harmonic approximation: Application to platinum and palladium hydrides. Physical Review B - Condensed Matter and Materials Physics **89**(6), 1–16 (2014) arXiv:1311.3083. https://doi.org/10.1103/PhysRevB.89.064302

[10] Baer, R., Neuhauser, D., Rabani, E.: Self-averaging stochastic kohn-sham density-functional theory. Physical Review Letters **111**(10), 1–5 (2013). https://doi.org/10.1103/PhysRevLett.111.106402

[11] Han, I., Malioutov, D., Shin, J.: Large-scale log-determinant computation through stochastic chebyshev expansions. In: International Conference on Machine Learning, pp. 908–917 (2015). PMLR

[12] Perdew, J.P., Burke, K., Ernzerhof, M.: Generalized gradient approximation made simple. Physical review letters **77**(18), 3865 (1996)

[13] Grimme, S., Bannwarth, C., Shushkov, P.: A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (z=1–86). J. Chem. Theory Comput. **13**(5), 1989–2009 (2017). https://doi.org/10.1021/acs.jctc.7b00118

[14] Ehrlich, S., Göller, A.H., Grimme, S.: Towards full quantum-mechanics-based protein–ligand binding affinities. ChemPhysChem **18**(8), 898–905 (2017)

[15] Humphrey, W., Dalke, A., Schulten, K.: Vmd: visual molecular dynamics. Journal of molecular graphics **14**(1), 33–38 (1996)

[16] Reiling, K.K., Endres, N.F., Dauber, D.S., Craik, C.S., Stroud, R.M.: Anisotropic dynamics of the JE-2147-HIV protease complex: Drug resistance and thermodynamic binding mode examined in a 1.09 Å structure. Biochemistry **41**(14), 4582–4594 (2002). https://doi.org/10.1021/bi011781z

[17] Mardirossian, N., Wang, Y., Pearlman, D.A., Chan, G.K., Shiozaki, T.: Novel algorithms and high-performance cloud computing enable efficient fully quantum mechanical protein-ligand scoring. arXiv preprint arXiv:2004.08725 (2020)

[18] Jaklič, J., Prelovšek, P.: Lanczos method for the calculation of finite-temperature quantities in correlated systems. Physical Review B **49**(7), 5065 (1994)

[19] Neese, F.: The ORCA program system. Wiley Interdisciplinary Reviews: Computational Molecular Science **2**(1), 73–78 (2012). https://doi.org/10.1002/wcms.81

[20] Neese, F.: Software update: the ORCA program system, version 4.0. Wiley Interdisciplinary Reviews: Computational Molecular Science **8**(1), 4–9 (2018). https://doi.org/10.1002/wcms.1327

[21] Bannwarth, C., Ehlert, S., Grimme, S.: GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. Journal of Chemical Theory and Computation **15**(3), 1652–1671 (2019). https://doi.org/10.1021/acs.jctc.8b01176

[22] Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., Grimme, S.: Extended tight-binding quantum chemistry methods. Wiley Interdisciplinary Reviews: Computational Molecular Science **11**(2), 1–49 (2021). https://doi.org/10.1002/wcms.1493

[23] Schindler, C.E.M., Baumann, H., Blum, A., Böse, D., Buchstaller, H.P., Burgdorf, L., Cappel, D., Chekler, E., Czodrowski, P., Dorsch, D., Eguida, M.K.I., Follows, B., Fuchß, T., Grädler, U., Gunera, J., Johnson, T., Lebrun, C.J., Karra, S., Klein, M., Knehans, T., Koetzner, L., Krier, M., Leiendecker, M., Leuthner, B., Li, L., Mochalkin, I., Musil, D., Neagu, C., Rippmann, F., Schiemann, K., Schulz, R., Steinbrecher, T., Tanzer, E.M., Lopez, A.U., Follis, A.V., Wegener, A., Kuhn, D.: Large-scale assessment of binding free energy calculations in active drug discovery projects. Journal of Chemical Information and Modeling **60**(11), 5457–5474 (2020). https://doi.org/10.1021/acs.jcim.0c00900

[24] Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., Simmerling, C.: ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of Chemical Theory and Computation **11**(8), 3696–3713 (2015). https://doi.org/10.1021/acs.jctc.5b00255

[25] Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A.: Development and testing of a general Amber force field. Journal of Computational Chemistry **25**(9), 1157–1174 (2004). https://doi.org/10.1002/jcc.20035

[26] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C., Kollman, P.A.: Amber 2021, Univeristy

of California, San Francisco (2021). https://ambermd.org/doc12/Amber21.pdf

[27] Schrödinger, L.: The PyMOL molecular graphics system, version 1.8. November (2015)

[28] Waaler, J., Leenders, R.G.G., Sowa, S.T., Alam Brinch, S., Lycke, M., Nieczypor, P., Aertssen, S., Murthy, S., Galera-Prat, A., Damen, E., Wegert, A., Nazaré, M., Lehtiö, L., Krauss, S.: Preclinical Lead Optimization of a 1,2,4-Triazole Based Tankyrase Inhibitor. Journal of Medicinal Chemistry **63**(13), 6834–6846 (2020). https://doi.org/10.1021/acs.jmedchem.0c00208

[29] Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C.: The protein data bank. Acta Crystallographica Section D: Biological Crystallography **58**(6 I), 899–907 (2002). https://doi.org/10.1107/S0907444902003451

[30] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E.: UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry **25**(13), 1605–1612 (2004). https://doi.org/10.1002/jcc.20084

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIforStochasticvibrationalthermodynamics.pdf
- stochasticthermorepo11.zip