

# Spectral decomposition of atomic structures in heterogeneous cryo-EM

簡述：

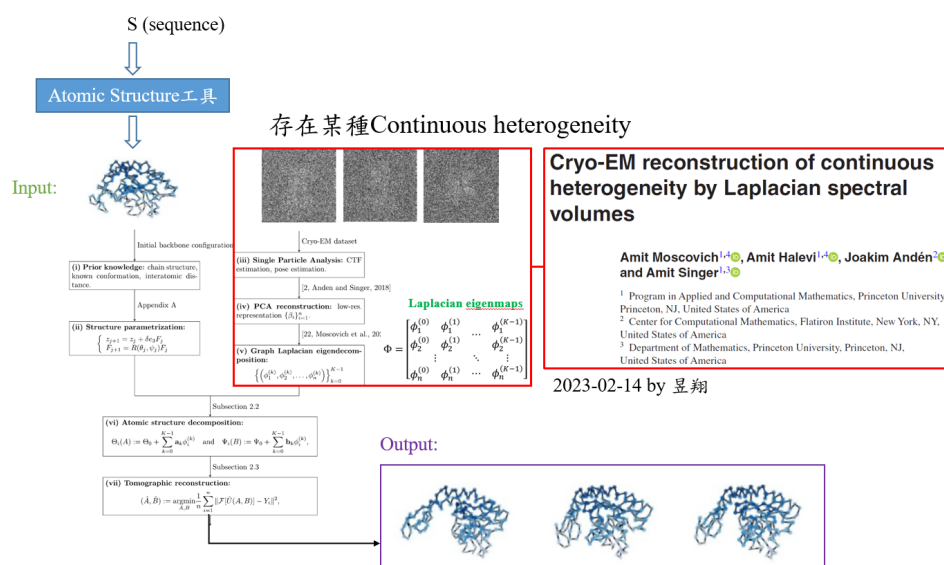


Figure 1：完整流程圖

從流程圖裡，可以很清楚的知道，裡頭包含了另外一篇論文裡的方法，並結合了該方法的精神。因此，我會先介紹 Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes 的方法與精神，然後再說明 Spectral decomposition of atomic structures in heterogeneous cryo-EM。

## 一、Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes

### ● 假設與目的：

在這篇論文當中，假設了我們所取得的 cryo-EM images 來自一個具有 continuous heterogeneity 的 particle，進而做了一個假設「每張 cryo-EM image 分別來自某個 conformation，這些 conformations 之間具有某種連續變化。」

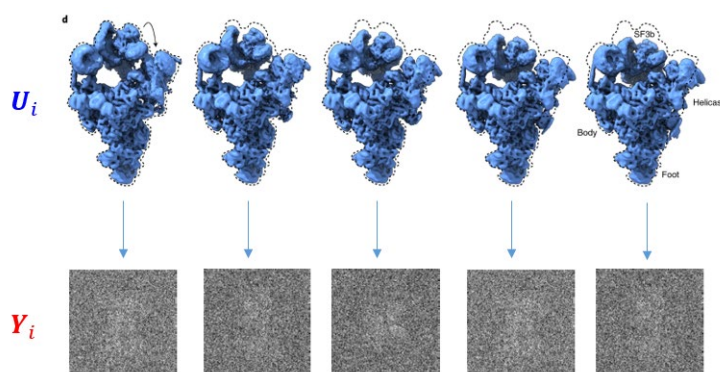


Figure 2：Continuous heterogeneity 示意圖

換句話說，每張 cryo-EM image 都是某個 conformation 的投影，而這個關係可以用式子更清楚的表示：

$$Y_i = h_i * T(U_i) + \varepsilon_i, i = 1, \dots, n$$

其中  $T$  表示電子束的垂直投影， $h_i$  為 point spread function (PSF)，在這篇論文當中，還假設  $T$  與  $h_i$  都是事先知道的。

這篇論文中所感興趣的部分為  $U_i$  之間的 conformation 變化，因此在實際上只能拿到  $Y_i$  的資訊的前提下，困難點有二：1、要如何從  $Y_i$  轉換成  $U_i$ ，並更進一步的 2、描述 continuous heterogeneity，

### ● 方法：

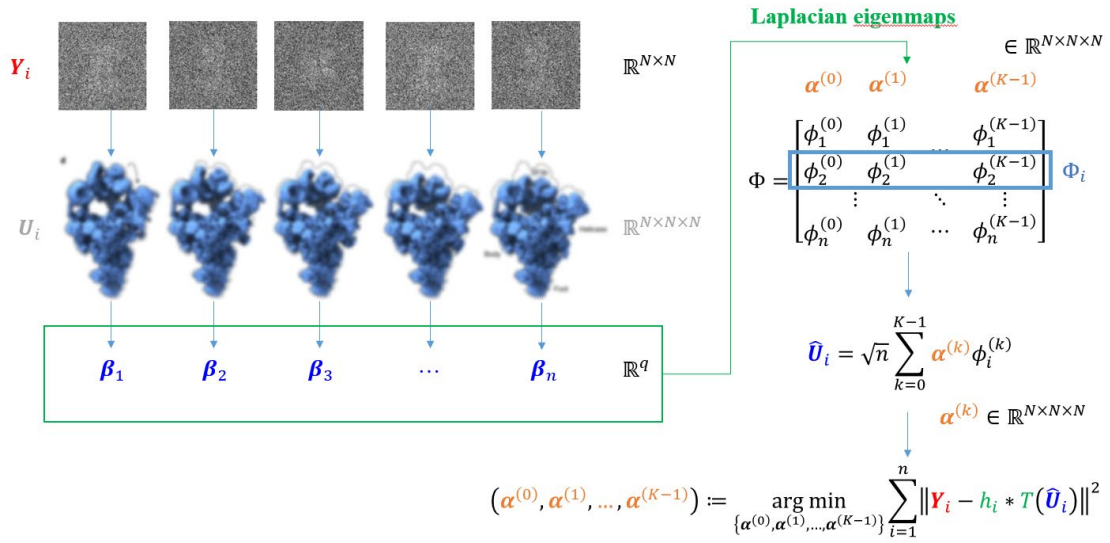


Figure 3：Cryo-EM reconstruction of continuous heterogeneity 流程圖

步驟：

1. 利用  $Y_i$  提取出可以代表  $U_i$  的 PCA 特徵係數  $\beta_i \in \mathbb{R}^q$
2. 使用  $\beta_i$  做 Laplacian eigenmaps 降至  $K$  維，即  $\Phi_i \in \mathbb{R}^K$ ，得到在  $\mathbb{R}^K$  空間中  $\beta_i$  ( $U_i$ ) 之間的關係。

在 Laplacian eigenmaps 步驟裡，有一個很重要的特點，就是我們得到的  $\Phi_i$ ，雖然可以用來畫在座標上，但是座標軸卻還沒有被賦予意義的！

3. 用  $\Phi_i$  重建出每個  $Y_i$  背後對應的  $\hat{U}_i$ ，也就是根據我們的目標，賦予  $\Phi_i$  每個 component 合適的意義，Moscovich et al. (2020) 主要目的是要重建出  $U_i$ ，所以讓每個 component 都對應到一個 3D volume。

步驟相關備註：

《步驟 1》如何從  $Y_i$  轉換成  $\beta_i$ ？

參考 Andén and Singer (2018)，或者可以嘗試其他辦法。

## 《步驟 2》Laplacian eigenmaps

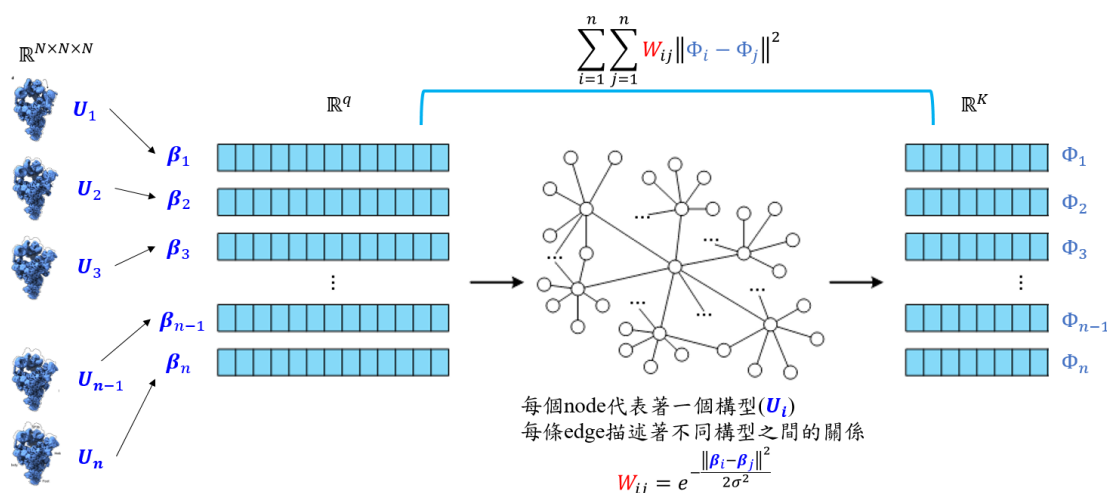


Figure 4：Continuous heterogeneity 之 Laplacian eigenmaps 示意圖

目的：用來描述不同 conformations 之間的關係，如果是具有連續變化的關係，例如有某個缺口會慢慢打開，或是慢慢閉合，那麼 conformations 可以用合適的 manifold 描述出來。

方法特色：著重在描述 local 的表現，希望降維後的資料可以充分保留下高維度的 local 關係，藉由一個一個 local 之間的連結，將所有資料結合在一起。

$$\min_{\Phi} \sum_{i=1}^n \sum_{j=1}^n W_{ij} \|\Phi_i - \Phi_j\|^2$$

如果我們將 Laplacian eigenmaps 最後降維的所有資料寫成一個  $n \times K$  的矩陣  $\Phi$ ， $\Phi$  的第  $i$  個 row 記為  $\Phi_i \in \mathbb{R}^K$ ， $\Phi$  的第  $j$  個 row 記為  $\phi_j \in \mathbb{R}^n$ ，且定義  $W =$

$[W_{ij}]_{n \times n}$  以及  $D = \text{diag}(d_{ii})$ ，其中  $d_{ii} = \sum_{j=1}^n W_{ij}$ ，則這個最佳化問題可以改寫為

$$\min_{\Phi} 2\text{tr}[\Phi^T(D - W)\Phi]$$

倘若限制求出來的  $\phi$  不能為 0，這將引導成為 general eigenvalue problem，即

$$L\phi = \lambda D\phi$$

其中  $L = D - W$ 。

會議中討論：

(1) 雖然 Laplacian eigenmaps 中涉及 general eigenvalue problem，但是求出來的 general eigenvector 並非是基底，直接就是座標軸的係數，這個非常重要!!!

(2) 對於利用越多個 eigenvectors  $\phi_j$ ，可以越  $\min_{\Phi} 2\text{tr}[\Phi^T(D - W)\Phi]$ ，是否

有更直觀的看法？（感謝定立老師的分享，希望我沒有理解錯誤）

當我們用越多的 $\phi_j$ ，相當於用更多個 components 一起去模擬高維度的資料，所以可以慢慢降低差距。舉例來說，如果我們只用一個 component，那麼最貼近的辦法，只能用 mean。

(3) 是否可以用 tSNE 或是其他的工具來替換 Laplacian eigenmaps？

Laplacian eigenmaps 在建立 graph 的時候，使用的是 knn 的架構，而最後求座標位置的時候，是解 general eigenvalue problem，依舊屬於線性範疇，較可以完整保留住高維度資料的資訊。

而 tSNE 跟 Laplacian eigenmaps 最不一樣的地方，在於低維度呈現時，會讓原本在高維度相近的點，在低維度時更靠近，在高維度距離遠的點，在低維度時更遠，這是因為這個方法更在乎的是視覺上的呈現，因此低維度時會誇大群聚的特色，進而有一定程度的扭曲。

另外一個相當不同的差別是，Laplacian eigenmaps 是唯一解，tSNE 會因為起始點的不同，使得 output 也不同。若是比較計算時間，Laplacian eigenmaps 也相對較快速。

《步驟 3》用 $\Phi_i$ 重建出每個 $Y_i$ 背後對應的 $\hat{U}_i$

這個步驟相當特別，因為它利用到 Laplacian eigenmaps 的 output 的 $\Phi$ ， $\Phi$ 每個 component 沒有意義的這個特性。

正因為 $\Phi$ 的 component 沒有意義，所以可以根據我們的目的，去賦予它意義。

譬如 Moscovich et al. (2020) 想要重建出 3D volume，所以利用 SGD 去找出對應每個軸合適的 3D volume，即 $\alpha^{(k)} \in \mathbb{R}^{N \times N \times N}$ ，如 Figure 5。

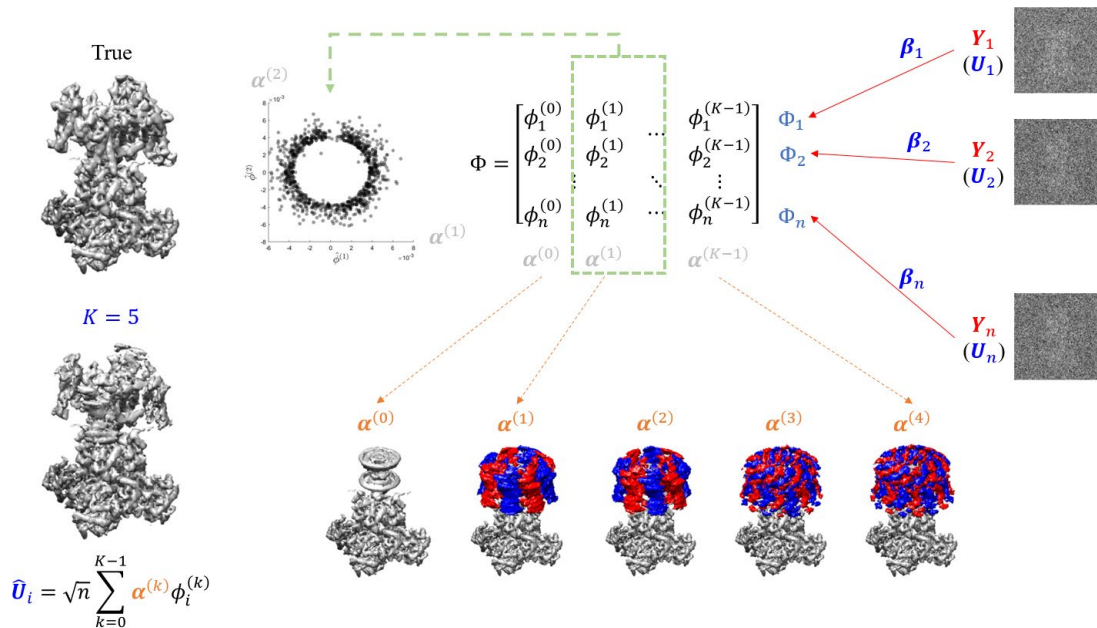


Figure 5：用 $\Phi_i$ 重建出每個 $Y_i$ 背後對應的 $\hat{U}_i$

## 二、Spectral decomposition of atomic structures in heterogeneous cryo-EM

- 目的：

流程圖 (Figure 1) 中的紅色部分，已經提供了「從 $Y_i$ 建立 conformation 的關係圖( $\Phi$ )，再從 $\Phi$ 重建出 $\hat{U}_i$ 」的完整流程，這相當於證實了確實可以從每個 $Y_i$ 找到背後所代表某個 $U_i$ 的途徑。

如果將這個想法，結合 atomic structure，也就是說，「atomic structure 也應該根據 $U_i$ 的變化而變化。」

- 需要預備的背景知識

### (1) Molecule Structure

為此篇論文解析 atomic structure 時所使用的主要方式，只針對 backbone 上的 $C_\alpha$ ，並借用 bond angle  $\psi$  與 torsion angle  $\theta$  概念建立空間關係。

對應 bond angle 的部分：

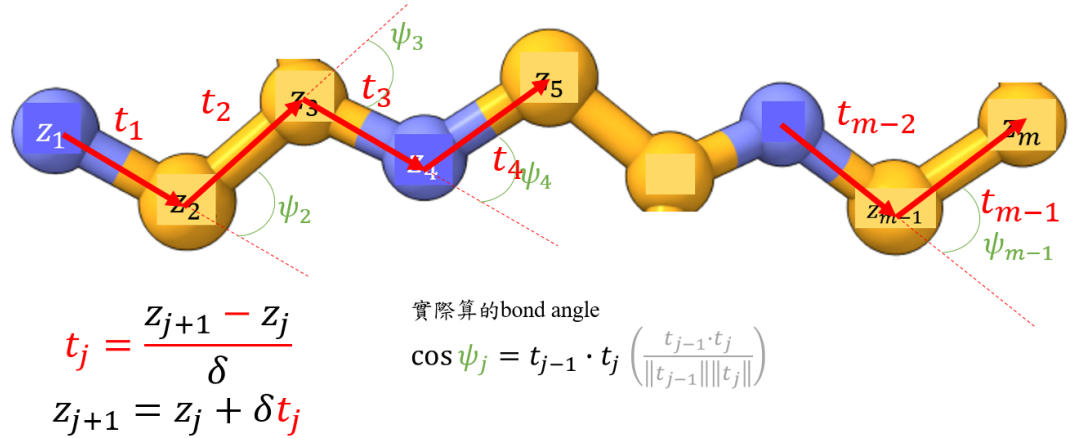


Figure 6 : bond angle  $\psi$

對應 torsion angle 的部分：

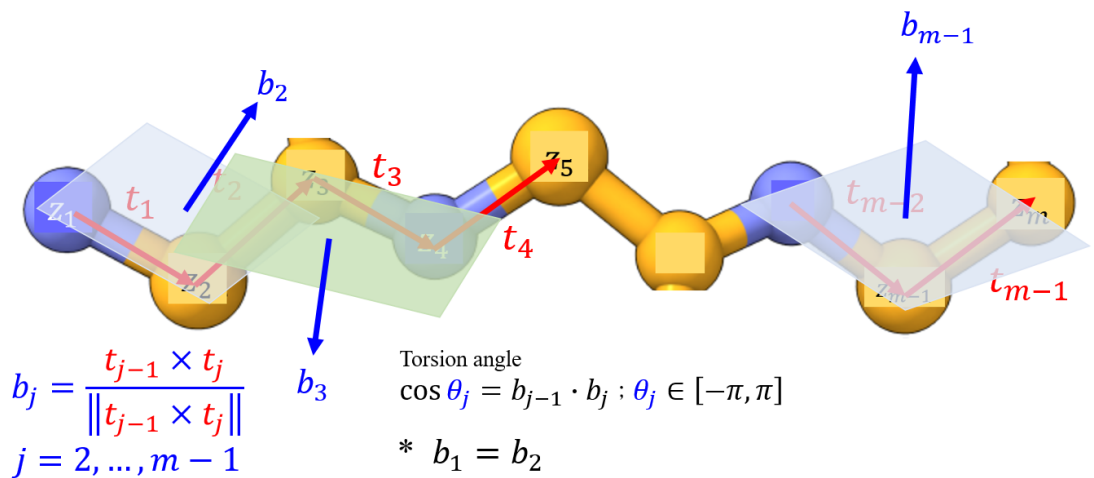


Figure 7 : torsion angle  $\theta$



## (2) Frenet Frame

上述的 molecule structure 剛好跟 Frenet frame 的架構重和度相當高。

可以將其視為一種離散型態的 Frenet frame 系統。

假設 atomic structure 上共有  $m$  個胺基酸，即有  $m$  個  $C_\alpha$ ，

每個  $C_\alpha$  的空間座標為  $z_j \in \mathbb{R}^{1 \times 3}$ ， $j = 1, \dots, m$ ，

而  $t_j \in \mathbb{R}^{1 \times 3}$  與  $b_j \in \mathbb{R}^{1 \times 3}$  如 Figure 6, 7 所定義。

每一個 Frenet frame 需要 3 個  $C_\alpha$  決定，並將 Frenet frame  $F_j$  紀錄為

$$F_j = \begin{bmatrix} n_j \\ b_j \\ t_j \end{bmatrix}_{3 \times 3}, \quad j = 2, \dots, m-1$$

若在  $z_1$  補上虛擬的  $b_1$  (平行於  $b_2$ )，則可以創造出  $F_1$ 。

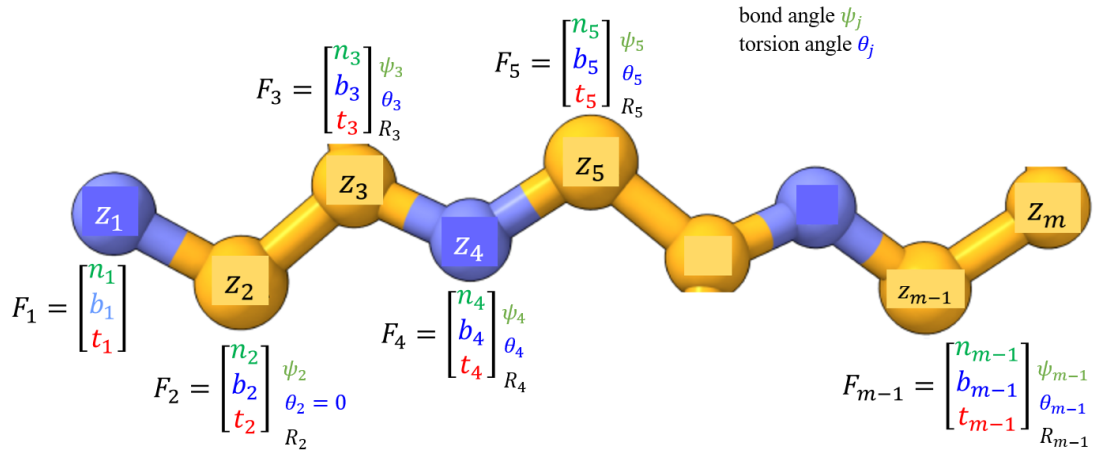


Figure 8：每個  $z_j$  可以擁有的 Frenet Frame 資訊

從建構出  $F_j$  的規則中，我們可以知道  $F_j$  與  $F_{j-1}$  之間存在一個旋轉矩陣的轉換關係，即  $F_j = R_j F_{j-1}$ ， $j = 2, \dots, m-1$ 。

$$R_j = \begin{bmatrix} \cos \psi_j & -\sin \psi_j & 0 \\ 0 & 1 & 0 \\ \sin \psi_j & \cos \psi_j & 0 \end{bmatrix} \begin{bmatrix} \cos \theta_j & \sin \theta_j & 0 \\ -\sin \theta_j & \cos \theta_j & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(證明見 Appendix A.1)

## (3) 表達 atomic structure 的方式

已有工具  $\Theta = (\theta_2, \theta_3, \dots, \theta_{m-1}) \in \mathbb{R}^{m-2}$  (torsion angle)

$\Psi = (\psi_2, \psi_3, \dots, \psi_{m-1}) \in \mathbb{R}^{m-2}$  (bond angle)

方法 1，the positions of the atoms

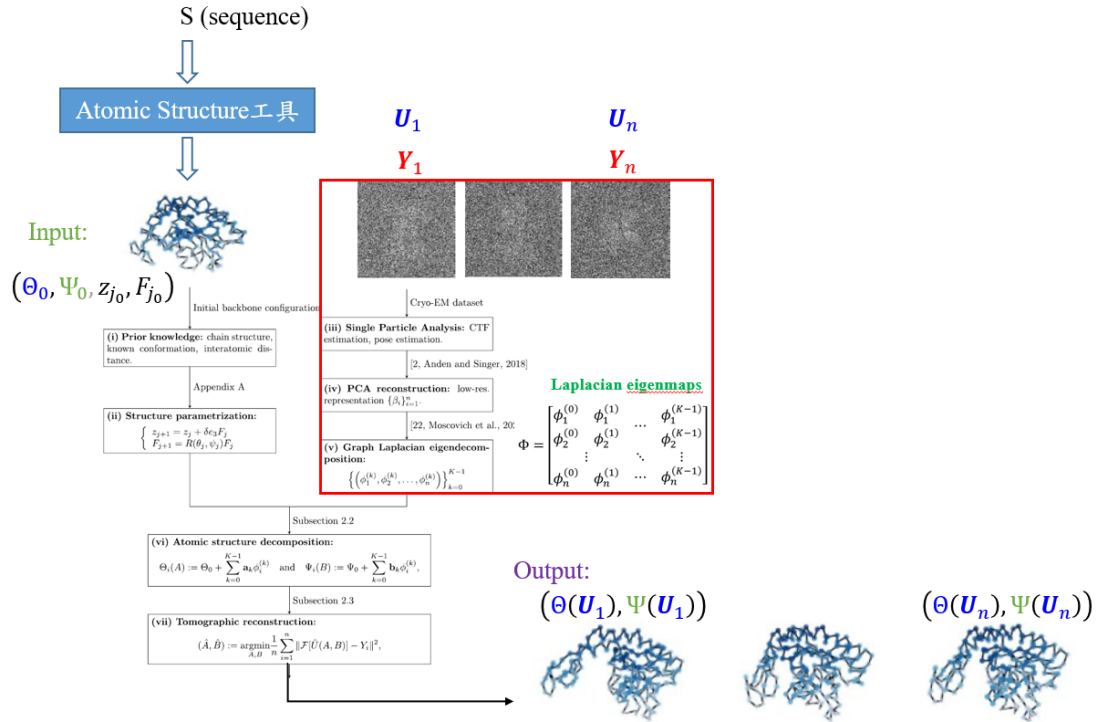
$$(z_1, z_2, \dots, z_m) \in \mathbb{R}^{3m}$$

方法 2，Frenet frame 與 the associated point cloud

只要有 torsion angles 與 bond angles 的資訊，再加上一個屬於中間部分的  $z_{j_0}$  與它的 Frenet frame  $F_{j_0}$ ，就可以依靠  $F_j = R_j F_{j-1}$  的關係，推估所有的 the positions of the atoms，即

$$\mathbf{Z}(\Theta, \Psi, z_{j_0}, F_{j_0}) = (z_1, z_2, \dots, z_m)$$

● 方法：



輸入：atomic structure (例如: Alphafold 2 的 output)，提取  $\Theta_0, \Psi_0, z_{j_0}, F_{j_0}$  訊息  
 n 張 cryo-EM images (每張  $Y_i$  的背後各自代表著某個  $U_i$ )

步驟：

1. 效法 Moscovich et al. (2020) 中的步驟 1, 2，從  $Y_i$ ，得到還未賦予軸意義的  $\Phi_i$ 。
2. 利用  $\Phi_i \in \mathbb{R}^K$  作為係數，分別作用在 torsion angles 與 bond angles 兩個系統上，假設在 torsion angles 的系統上，可以找到對應  $\Phi_i$  每個軸上背後可以對應的 torsion angle 的成分組合  $a_k \in \mathbb{R}^{m-2}$ ；同樣的，在 bond angles 的系統上，每個軸對應到的 bond angle 的成分組合  $b_k \in \mathbb{R}^{m-2}$ 。

$$\Theta(U_i) = \Theta_0 + \sum_{k=0}^{K-1} a_k \phi_i^{(k)} = \Theta_0 + A_{(m-2) \times K} \Phi_i$$

$$\Psi(U_i) = \Psi_0 + \sum_{k=0}^{K-1} b_k \phi_i^{(k)} = \Psi_0 + B_{(m-2) \times K} \Psi_i$$

(注意事項) 這是一個「假設」步驟，假設我們如果找到軸對應的意義後，會怎麼去重構各個  $U_i$  的  $\Theta(U_i)$ 、 $\Psi(U_i)$ ，因此在假設的前提下，這步驟時我們還不知道 A、B 為何，因此也不知道  $\Theta(U_i)$ 、 $\Psi(U_i)$  是怎樣的。

3. 從 atomic structure 去模擬 cryo-EM 的 density map

「假設」我們已經有了調整後的  $\Theta(U_i)$ 、 $\Psi(U_i)$ ，在根據一早選定的  $z_{j_0}, F_{j_0}$ ，根據 the associated point cloud，

$$\Gamma_i(A, B) = \mathbf{Z}(\Theta(\mathbf{U}_i), \Psi(\mathbf{U}_i), z_{j_0}, F_{j_0}) = (z_1, z_2, \dots, z_m)$$

這相當於 cryo-EM density map 的中心骨幹，接著可以模擬出骨感周圍的 density map

$$\hat{\mathbf{U}}_i(A, B)(x) = \sum_{z_j \in \Gamma_i(A, B)} v \exp \left[ -\frac{\|x - z_j\|^2}{2\sigma^2} \right]$$

(注意事項) 這是一個相當重要的步驟，因為它架起了從 atomic structure 到 cryo-EM structure 的橋樑，並且別忘了，我們實際上擁有可用的資料，其實只有 cryo-EM images  $\mathbf{Y}_i$ 。

#### 4. 利用 $\hat{\mathbf{U}}_i$ 去反推 A、B

由於  $\hat{\mathbf{U}}_i$  連結起 atomic structure 與 cryo-EM structure，且其中未知參數只有 A、B，所以用估計對應  $\mathbf{Y}_i$  的投影  $\hat{\mathbf{Y}}_i$ ，

$$\hat{\mathbf{Y}}_i(A, B) = h_i * T(\hat{\mathbf{U}}_i)$$

則求得最佳的 A、B，相當於

$$[\hat{A}, \hat{B}] = \underset{A, B}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n \|\hat{\mathbf{Y}}_i(A, B) - \mathbf{Y}_i\|_2^2$$

求解方法為 Stochastic gradient descent (SGD) algorithm

### ● 相關討論議題(以下只是我的看法，也歡迎大家一起提出各自的觀點)

#### 1. 為什麼要使用兩次降維方法？(PCA 與 Laplacian eigenmaps)

(感謝思齊的提問。)

在 Moscovich et al.(2020) 中，從  $\mathbf{Y}_i$  到提取出  $\beta_i$ ，雖然這篇裡沒有使用到每個  $\beta_i$  component 背後的 PC，但理論上來說，我們也可以用  $\beta_i$  來重建出  $\tilde{\mathbf{U}}_i$ 。從流程圖中，我們可以知道  $\Phi_i$  的維度比  $\beta_i$  更低，不難推測出，由  $\Phi_i$  所建構出的  $\tilde{\mathbf{U}}_i$  的解析度比  $\tilde{\mathbf{U}}_i$  更低。

這邊提出一個我的小小的想法：如果我們想要看到相對高解析度的重建，我們可以只使用  $\Phi_i$  提供的 conformations 之間的關係，當我們想要知道背後的  $\mathbf{U}_i$  實際上模樣時，或許可以回到由  $\beta_i$  所重建出的  $\tilde{\mathbf{U}}_i$ 。

另外，雖然  $\Phi_i$  只能建出相對低解析度的  $\tilde{\mathbf{U}}_i$ ，但是這個方法背後引伸出了一個很創新的想法，就是「根據 conformations 之間的關係，重新賦予軸意義」，而這個意義完全可以視需求而定。

正因為 Moscovich et al.(2020) 成功了，所以才有了 Esteve-Yagüe et al. (2022)。這也使得第二步的降維(Laplacian eigenmaps)在這個方法中扮演相當重要的角色。

#### 2. 在 Esteve-Yagüe et al. (2022) 中，從 atomic structure 過渡到 cryo-EM structure 時，使用的 $\hat{\mathbf{U}}_i(x)$ 模型跟我們在 Q-scores 看到的模型相同，而



我們正致力於找出是否有更貼近現實的 $\sigma$ 。

Reference:

Andén, J., and Singer, A. (2018). Structural variability from noisy tomographic projections. *SIAM Journal on Imaging Sciences*, 11(2), 1441-1492.

Moscovich, A., Halevi, A., Andén, J., and Singer, A. (2020). Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. *Inverse Problems*, 36(2), 024003.

Esteve-Yagüe, C., Diepeveen, W., Öktem, O., and Schönlieb, C. B. (2022). Spectral decomposition of atomic structures in heterogeneous cryo-EM. *Inverse Problems*.

Appendix

A.1

一般來說，在三維座標中， $v_1 \in \mathbb{R}^{3 \times 1}$ 經過旋轉矩陣 $R \in \mathbb{R}^{3 \times 3}$ 作用後，變成 $v_2 \in \mathbb{R}^{3 \times 1}$ ，可寫為

$$v_2 = R \cdot v_1$$

因此， $F_{j-1}$ 要轉至 $F_j$ ，可以分成兩步驟：

第一步，固定 $b_{j-1}$ ，在 $n_{j-1}$ 與 $t_{j-1}$ 所構成的平面上，將 $t_{j-1}$ 轉 $\psi_j$ (bond angle)至 $t_j$ ，並更新 $\tilde{b}_j$ 與 $\tilde{n}_j$ 。

第二步，固定 $t_j$ ，在 $\tilde{n}_j$ 與 $\tilde{b}_j$ 所構成的平面上，將 $\tilde{b}_j$ 轉 $\theta_j$ (torsion angle)至 $b_j$ 。

綜合上述，可以寫成

$$\begin{bmatrix} n_j^T & b_j^T & t_j^T \end{bmatrix} = \begin{bmatrix} n_{j-1}^T & b_{j-1}^T & t_{j-1}^T \end{bmatrix} \begin{bmatrix} \cos \theta_j & -\sin \theta_j & \\ \sin \theta_j & \cos \theta_j & \\ & & 1 \end{bmatrix} \begin{bmatrix} \cos \psi_j & \sin \psi_j & \\ & 1 & \\ -\sin \psi_j & \cos \psi_j & \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

等號兩邊取轉置，

$$F_j = \begin{bmatrix} \cos \psi_j & -\sin \psi_j & \\ & 1 & \\ \sin \psi_j & \cos \psi_j & \end{bmatrix} \begin{bmatrix} \cos \theta_j & \sin \theta_j & \\ -\sin \theta_j & \cos \theta_j & \\ & & 1 \end{bmatrix} F_{j-1}$$

可得到

$$R_j = \begin{bmatrix} \cos \psi_j & -\sin \psi_j & \\ & 1 & \\ \sin \psi_j & \cos \psi_j & \end{bmatrix} \begin{bmatrix} \cos \theta_j & \sin \theta_j & \\ -\sin \theta_j & \cos \theta_j & \\ & & 1 \end{bmatrix}$$