

分子モデリング

山内 仁喬

2021 年 10 月 31 日

1 濃度換算

分子モデリングをする際には、実験で使用されたイオン濃度をシミュレーションでも再現することが多い。その際に、何個のイオン原子を含めるかを計算する必要がある。この章では、濃度換算についてまとめる。

1.1 モル濃度

モル濃度の定義は以下の通りである。

- 単位体積の溶液中の溶質の物質質量
- SI 単位系で mol/m^3

モル濃度をモラー ($M = \text{mol/L}$) へ単位変換は以下の通りである。

$$\text{mol/m}^3 = 10^{-3} \text{ mol/dm}^3 \quad (1)$$

$$= 10^{-3} \text{ mol/L} \quad (2)$$

$$= 10^{-3} M \quad (3)$$

$$= 1 \text{ mM} \quad (4)$$

1.1.1 例 1: 2.00 mol/L の NaCl 水溶液を 100 mL 調整する

NaCl のモル濃度は 58.4 g/mol である。必要な NaCl は

$$2.00 \text{ (mol/L)} \times 58.4 \text{ (g/mol)} = 116.8 \text{ (g/L)} \quad (5)$$

$$116.8 \text{ (g/L)} = 0.1168 \text{ (g/10}^{-3}\text{L)} = 0.1168 \text{ (g/mL)} \quad (6)$$

$$0.1168 \text{ (g/mL)} \times 100 \text{ (mL)} = 11.7 \text{ (g)} \quad (7)$$

故に 11.7 g の NaCl を 100mL になるまで水を足せば、2.00 mol/L の NaCl 水溶液が完成する。

1.1.2 例 2: NaCl 水溶液における NaCl の質量分率, 体積, モル濃度の計算

100 mL の水に 11.6 g の NaCl が溶解している。溶液の密度は 1.07 g/mL, 水の密度を 1.00 g/mL とする。この時, NaCl の質量分率は

$$\frac{11.6 \text{ (g)}}{11.6 \text{ (g)} + 100 \text{ (g)}} \times 100\% = 10.5\% \quad (8)$$

一方で水 (H_2O) の質量分率は

$$\frac{100 \text{ (g)}}{11.6 \text{ (g)} + 100 \text{ (g)}} \times 100\% = 89.6\% \quad (9)$$

である。溶液の密度から溶液の体積は、

$$\frac{11.6 \text{ (g)} + 100 \text{ (g)}}{1.07 \text{ (g/mL)}} = 104 \text{ (mL)} \quad (10)$$

NaCl のモル濃度は

$$11.6 \text{ (g)} \times \frac{1}{58.4 \text{ (g/mol)}} \times \frac{1}{104 \text{ (mL)}} \times 1000 = 1.91 \text{ M} \quad (11)$$

1.1.3 例 3: 水溶液に含まれているイオンの数を換算する

一辺 100 \AA の箱に NaCl が $150 \text{ mM} = 150 \text{ mol/m}^3$ 入っているとす。この時に、NaCl がいくつ含まれているかを数える。まず体積を SI 単位系で表す。

$$100 \text{ \AA} = 100 \times 10^{-10} \text{ m} = 10^{-8} \text{ m} \quad (12)$$

だから、

$$(100 \text{ \AA})^3 = (10^{-8} \text{ m})^3 = 10^{-24} \text{ m}^3 \quad (13)$$

150 mol/m^3 の濃度の時、1 立方メートルあたりに含まれる NaCl のイオンの数は、

$$150 \text{ (mol/m}^3\text{)} = 150 \times 6.0 \times 10^{23} \text{ (m}^{-3}\text{)} \quad (14)$$

$$= 900 \times 10^{23} \text{ (m}^{-3}\text{)} \quad (15)$$

$$= 9 \times 10^{25} \text{ (m}^{-3}\text{)} \quad (16)$$

したがって、一辺が 100 \AA の箱に含まれる NaCl の数は、

$$9 \times 10^{25} \text{ (m}^{-3}\text{)} \times 10^{-24} \text{ (m}^3\text{)} = 90 \quad (17)$$

と計算される。つまり、90 個の NaCl が含まれている。

1.1.4 例 4: 体積が $V \text{ (\AA}^3\text{)}$ の箱にモル濃度 $x \text{ mM}$ の NaCl を入れる

単位を mM から mol/m^3 に変換すると、

$$x \text{ (mM)} = x \text{ (mol/m}^3\text{)} \quad (18)$$

体積の単位を SI 単位系に直すと、

$$V \text{ (\AA}^3\text{)} = V \times 10^{-30} \text{ (m}^3\text{)} \quad (19)$$

したがって、

$$x \text{ (mM)} = x \text{ (mol/m}^3\text{)} \quad (20)$$

$$= x \times 6.022 \times 10^{23} \text{ (mol}^{-1}\text{)}(\text{mol/m}^3\text{)} \quad (21)$$

$$= x \times 6.022 \times 10^{23} \text{ (m}^{-3}\text{)} \quad (22)$$

$$= 6.022x \times 10^{23} \text{ (m}^{-3}\text{)} \quad (23)$$

よって、体積が $V \text{ (\AA}^3\text{)}$ の箱に含まれるイオンの数は、

$$6.022x \times 10^{23} \times V \times 10^{-30} = 6.022xV \times 10^{-7} \text{ (個)} \quad (24)$$

と計算される。

2 水の初期配置について

常温常圧における水の密度に合わせて水を配置するには、どのような間隔で並べればいいのかを考える。水の質量、常温常圧での密度は

- 水の質量 = 18 (g/mol)
- 水の密度 = 998.233 (kg/m³) \simeq 1.0 (g/cm³)
- アボガドロ定数 = $6.02214086 \times 10^{23}$ (mol⁻¹)

である。1 cm = 1.0×10^8 Å であるので、

$$1.0 \text{ (g/cm}^3\text{)} = \frac{6.02214086 \times 10^{23}}{18} \text{ (cm}^{-3}\text{)} \quad (25)$$

$$= \frac{6.02214086 \times 10^{23}}{18} \times \frac{1}{10^{24}} \text{ (Å}^{-3}\text{)} \quad (26)$$

$$= 0.334 \times \frac{1}{10} \text{ (Å}^{-3}\text{)} \quad (27)$$

$$= 0.0334 \text{ (Å}^{-3}\text{)} \quad (28)$$

すなわち、1 Å³ に 0.0334 個の水分子が存在するような密度であるので、1 Å に 0.3220 個の間隔で水分子を置けばよいという計算となる。よって、3.104 Å に 1 個の水を置けば良い。

3 一般化螺旋集合 (GSS: Generalized Spiral Set)

ミセルの初期構造を配置するなど、任意の点数を球面上にできるだけ等間隔にプロットするためのアルゴリズムを考える。これを実現する一つの方法として、螺旋を球面状に射影することが挙げられる。このような点の集合を一般化螺旋集合という。

3.1 アルゴリズム

区間 $[-1, 1]$ を $(N - 1)$ 等分した離散パラメータ h は自然数 $0 \leq k \leq N - 1$ を用いれば、

$$h_k = -1 + 2 \frac{k}{N - 1} \quad (29)$$

と書くことができる。一般化螺旋集合は以下のような漸化式で表される偏角を持つような点の集合である。

$$\theta_k = \arccos(h_k) \quad (30)$$

$$\phi_0 = 0 \quad (31)$$

$$\phi_{k+1} = \phi_k + \frac{3.6}{\sqrt{N}} \frac{1}{\sqrt{1 - h_k^2}} \quad (32)$$

$h_k = \pm 1$, つまり $k = 0, N - 1$ のとき ϕ_{k+1} が発散してしまうのに注意。

4 RESPAC: 粗視化粒子に小数電荷を割り当てるアルゴリズム

粗視化モデルを使用した場合、分子の持つ電荷は粗視化粒子の座標点に置かれることがあるが、このような電荷の置き方はしばしば適切ではない。例えば、アミノ酸 1 残基を 1 粒子に粗視化して粒子の代表点を C_α 原子上に定めた場合、 ± 1 or 0 の電荷を C_α 原子上に置くことになる。しかし、荷電アミノ酸は側鎖の先端に電荷を持っているため、 C_α 原子上に電荷を置くような取り扱いは実際の分子描像と解離がある。このような問題点を解決するためのアルゴリズムとして RESPAC [1] が提案されている。RESPAC では全原子レベルの部分電荷によって計算された静電ポテンシャルを再現するように粗視化粒子の部分電荷を決定する。量子化学計算で得られた静電ポテンシャルを再現するように全原子の部分電荷を決定するアルゴリズム RESP [2, 3] に由来して、RESPAC と名前がつけられている。

4.1 RESPAC の流れ

RESPAC の一連の流れは次の通りである:

1. 全原子モデル構造の各原子に対して部分電荷を割り振る (PDB2PQR)
2. 全原子の部分電荷に基づいて静電ポテンシャルを計算する (APBS)
3. 粗視化構造について、蛋白質表面のアミノ酸残基を特定する (surface)
4. 蛋白質表面のアミノ酸残基について、全原子モデルから得られた静電ポテンシャルを再現するように、最適な粗視化粒子の部分電荷を決定する (pdc)

4.2 RESPAC の理論

4.2.1 全原子モデルの部分電荷が作り出す静電ポテンシャルを計算する

連続誘電体モデルにおいて、全原子モデルの部分電荷が作り出す静電ポテンシャル ϕ_{ref} は Poisson Boltzmann 方程式

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi_{\text{ref}}(\mathbf{r})] - \epsilon(\mathbf{r}) \kappa \sinh[\phi_{\text{ref}}(\mathbf{r})] + \frac{4\pi\rho(\mathbf{r})}{k_B T} = 0 \quad (33)$$

を解くことで得られる。ここで、 \mathbf{r} は分子の周りのグリッド点の位置ベクトル、 $\epsilon(\mathbf{r})$ は位置に依存した誘電率、 κ はデバイ長の逆数、 $\rho(\mathbf{r})$ は原子電荷密度、 k_B はボルツマン定数、 T は温度である。

4.2.2 粗視化粒子が作り出す静電ポテンシャル

Debye-Hückel 近似を適用している場合、粗視化粒子による静電ポテンシャルは

$$\phi(\mathbf{r}) = \sum_{i=1}^N q_i \frac{\exp(-\kappa|\mathbf{r} - \mathbf{r}_i|)}{\epsilon|\mathbf{r} - \mathbf{r}_i|} \quad (34)$$

と計算できる。ここで \mathbf{r}_i は粗視化粒子の位置ベクトル、 ϵ は溶媒の誘電率である。

4.2.3 粗視化粒子の部分電荷の決定

粗視化粒子の部分電荷は次の評価関数が最小となるように決定する:

$$\mathcal{L}(\phi_{\text{ref}}^{\text{PB}}, \mathbf{q}) = \int_{\Omega} d\mathbf{r} \left[\phi_{\text{ref}}^{\text{PB}} - \sum_{i=1}^N q_i \frac{\exp(-\kappa|\mathbf{r} - \mathbf{r}_i|)}{\epsilon|\mathbf{r} - \mathbf{r}_i|} \right]^2 + \delta \sum_{i=1}^N (q_i - q'_i)^2 + \lambda (q_{\text{tot}} - \sum_{i=1}^N q_i)^2 \quad (35)$$

ここで, Ω は分子に近接する空間を表す. また $\phi_{\text{ref}}^{\text{PB}}$ は全原子モデルから計算されて参照静電ポテンシャル, q は粗視化粒子の電荷の集合を表す. q'_i はターゲット電荷, q_{tot} は分子の総電荷量である. δ, λ はラグランジュの未定乗数である.

評価関数の第 1 項目は全原子から求めた参照静電ポテンシャルが粗視化粒子が形成する静電ポテンシャルと一致することを要請する束縛条件である. 第 2 項目は各粗視化粒子の部分電荷に対する束縛条件である. この条件は RESP 論文において部分電荷のオーバーフィッティングを避けるために導入された束縛条件である. 特に理由がなければ RESP 論文で述べられているように $q'_i = 0$ と設定することが多い. 第 3 項目は分子の総電荷量を固定するための束縛条件である.

続いて評価関数が極値を取る条件から粗視化粒子の部分電荷を計算する方法を見ていく. 実際の計算では離散化していた方が便利なので, 評価関数 (35) の積分を和で置き換える:

$$\mathcal{L}(\phi_{\text{ref}}^{\text{PB}}, q) = \sum_{i=1}^M \left[\phi_i^{\text{PB}} - \sum_{j=1}^N q_j \frac{\exp(-\kappa|\mathbf{r} - \mathbf{r}_j|)}{\epsilon|\mathbf{r} - \mathbf{r}_j|} \right]^2 + \delta \sum_{j=1}^N (q_j - q'_j)^2 + \lambda (q_{\text{tot}} - \sum_{j=1}^N q_j)^2 \quad (36)$$

ここで i はグリッド点, j は粗視化粒子のインデックスを表していることに注意すること. 評価関数が極値を取る条件

$$\frac{\partial \mathcal{L}}{\partial q_k} = 0 \quad (37)$$

より

$$\sum_{i=1}^M 2 \left[\phi_i^{\text{PB}} - \sum_{j=1}^N q_j \frac{\exp(-\kappa r_{ij})}{\epsilon r_{ij}} \right] \frac{\exp(-\kappa r_{ik})}{\epsilon r_{ik}} + 2\delta \sum_{j=1}^N (q_j - q'_j) + 2\lambda (q_{\text{tot}} - \sum_{j=1}^N q_j) = 0 \quad (38)$$

を得る. これを整理すると,

$$\sum_{i=1}^M \sum_{j=1}^N \left\{ \frac{\exp[-\kappa(r_{ij} + r_{ik})]}{\epsilon^2 r_{ij} r_{ik}} - \lambda \right\} q_j - \delta q_k = \sum_{i=1}^M \frac{\exp(-\kappa r_{ik})}{\epsilon r_{ik}} \phi_i^{\text{PB}} - \delta q'_k - \lambda q_{\text{tot}} \quad (39)$$

となる. 全ての粒子の電荷 q_k , $k = 1, \dots, N$ について同様の方程式を得ることができるので, 行列形式で書き直すと

$$A_{jk} = \sum_{i=1}^M \frac{\exp[-\kappa(r_{ij} + r_{ik})]}{\epsilon r_{ij} r_{ik}} + \lambda' \quad (j \neq k) \quad (40)$$

$$A_{kk} = \sum_{i=1}^M \frac{\exp[-2\kappa r_{ik}]}{\epsilon r_{ik}^2} + \delta' + \lambda' \quad (j = k) \quad (41)$$

$$B_j = \sum_{i=1}^M \frac{\exp[-\kappa r_{ik}]}{\epsilon r_{ik}} \phi_i^{\text{PB}} + \delta' q'_k + \lambda' q_{\text{tot}} \quad (42)$$

を導入すると

$$\mathbf{A} \mathbf{q} = \mathbf{B} \quad (43)$$

の連立方程式が得られ,

$$\mathbf{q} = \mathbf{A}^{-1} \mathbf{B} \quad (44)$$

のように連立方程式を解くことで最適な部分電荷の値を得ることができる. なお, 途中で数式の便利のために $\lambda' = -\lambda$, $\delta' = -\delta$ のように置き直した.

4.2.4 蛋白質表面のアミノ酸残基の特定

RESPAC の論文では 1 つのビーズが 1 つのアミノ酸に対応する粗視化表現を使用しており、表面のアミノ酸を表す粒子に対してゼロではない電荷を割り振るように提案している。これは、RESP 法において蛋白質内部に埋もれた電荷がうまく決定されない傾向があり、タンパク質構造の予期せぬ不安定化を引き起こす可能性があることが知られているためである。したがって、評価関数 (35) では表面に露出しているアミノ酸のみを考慮する。蛋白質表面に露出しているアミノ酸は以下のようにして決定することができる:

1. 1 \AA で区切られた 3 次元格子をターゲット蛋白質周辺に設定する。
2. 全てのグリッド点について各原子との距離を計算して、それが原子 + プローブ球より離れている点については 1, 小さい場合は 0 のフラグを立てる。つまり、蛋白質 + プローブ球の内部にあるグリッド点を 0, その他の蛋白質外部にあるグリッド点を 1 に設定する。
3. 1 のフラグを立てたグリッドについて、グリッド座標と原子の距離を全原子について計算して、最も距離の近い原子を持つアミノ酸残基を表面アミノ酸残基として定義する。

4.2.5 スケーリング係数 δ と λ の決定方法

評価関数 (35) を最小化するためには、適切なスケーリング係数 δ , λ を選択する必要がある。分子の総電荷に対する束縛条件のスケーリング係数である λ は、得られる有効電荷にあまり影響しないため、十分大きい値を設定しておけばいい。例えば RESPAC 論文では $\lambda = 10^5$ と設定している。

一方で、各粒子に対する参照電荷の束縛条件のスケーリング係数 δ の値は RESPAC で得られる有効電荷に大きく影響を与える。最適な δ の値を得るためのプロトコルとして次のようなものが提案されている: (i) 全原子モデルに基づいた短い MD シミュレーションを実行して、10 個の構造をサンプリングする。 (ii) 各構造に対して、評価関数 (35) を最小化するようにして粗視化電荷を決定する。 (iii) 得られた粗視化電荷 q に対して、規格化された誤差 $\Delta(q(\delta))$

$$\Delta(q(\delta)) = \frac{\int_{\Omega} d\mathbf{r} \left[\phi_{\text{ref}}^{\text{PB}} - \sum_i^N q_i \frac{\exp(-\kappa|\mathbf{r}-\mathbf{r}_i|)}{\epsilon|\mathbf{r}-\mathbf{r}_i|} \right]^2}{\int_{\Omega} d\mathbf{r} [\phi_{\text{ref}}^{\text{PB}}]^2} \quad (45)$$

を計算する。 (iv) サンプリングした 10 個の構造に対して誤差の平均 $\langle \Delta(q(\delta)) \rangle$ を計算し、誤差の平均が最小となるような δ を採用する。

参考文献

- [1] Tsuyoshi Terakawa and Shoji Takada. RESPAC: Method to Determine Partial Charges in Coarse-Grained Protein Model and Its Application to DNA-Binding Proteins. *Journal of Chemical Theory and Computation*, Vol. 10, No. 2, pp. 711–721, feb 2014.
- [2] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, Vol. 97, No. 40, pp. 10269–10280, oct 1993.
- [3] Brent H. Besler, Kenneth M. Merz, and Peter A. Kollman. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry*, Vol. 11, No. 4, pp. 431–439, may 1990.