SAMPLE SOLUTION

1 Concepts and Understanding (S2021.1)

Note that the answers given below are more detailed (for didactical purposes!) than what was actually expected from the students during the examination.

a. To assess the quality of a pseudo-random number generator producing real values over the interval $[0, 1[$, one should calculate two type of functions. First, the probability distribution (histogram) function $p(x)$ of the generated numbers x, which should ideally be homogeneous over the interval $[0, 1[$.

> ➢ Uniformity test (X^2 test):
>> ○ Idea: If M random numbers are generated, are these spread out reasonably of the interval?
>> ○ Divide the interval in B bins of equal length
>>
>> $$X^2 = \frac{1}{B} \sum_{i=1}^{B} \frac{(n_i - E_i)^2}{E_i}$$
>>
>> n_i: number of random numbers in bin i
>> E_i: expectation value of n in bin $i \rightarrow E_i = \frac{M}{B}$
>>
>> ○ If $X^2 \approx 0 \rightarrow$ agreement between expected and generated distribution
>> ○ Rule of thumb (for $M > 10B$): $\quad 1 - X^2 < \frac{2}{\sqrt{B}}$
>> ○ Repeat X^2 test a few times

Second, the autocorrelation (sequence correlation) function $c(\Delta k)$, where Δk indicates the interval between two generated numbers along the sequence, which should ideally be zero for any $\Delta k \neq 0$.

> ➢ Correlation test:
>> ○ Idea: Are there "time" correlations in the sequence of random numbers?
>> ○ Calculate the correlation function
>>
>> $$C(k) = \frac{\langle x_i x_{i+k} \rangle_i - \langle x_i \rangle_i \langle x_{i+k} \rangle_i}{\langle x_i x_i \rangle_i - \langle x_i \rangle_i \langle x_i \rangle_i} \qquad \text{with } \langle x_i x_{i+k} \rangle_i \equiv \frac{1}{B-k} \sum_{i=1}^{B-k} x_i x_{i+k}$$
>>
>> ○ If there is no correlation: $\langle x_i x_{i+k} \rangle_i \langle x_i \rangle_i \langle x_{i+k} \rangle_i \rightarrow C(k) = 0$

Briefly stated, the pseudo-random numbers should be homogeneously distributed over the interval $[0, 1[$ as well as uncorrelated along the sequence.

b. The correspondence principle associates classical-mechanical variables to quantum-mechanical operators. For a single particle in three dimensions, the operators associated with the classical position vector ($\mathbf{r}$), momentum vector ($\mathbf{p}$), and total energy (E) are

$$\mathbf{r} \rightarrow \mathbf{r} \ , \quad \mathbf{p} \rightarrow -i\hbar \nabla_{\mathbf{r}} \quad \text{and} \quad E \rightarrow i\hbar \frac{\partial}{\partial t} \ ,$$

where $\hbar = (2\pi)^{-1}h$, h being the Planck constant, and $\nabla_{\mathbf{r}}$ is the gradient operator. The classical energy conservation reads

$$\mathcal{H}(\mathbf{r}, \mathbf{p}) = \mathcal{K}(\mathbf{p}) + \mathcal{V}(\mathbf{r}) = E \ ,$$

where $\mathcal{H}$ is the Hamiltonian function, $\mathcal{K}$ the kinetic-energy function, $\mathcal{V}$ the potential-energy function, and E the total energy. The corresponding quantum-mechanical equation *via* the correspondence principle is the time-dependent Schrödinger equation, which reads

$$\hat{\mathcal{H}} \Psi = (\hat{\mathcal{K}} + \mathcal{V}) \Psi = i\hbar \frac{\partial}{\partial t} \Psi \ ,$$

where $\hat{\mathcal{H}}$ is the Hamiltonian operator, $\hat{\mathcal{K}}$ the kinetic-energy operator, and Ψ the wavefunction. Using the expression for the classical kinetic-energy function
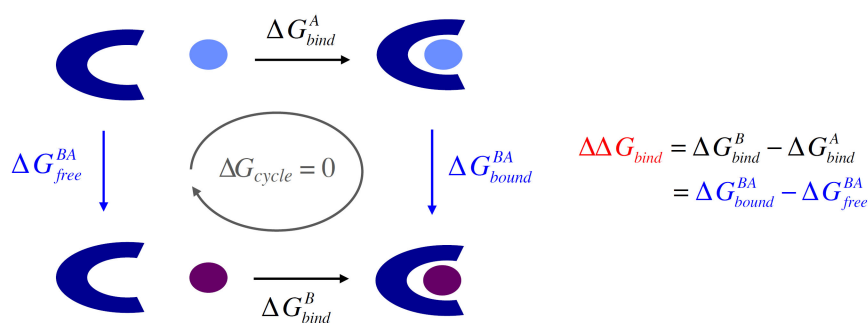
$$\mathcal{K} = \sum_i \frac{\mathbf{p}_i^2}{2m_i} \ ,$$

one may write the corresponding quantum-mechanical operator using the correspondence principle as

$$\hat{\mathcal{K}} = -\sum_i \frac{\hbar^2}{2m_i} \nabla^2 \ ,$$

where ∇^2 is the Laplacian operator.

c. The schematic drawing of the thermodynamic cycle, along with the equation for the relative binding free energy $\Delta G_{\text{bind}}^B - \Delta G_{\text{bind}}^A$, are shown below

The horizontal arrows correspond to the two conformational changes and the vertical ones to the two alchemical changes. The alchemical changes typically require much less simulation time to reach convergence up to a given accuracy compared to the conformational ones. This is because bringing a ligand from a far distance into the active site of a protein (including possible protein and ligand rearrangements) is generally much more difficult to sample than performing limited topology changes in an already bound ligand (smaller protein and ligand rearrangements). The alchemical free-energy differences can be calculated using *e.g.* thermodynamic integration (TI), free-energy perturbation (FEP), or enveloping distribution sampling (EDS). If the ligand state A is chosen to involve the ligand with its normal intramolecular interactions but no non-bonded interactions with the environment (*i.e.* a so-called dummy skeleton), then $\Delta G_{bind}^A = 0$ and the calculated relative binding free energy is equal to the absolute binding free energy ΔG_{bind}^B of ligand B.

2 Fundamental Equations (S2021.2)

Note that the answers given below are more detailed (for didactical purposes!) than what was actually expected from the students during the examination. However (unfortunately!), many students completely overlooked the requirements "**explain the meaning** of all the involved symbols" and "**state the SI units** of these quantities".

   a. The equation for the normalized probability density $p(x)$ corresponding to a normal (Gaussian) distribution of mean μ and standard deviation σ in a variable x reads

$$p(x) = \frac{1}{(2\pi)^{1/2}\sigma}\, e^{-(x-\mu)^2/(2\sigma^2)} \;.$$

The word "normalized" refers to the condition

$$\int_{-\infty}^{\infty} dx\, p(x) = 1$$

imposed to $p(x)$. It is a common mistake to think that a probability distribution is "unitless" or with "arbitrary units". If the distribution is normalized, the above normalization condition imposes that the product of the units of dx and $p(x)$ is unitless (like the resulting value of "one"). So, the units of $p(x)$ are inverse to the units of x.

   b. The Newtonian equation of motion reads

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = m\,\frac{d^2\mathbf{r}}{dt^2} \;.$$

Here, $\mathbf{F}$ is the force acting on the particle (SI unit: N = J·m^{-1}), t is time (SI unit: s), $\mathbf{r}$ is the position vector of he particle (SI unit: m) and $\mathbf{p}$ is the corresponding momentum vector (SI unit: kg·m·s^{-1} = J·s·m^{-1}). When considering a conservative force field, the force is given by the gradient of a potential-energy function $\mathcal{V}(\mathbf{r})$, *i.e.*

$$\mathbf{F} = -\nabla_{\mathbf{r}}\mathcal{V}(\mathbf{r}) \;.$$

The Newtonian equations of motion are only valid in a Cartesian coordinate system, *i.e.* $\mathbf{r}$ must be the Cartesian coordinate vector and $\mathbf{p}$ must be the Cartesian momentum vector, satisfying

$$\mathbf{p} = m\,\frac{d\mathbf{r}}{dt} \;.$$

The three types of quantities often conserved by the Newtonian dynamics (constants of the motion) are the total energy (conserved in the case of a conservative force field, *i.e.* when the force derives from a potential energy that solely depends of $\mathbf{r}$), the linear momentum of the center of mass (when the total force acting on the system is zero) and the angular momentum about the center of mass (when the total torque of the forces acting on the system is zero).

   c. An expression for the instantaneous observable $\mathcal{T}$ can be derived from the equipartition principle, which states that the average kinetic energy associated with each degree of freedom of a system at equilibrium is equal to $(1/2)k_B T$, where k_B is the Boltzmann constant and T the absolute temperature. So, a possible definition reads

$$\mathcal{T}(\mathbf{p}) = \frac{2}{k_B N_{\text{dof}}}\mathcal{K}(\mathbf{p}) \;,$$

where $\mathcal{K}$ is the kinetic energy and N_{dof} the total number of degrees of freedom of the system. This definition is not unique. For instance, we could use a subset of the degrees of freedom and the associated kinetic energy. The instantaneous variable $\mathcal{T}$ is connected to its canonical

ensemble average T *via*

$$T = \langle \mathcal{T} \rangle = \frac{\int d\mathbf{r} d\mathbf{p} \, \mathcal{T} \, e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})}}{\int d\mathbf{r} d\mathbf{p} \, e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})}}$$

where $\langle \dots \rangle$ indicates a canonical ensemble average and $\mathcal{H}$ is the Hamiltonian.

3 Derivations (S2021.3)

a. Since the coupling is linear, one has

$$\frac{\partial \mathcal{H}}{\partial \lambda} = \mathcal{H}_B - \mathcal{H}_A \ .$$

Thus, the TI formula reads

$$\Delta G_{A\to B} = \int_0^1 d\lambda \, \langle \frac{\partial \mathcal{H}}{\partial \lambda} \rangle_\lambda = \int_0^1 d\lambda \, [\langle \mathcal{H}_B \rangle_\lambda - \langle \mathcal{H}_A \rangle_\lambda] \ .$$

Inserting the fitted equations leads to the result

$$\Delta G_{A\to B} = \int_0^1 d\lambda \, [c\,\lambda + d - a\,\lambda - b] = d - b + (c-a)/2.$$

b. The suggested expansion gives

$$\frac{\partial}{\partial \beta} \left(\frac{1}{Z} \frac{\partial Z}{\partial \beta} \right) = \frac{1}{Z} \frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2} \left(\frac{\partial Z}{\partial \beta} \right)^2 \ .$$

Using the definition of Z, we easily derive

$$\frac{\partial Z}{\partial \beta} = - \int d\mathbf{x} \, \mathcal{H} \, e^{-\beta \mathcal{H}} = -Z \langle \mathcal{H} \rangle$$

and

$$\frac{\partial^2 Z}{\partial \beta^2} = \int d\mathbf{x} \, \mathcal{H}^2 \, e^{-\beta \mathcal{H}} = Z \langle \mathcal{H}^2 \rangle \ .$$

Inserting these expressions into the definition of σ_E^2, this gives

$$\sigma_E^2 = \frac{1}{Z} \frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2} \left(\frac{\partial Z}{\partial \beta} \right)^2 \ .$$

Using the above expansion, we get

$$\sigma_E^2 = \frac{\partial}{\partial \beta} \left(\frac{1}{Z} \frac{\partial Z}{\partial \beta} \right) \ .$$

Finally, inserting the expression for $\langle \mathcal{H} \rangle$, the definition of β, and the definition of C_V, one obtains

$$\sigma_E^2 = -\frac{\partial}{\partial \beta} \langle \mathcal{H} \rangle = k_B T^2 \frac{\partial}{\partial T} \langle \mathcal{H} \rangle = k_B T^2 C_V \ .$$

So, the energy fluctuations in the canonical ensemble are directly related to the isochoric heat capacity of the system.

4 Explicit Calculations (S2021.4)

a. Using the equation for the harmonic potential energy one has

$$V(b_o \pm \Delta b; b_o) = \frac{1}{2}k_b(\Delta b)^2.$$

Inserting the numerical values $k_b = 10^5$ kJ·mol·nm^{-2} and $\Delta b = 0.01$ nm, one gets $V = 5$ kJ·mol^{-1}. In comparison, the thermal energy RT at $T = 300$ K is 2.49 kJ·mol^{-1}. So, a bond-length change of 0.01 nm already corresponds to an energy increase by about $2RT$. As a result, the typical bond-length fluctuation in molecular dynamics simulations are on the order of 0.01 nm or less (RT would be $0.01/2^{1/2} \approx 0.007$ nm). Compared to the reference C-C bond length of 0.15 nm, this is only 6.7%.

5   Algorithms and Implementation (S2021.5)

a. A possible code for the function `MeanSquareDispl` is given below

```
void MeanSquareDispl ( int N, int K, double x[3*N][], double s[] ) {
    int k,kk,i;
    double dif,sum;
    for ( k = 0; k < K; k++ ) { // loop over lag times k*dt
        sum = 0;
        for ( kk = 0; kk < K-k; kk++ ) { // loop over time origins kk*dt
            for ( n = 0; n < 3*N; n++ ) { // loop over particles and components
                dif = r[n][k+kk] - r[n][k];
                sum += dif*dif;
            }
        }
        s[k] = sum/(N*(K-k));
    }
}
```

According to the Einstein equation, $s(\tau)$ should become linear at long times, its slope being connected to the diffusion constant D of the system as

$$\lim_{\tau \to \infty} \tau^{-1} s(\tau) = 6D \ .$$

In the absence of specific "cage effects", the curve should actually be linear already from zero (or very small) lag times onward. However, the number of time origins available to calculate $s(k\Delta t)$ for a given k is given by $K - k$. As a result, when τ approaches T, the statistics become poor and the curve becomes more and more "noisy". When using periodic boundary conditions, one should make sure that $\mathbf{r}_n(t)$ follows the particle n across periodic boundaries, *i.e.* that it is a continuous and unbounded trajectory within the infinite periodic system.