Metropolis Acceptance Criteria as Implemented in Cassandra

All Monte Carlo moves are implemented in Cassandra to preserve detailed balance between each pair of microstates m and n

$$\Pi_{mn}\ \alpha_{mn}\ p_m = \Pi_{nm}\ \alpha_{nm}\ p_n \tag{1}$$

where Π_{mn} is the probability of accepting the move from microstate m to microstate n, α_{mn} is the probability of attempting the move that will form n from m, and p_m is the probability of m in the ensemble of interest.

In Cassandra, detailed balance is enforced via the Metropolis criterion

$$\Pi_{mn} = \min\left(1, \frac{\alpha_{nm}}{\alpha_{mn}}\frac{p_n}{p_m}\right) \tag{2}$$

The ratio in Eq. (2) will often involve an exponential, e.g. $e^{-\beta\Delta U}$. To preserve precision in the energy calculation, the acceptance probability is computed

$$\Pi_{mn} = \exp\left\{-\max\left[0, \ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right)\right]\right\} \tag{3}$$

The logarithm, codenamed ln_pacc, is tested in accept_or_reject.f90 before being exponentiated. If ln_pacc is greater than 0 and less than a maximum numerical value, Π_{mn} is computed and compared to a random number.

Code 1: accept_or_reject.f90

```
47  accept = .FALSE.
48
49  IF (ln_pacc <= 0.0_DP) THEN
50
51     accept = .TRUE.
52
53  ELSE IF ( ln_pacc < max_kBT) THEN
54
55     pacc = DEXP(-ln_pacc)
56
57     IF ( rranf() <= pacc ) THEN
58
59        accept = .TRUE.
60
61     END IF
62
63  END IF
```

1 Canonical Monte Carlo

In the canonical ensemble, the number of molecules N, the volume V and temperature T are all constant. The position, orientation and conformation of a molecule containing M atoms is given by a $2M+1$-dimensional vector $\mathbf{q}$. The positions, orientations and conformations of all N molecules are denoted $\mathbf{q}^N$. The probability of observing microstate m with configuration $\mathbf{q}_m^N$ is

$$p_m = \frac{e^{-\beta U\left(\mathbf{q}_m^N\right)}}{Z(N,V,T)}\, d\mathbf{q}^N \tag{4}$$

where β is the inverse temperature $1/k_BT$, U is the potential energy, the differential volume $d\mathbf{q}^N$ is included to make p_m dimensionless and Z is the configurational partition function

$$Z(N,V,T) = \int e^{-\beta U(\mathbf{q}^N)} d\mathbf{q}^N \tag{5}$$

The ratio of microstate probabilities follows from Eq. (4)

$$\begin{aligned} \frac{p_n}{p_m} &= \frac{e^{-\beta U\left(\mathbf{q}_n^N\right)} d\mathbf{q}^N / Z(N,V,T)}{e^{-\beta U(\mathbf{q}_m^N)} d\mathbf{q}^N / Z(N,V,T)} \\ &= e^{-\beta(U_n - U_m)} = e^{-\beta \Delta U_{mn}} \end{aligned} \tag{6}$$

The configurational partition function Z and differential volume $d\mathbf{q}^N$ both cancel, leaving only the ratio of Boltzmann factors.

New configurations are generated by attempting moves that translate, rotate and regrow a randomly selected molecule.

1.1 Translating a Molecule

A molecule is translated by moving its center of mass in each Cartesian direction by a random amount chosen from the uniform distribution on the interval [-$\delta r_{max}, \delta r_{max}$]. The maximum displacement δr_{max} must be given in the input file. The translation move is symmetric in forward and reverse directions. That is, either microstate n can be formed from microstate m and vice versa by moving one molecule within δr_{max} in each Cartesian direction, or microstate n cannot be formed at all. As a result, $\alpha_{mn} = \alpha_{nm}$.

The acceptance probability for a translation move follows from Eq. (6)

Table 1: Variable symbols and code names for translating and rotating a molecule

Symbol	**Code name**
β	beta(this_box)
ΔU_{mn}	delta_e

$$\ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right) = \ln\left(\frac{p_m}{p_n}\right) = \beta\Delta U_{mn} \tag{7}$$

In Cassandra, the translation move is implemented in the subroutine Translate defined in translate.f90. The relevant lines from version 1.1 are quoted below. The variable names in the translate.f90 code are identified with the symbols from Eq. (7) in Table 1.

Code 2: translate.f90

```
274 ln_pacc = beta(this_box) * delta_e
275 accept = accept_or_reject(ln_pacc)
```

1.2 Rotating a Molecule

A linear molecule is rotated differently than a non-linear molecule. If the molecule is linear:

1. Pick three random angles: ϕ on $[-\pi, \pi]$, $\cos(\theta)$ on [-1,1], and ψ on $[-\pi, \pi]$.
2. With the origin at the molecule's center of mass, rotate by ϕ around z, rotate by θ around x', and rotate by ψ around z'', as shown in Fig. 1.

If the molecule is non-linear:

1. Randomly select an axis: x, y, or z.

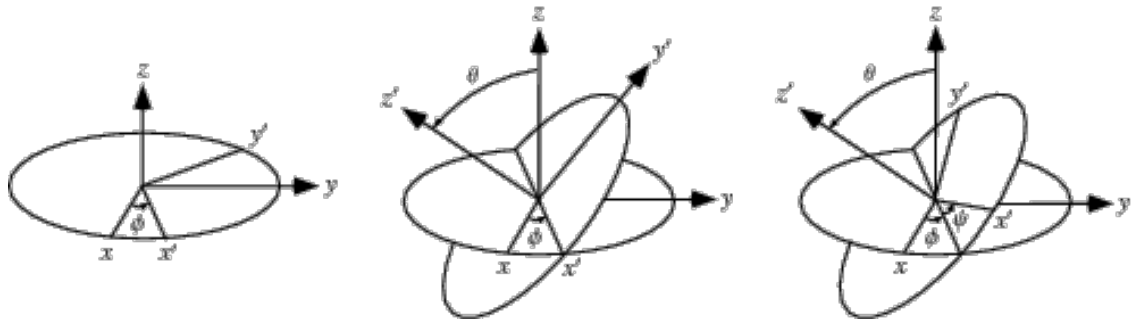

Figure 1: Procedure for rotating linear molecules.
Image from mathworld.wolfram.com/EulerAngles.html.

2. Choose a random angular displacement $\delta\theta$ from $[-\delta\theta_{max}, \delta\theta_{max}]$. $\delta\theta_{max}$ must be given in the input file.

3. Rotate the molecule around a vector parallel to the selected axis and through its center of mass by $\delta\theta$.

In either case, the rotation move is symmetric, $\alpha_{mn} = \alpha_{nm}$, and the acceptance criteria is given by Eq. (7). The rotation move is implemented in subroutine Rotate defined in rotate.f90.

Code 3: rotate.f90

```
261 ln_pacc = beta(this_box) * delta_e
262 accept = accept_or_reject(ln_pacc)
```

1.3 Regrowing a Molecule

Internal degrees of freedom in flexible molecules are sampled by cutting a bond and regrowing the severed fragments using Configurational Bias Monte Carlo (CBMC).

[SECTION INCOMPLETE]

1.4 Canonical Partition Function

In Sections 1.1-1.3, the microstate probability is normalized by the configuration partition function Z because the only relevant degrees of freedom are configurational. In other ensembles, the full partition function Q appears, integrated over both configuration space $\mathbf{q}^N$ and momenta space $\mathbf{p}_q^N$

$$Q(N,V,T) = \frac{1}{h^{N(2M+1)}N!}\int e^{-\beta H(\mathbf{p}_q^N,\mathbf{q}^N)}\; d\mathbf{p}_q^N d\mathbf{q}^N \tag{8}$$

where the $2M+1$ momenta $\mathbf{p}_q$ are conjugate to the generalized coordinates $\mathbf{q}$. The momenta and configuration integrals are separable, and the single molecule momenta integrals are all identical.

$$\begin{aligned} Q(N,V,T) &= \frac{1}{N!}\left[\int e^{-\beta U(\mathbf{q}^N)} d\mathbf{q}^N\right]\left[\frac{1}{h^{2M+1}}\int e^{-\beta K(\mathbf{p}_q)}\; d\mathbf{p}_q\right]^N \\ &= \frac{1}{N!}Z(N,V,T)\left[\frac{Q(1,V,T)}{Z(1,V,T)}\right]^N \end{aligned} \tag{9}$$

where $Q(1,V,T)$ is the partition function of a single molecule in a box. The center of mass integrals for a single molecule are separable from the integrals over rotational and internal degrees of freedom:

$$Q(1,V,T) = Q_{com}Q_{rot+int} = V\Lambda^{-3}Q_{rot+int} \tag{10}$$

where Λ is the de Broglie wavelength of the molecule and $Q_{rot+int}$ equals 1 for a point particle. In general, the rotational and internal momenta are not separable since the moments of inertia will depend on the conformation adopted by the molecule. The configurational

partition function is further separable into center of mass (translational), orientational and internal degrees of freedom:

$$Z(1, V, T) = Z_{com} Z_{rot} Z_{int} \tag{11}$$

where Z_{com} equals the volume V and Z_{rot} equals 4π for a linear molecule and $8\pi^2$ for a non-linear molecule.

2 Isothermal-Isobaric Monte Carlo

In the isothermal-isobaric ensemble, the number of particles N, the pressure P and temperature T are all constant while the volume V and energy E fluctuate. The partition function is

$$\Delta(N, P, T) = \int e^{-\beta PV} Q(N, V, T) dV \tag{12}$$

Note that Q is dimensionless and Δ has dimensions of volume. The probability of the system having volume V is

$$p(V) = \frac{Q(N, V, T) e^{-\beta PV}}{\Delta(N, P, T)} \tag{13}$$

The probability of observing microstate m with configuration $\mathbf{q}_m^N$ and volume V_m is

$$
\begin{aligned}
p_m &= \frac{e^{-\beta U(\mathbf{q}_m^N)} d\mathbf{q}^N}{Z(N,V,T)} \; \frac{Q(N,V,T) e^{-\beta P V_m} dV}{\Delta(N,P,T)} \\
&= \frac{e^{-\beta U_m - \beta P V_m}}{\Delta(N,P,T)} \; \left(\frac{Q(1,V,T)}{Z(1,V,T)} \, d\mathbf{q} \right)^N dV
\end{aligned} \tag{14}
$$

All terms but the exponentials cancel out of the ratio of microstate probabilities

$$
\frac{p_n}{p_m} = e^{-\beta(U_n - U_m) - \beta P (V_n - V_m)} = e^{-\beta \Delta U_{mn} - \beta P \Delta V_{mn}} \tag{15}
$$

2.1 Scaling the Volume

In Cassandra, new volumes are sampled as follows:

1. Pick a random volume ΔV from the interval $[-\delta V_{max}, \, \delta V_{max}]$. The trial volume is $V + \Delta V$.
2. Scale the box lengths uniformly.
3. Scale the center of mass of each molecule uniformly.

The probability α_{mn} is the probability of selecting ΔV times the effect of scaling the configurations. Scaling the configurations changes the differential element $d\mathbf{q}^N$ surrounding configuration $\mathbf{q}_m^N$. Only the molecular centers of mass change, so we can separate $d\mathbf{q}$ into 3 center of mass coordinates $d\mathbf{r}_{com}$ and $2M$-2 orientational and internal coordinates $d\mathbf{q}_{rot+int}$. The scaled center of mass positions are held constant, making $d\mathbf{r}_{com} = V d\mathbf{s}_{com}$.

$$
\frac{\alpha_{nm}}{\alpha_{mn}} = \frac{dV/2\delta V_{max}}{dV/2\delta V_{max}} \; \frac{d\mathbf{q}_n^N}{d\mathbf{q}_m^N} = \left(\frac{V_n}{V_m} \right)^N \tag{16}
$$

Table 2: Variable symbols and code names for volume scaling move.

Symbol	**Code name**
β	beta(this_box)
ΔU_{mn}	delta_e
P	pressure(this_box)
ΔV	delta_volume
N	total_molecules
V_n	box_list(this_box)%volume
V_m	box_list_old%volume

The acceptance probability for a volume scaling move is

$$\ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right) = \beta\Delta U_{mn} + \beta P\Delta V_{mn} - N\ln(V_n/V_m) \tag{17}$$

The volume scaling move is implemented in subroutine Volume_Change defined in volume_change.f90.

Code 4: volume_change.f90

```
475 ln_pacc = beta(this_box) * delta_e &
476         + beta(this_box) * pressure(this_box) * delta_volume &
477         - total_molecules *
               DLOG(box_list(this_box)%volume/box_list_old%volume)
478 accept = accept_or_reject(ln_pacc)
```

3 Grand Canonical Monte Carlo

In the grand canonical ensemble, the chemical potential μ, the volume V and temperature T are held constant while the number of molecules N and energy E fluctuate. The partition function is

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} Q(N, V, T)\ e^{\beta\mu N} \tag{18}$$

The probability of the system having N molecules is

$$p(N) = \frac{Q(N, V, T)e^{\beta\mu N}}{\Xi(\mu, V, T)} \tag{19}$$

The probability of observing microstate m with configuration $\mathbf{q}_m^{N_m}$ and number of molecules N_m is

$$\begin{aligned} p_m &= \frac{e^{-\beta U(\mathbf{q}_m^{N_m})} d\mathbf{q}^{N_m}}{Z(N, V, T)} \ \frac{Q(N, V, T)e^{\beta\mu N_m}}{\Xi(\mu, V, T)} \\ &= \frac{e^{-\beta U_m + \beta\mu N_m}}{\Xi(\mu, V, T)} \ \left[\frac{Q(1, V, T)}{Z(1, V, T)}\ d\mathbf{q}\right]^{N_m} \end{aligned} \tag{20}$$

Note that Eq. (20) does not contain the factorial $N_m!$ that accounts for indistinguishable particles. In a simulation, particles *are* distinguishable: they are numbered and specific particles are picked for MC moves. The ratio of microstate probabilities is

$$\frac{p_n}{p_m} = e^{-\beta\Delta U_{mn} + \beta\mu\Delta N_{mn}} \ \left[\frac{Q(1, V, T)}{Z(1, V, T)}\ d\mathbf{q}\right]^{\Delta N_{mn}} \tag{21}$$

Fluctuations in the number of molecules are achieved by inserting and deleting molecules. A successful insertion increases the number of molecules from N to $N + 1$, i.e. $\Delta N_{mn} = 1$. A successful deletion decreases the number of molecules from N to N - 1, i.e. $\Delta N_{mn} = -1$.

3.1 Inserting a Molecule Randomly

To insert a molecule, a position, orientation and molecular conformation must each be selected. The probability of inserting the new molecule at a random location is $d\mathbf{r}/V$, where $d\mathbf{r}$ is a Cartesian volume element of a single atom. The probability of choosing the molecule orientation is $d\mathbf{q}_{rot}/Z_{rot}$. For a linear molecule $\mathbf{q}_{rot}$ is 2-dimensional and Z_{rot} is 4π. For a nonlinear molecule $\mathbf{q}_{rot}$ is 3-dimensional and Z_{rot} is $8\pi^2$. The probability of the molecule conformation only depends on the remaining internal angles and dihedrals $\mathbf{q}_{int}$. A thermal ensemble of configurations is Boltzmann distributed $e^{-\beta U(\mathbf{q}_{int})}/Z_{int}$. The overall probability α_{mn} is

$$\alpha_{mn} = \frac{d\mathbf{r}}{V} \frac{d\mathbf{q}_{rot}}{Z_{rot}} \frac{e^{-\beta U(\mathbf{q}_{int})}}{Z_{int}} d\mathbf{q}_{int} = \frac{e^{-\beta U(\mathbf{q})}}{Z(1,V,T)} d\mathbf{q}. \tag{22}$$

where we have used Eq. (11) to recover $Z(1, V, T)$. For a point particle with no rotational or internal degrees of freedom, α_{mn} reduces to $d\mathbf{r}/V$. For molecules with internal flexibility, a library of configurations distributed according to $e^{-\beta U(\mathbf{q}_{int})}/Z_{int}$ can be generated from a single molecule MC simulation. In the reverse move, 1 of the $N + 1$ particles is randomly selected for deletion. The probability α_{nm} of picking the molecule we just inserted is

$$\alpha_{nm} = \frac{1}{N+1} \tag{23}$$

The acceptance probability for a random insertion move is

$$\ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right) = \beta\left[\Delta U_{mn} - U(\mathbf{q}_{N+1})\right] - \beta\mu + \ln\left(\frac{N+1}{Q(1,V,T)}\right) \tag{24}$$

where $U(\mathbf{q}_{N+1})$ is the intramolecular potential energy of the inserted molecule. $Q(1,V,T)$ is typically not known *a priori*, nor is it easily estimated. Substituting Eq. (10) into Eq. (24) and absorbing $Q_{rot+int}$ into a shifted chemical potential μ'

$$\mu' = \mu - k_B T \ln(Q_{rot+int}) \tag{25}$$

gives the acceptance criteria for inserting a molecule

$$\ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right) = \beta\left[\Delta U_{mn} - U(\mathbf{q}_{N+1})\right] - \beta\mu' + \ln\left(\frac{(N+1)\Lambda^3}{V}\right). \tag{26}$$

The terms absorbed into μ' are intensive and therefore GCMC simulations using Eq. (26) will converge to a specific average density. However, the value of μ' that corresponds to the converged density will *not* match tabulated values of μ computed from experimental data.

Alternatively, Eq. (24) can be recast to use the fugacity f instead of the chemical potential μ. The relationship between μ and f is

$$\mu = -k_B T \ln\left(\frac{Q(1,V,T)}{N}\right) = k_B T \ \ln\left(\frac{\beta f V}{Q(1,V,T)}\right) \tag{27}$$

Substituting Eq. (27) into Eq. (24) gives

$$\ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right) = \beta\left[\Delta U_{mn} - U(\mathbf{q}_{N+1})\right] + \ln\left(\frac{N+1}{\beta f V}\right) \tag{28}$$

where no terms have been absorbed into the fugacity f. Note also that the partition function has completely been eliminated from the acceptance criteria.

3.2 Deleting a Molecule Inserted Randomly

The probability α_{mn} of choosing a molecule to delete is

$$\alpha_{mn} = \frac{1}{N} \tag{29}$$

The probability α_{nm} of inserting that molecule back in is

$$\alpha_{nm} = \frac{e^{-\beta U(\mathbf{q})}}{Z(1,V,T)}\, d\mathbf{q} \tag{30}$$

The acceptance probability for deleting a molecule inserted randomly is

$$\ln\left(\frac{\alpha_{mn}}{\alpha_{nm}}\frac{p_m}{p_n}\right) = \beta\left[\Delta U_{mn} + U(\mathbf{q}_N)\right] + \beta\mu' + \ln\left(\frac{V}{N\Lambda^3}\right) \tag{31}$$

$$= \beta\left[\Delta U_{mn} + U(\mathbf{q}_N)\right] + \ln\left(\frac{\beta f V}{N}\right) \tag{32}$$

Note that in ΔU_{mn} is defined differently in Eqs. (26) and (28) than in Eqs. (31) and (32). In the former, the new configuration has more molecules, $\Delta U_{mn} = U(\mathbf{q}_n^{N+1}) - U(\mathbf{q}_m^N)$. In the latter, the new configuration has fewer molecules, $\Delta U_{mn} = U(\mathbf{q}_n^{N-1}) - U(\mathbf{q}_m^N)$.

Random insertions and deletions in the liquid phase typically have very high ΔU_{mn} due to core overlap and dangling bonds, respectively, making the probability of acceptance very low. Instead, insertions in Cassandra are attempted using Configurational Bias Monte Carlo.

3.3 Inserting a Molecule with Configurational Bias Monte Carlo

In Configurational Bias Monte Carlo (CBMC), the molecular conformation of the inserted molecule is molded to the insertion cavity. First, the molecule is parsed into fragments such that each fragment is composed of (a) a central atom and the atoms directly bonded to it, or (b) a ring of atoms and all the atoms directly bonded to them. Then, a position, orientation and molecular conformation of the molecule to be inserted are selected via the following steps:

1. Select which fragment will be placed first f_1 with uniform probability $1/N_{frag}$.
2. Select a fragment conformation with Boltzmann probability $e^{-\beta U(\mathbf{q}_{int,f_1})} d\mathbf{q}_{int,f_1}/Z_{int,f_1}$, where $\mathbf{q}_{int,f_1}$ are the internal degrees of freedom associated with fragment f_1.
3. Select an orientation with uniform probability $d\mathbf{q}_{rot}/Z_{rot}$.
4. Select a coordinate for the center of mass (COM) of fragment f_1:
   (a) Select κ_{ins} trial coordinates, each with uniform probability $d\mathbf{r}/V$. Since one of the trial coordinates will be selected later, the individual probabilities are additive. The probability of the collection of trial coordinates is $\kappa_{ins} d\mathbf{r}/V$.
   (b) Select one of the trial coordinates with probability $e^{-\beta U(\mathbf{r}|\mathbf{q}_m^N)} / \sum_i e^{-\beta U(\mathbf{r}_i|\mathbf{q}_m^N)}$.
5. For each additional fragment f_j:
   (a) Select a fragment conformation with Boltzmann probability $e^{-\beta U(\mathbf{q}_{int,f_j})} d\mathbf{q}_{int,f_j}/Z_{int,f_j}$
   (b) Select κ_{dih} trial dihedrals, each with uniform probability $d\phi/2\pi$. The probability of the collection of trial dihedrals is $\kappa_{dih} d\phi/2\pi$.
   (c) Select one of the trial dihedrals with probability $e^{-\beta U(\phi|\mathbf{r},\mathbf{q}_m^N)} / \sum_i e^{-\beta U(\phi_i|\mathbf{r},\mathbf{q}_m^N)}$.

The overall probability α_{mn} of attempting the insertion with the selected position, orientation and conformation is

$$\alpha_{mn} = \frac{1}{N_{frag}} \frac{d\mathbf{q}_{rot}}{Z_{rot}} \frac{\kappa_{ins} d\mathbf{r}}{V} \frac{e^{-\beta U(\mathbf{r}|\mathbf{q}_m^N)}}{\sum_i e^{-\beta U(\mathbf{r}_i|\mathbf{q}_m^N)}} \times \left[\prod_{i=1}^{N_{frag}} \frac{e^{-\beta U(\mathbf{q}_{int,f_i})} d\mathbf{q}_{int,f_i}}{Z_{int,f_i}} \right] \left[\prod_{j=1}^{N_{frag}-1} \frac{\kappa_{dih} d\phi}{2\pi} \frac{e^{-\beta U(\phi|\mathbf{r},\mathbf{q}_m^N)}}{\sum_i e^{-\beta U(\phi_i|\mathbf{r},\mathbf{q}_m^N)}} \right] \tag{33}$$

In Cassandra, the insertion move is implemented in the subroutine insertion.f90. The relevant lines from version 1.1 are quoted below. The variable names in the insertion.f90 code are identified with symbols in Table 3.

Code 5: insertion.f90

```
321     ln_pacc = beta(this_box)*delta_e

324   !Note that in insertion P_reverse equals 1
325   ln_pacc = ln_pacc + DLOG(P_forward) +
         DLOG(REAL(nmols(is,this_box)+1,DP))
326
327   IF(lchempot) THEN
328      ! chemical potential is input
329      ln_pacc = ln_pacc - species_list(is)%chem_potential *
            beta(this_box) &
330         - DLOG(box_list(this_box)%volume) &
331         + 3.0_DP*DLOG(species_list(is)%de_broglie(this_box))
332   ELSE
333      ! need to come back and look at it carefully.
334      ln_pacc = ln_pacc - DLOG(species_list(is)%fugacity) &
335                        - DLOG(beta(this_box)) &
336                        - DLOG(box_list(this_box)%volume) &
337                        + DLOG(species_list(is)%zig_by_omega)
338   END IF
339
340
341   accept = accept_or_reject(ln_pacc)
```

Table 3: Variable symbols and code names for inserting a molecule

Symbol	**Code name**
β	beta(this_box)
ΔU	delta_e
α_{mn}	P_forward
α_{nm}	P_reverse
μ	species_list(is)%chem_potential
N	nmols(is,this_box)
V	box_list(this_box)%volume
Z_{ig}/Ω	species_list(is)%zig_by_omega
Λ	species_list(is)%de_broglie(this_box)
f	species_list(is)%fugacity