arXiv:2611.28961v1 [physics.bio-ph] 21 Nov 2026

时间分辨晶体学中的外推因子估计

SebastianBielfeldt [a,b]，HenryN.Chapman [a,c,d]，以及 ThomasJ. Lane ✉

[a] 自由电子激光科学中心 CFEL，德国电子同步加速器 DESY，Notkestr. 85，22607 Hamburg，Germany [b] 汉堡工业大学，算法与复杂性研究所，Hamburg，Germany [c] 汉堡超快成像中心，Luruper Chaussee 149，22761 Hamburg，Germany [d] 汉堡大学物理系，Luruper Chaussee 149，22761 Hamburg，Germany

摘要

时间分辨晶体学（TRX）能够捕捉处于作用过程中的大分子，但触发态晶体包含处于未知状态的触发态蛋白质与处于参考态的未触发蛋白质的混合物。要分离这两者，从而提取所关注的动态过程，必须解决一个欠定的数据分析问题。对于包含一个触发态和一个参考态的最简单情况，已有一系列方法，但它们的假设、准确性和适用范围尚未得到系统研究。在本文中，我们采用统一的符号体系，对当前使用的方法进行拆解和综述。我们明确阐述这些方法的假设，并在模拟数据集上对每种方法进行基准测试。一个关键点是，许多方法估计的并不是晶体中所有触发态分子的比例，而是我们称为外推因子的潜在参数；该参数与所谓的“占有率”相关，但并不相同。

正如我们的模拟所揭示的，当前方法并不总能得出准确结果，尤其是在数据噪声较大、触发的蛋白质比例较低时。受前人工作的启发，我们提出了真空能级匹配，这种方法以一项物理约束为基础进行估计：电子密度是非负量，不会低于真空水平。真空能级匹配逐体素进行处理，以数据驱动的流程取代了以往方法中的人工选择，并报告质量控制指标，以标示无法支持可靠预测的数据。对于相同的模拟数据，该方法优于当前最先进的方法；对于两个实验数据集——rsEGFP2 和多时间点光裂合酶系列——该方法能够恢复已知的结构变化，并在数据较弱时标示出问题。最后，我们报告了真空能级匹配的一种实现形式：一个小型、可复用的

Python 软件包。

1 引言

近年来，随着X射线光源和探测器的发展，从静态结构快照逐步转向观察生物大分子动态过程变得越来越可行，时间分辨晶体学（TRX）因此受到了广泛关注。通过捕捉在飞秒至秒级时间尺度上形成的亚稳态中间体，研究人员旨在直观呈现配体结合、催化和信号转导所需的精确原子运动（Khusainov 等，2024；Br¨and´en & Neutze，2021）。

近期研究充分展现了这一方法的强大能力。我们利用 TRX 揭示了光裂合酶修复 DNA 的机制，使我们得以观察光激发、随后胸腺嘧啶二聚体的断裂以及产物释放（Christou 等，2023年）。另一个案例中，研究人员将 TRX 与电场刺激结合使用，以观察 NaK2K通道中的钠离子传导机制（Lee等，2025年）。

尽管 TRX 长期以来主要由少数专家开展研究，但随着各种赋能技术日趋成熟，这一领域正在不断发展。例如，过去十年的大部分研发工作虽由 X射线自由电子激光器（XFEL）设施推动，但机时供应更充足的同步辐射光源正日益建设性能完善的 TRX专用光束线（Pearson & Mehrabi, 2020年）。与此同时，串行晶体学的样品递送技术不断创新；这种方法将来自大量微晶体的衍射数据合并为一个数据集。借助液体射流和固定靶，样品消耗需求已降低，可研究的蛋白质体系范围也因此扩大（Barends等，2022年；Banari等，2025年）。最后，支持 TRX 的程序和数据分析流程已发展成熟，形成了可供实际使用的软件包（Grosse-Kunstleve 等，2002年；White等，2012年；De Zitter等，2022年；Fadini等，2025年）。因此，人们有理由期待，不具备 TRX 专业知识的用户也能常规运用这项技术来探索新的生物学问题。

然而，仍存在重大障碍。一个重要的挑战在于分析TRX数据中混合的结构构象。在TRX实验中，通常会通过光学激光器发出的闪光、与配体快速混合、施加外场或其他合适的扰动来触发动态过程。这一脉冲会启动所关注的反应或运动，并可通过晶体学方法加以追踪。遗憾的是，采用这种方法几乎总是无法获得完全触发态的样品，即晶体中的所有蛋白质都被触发，并在所关注的动态过程中保持结构同步。首先，触发通常只能激活晶体中一小部分蛋白质，许多未受扰动的分子仍处于基态或参考态。其次，由于触发态蛋白质以随机方式经历反应中间体，它们在任一给定时刻可能并未都采取相同的结构构象。因此，一般而言，TRX实验中的触发晶体包含多种蛋白质构象的混合物。分离这种混合物，以便解析其中每种有贡献的中间体的结构，是所有TRX研究者的一项核心数据分析任务。

与之对应的最简单晶体学信号模型认为，观测到的复结构因子可以用线性组合来描述，

$$\mathbf{F}_{\mathrm{mix}}(t) = \alpha_0(t)\mathbf{F_0} + \alpha_1(t)\mathbf{F_1} + \alpha_2(t)\mathbf{F}_2 + \cdots \tag{1}$$

其中我们将结构因子写作 $\mathcal{F}\{\rho(\mathbf{r})\} = \mathbf{F(q)}$。复数值用

粗体表示，并省略 $\mathbf{F}$ 的自变量以简化记号。

这一线性模型建立在关于样品的若干假设之上，而这些假设很少被明确说明。首先，该模型假设每个晶格位点均由一种构象占据，并按照各构象的占有率随机选定，且这一选择在统计上独立于相邻晶格位点。于是，在布拉格峰位置观测到的衍射由平均晶胞电子密度的傅里叶振幅给出。各个构象的结构因子根据其占有率进行加权，并相干相加。相反，如果晶格位点之间存在相关性，使其分隔成不同的畴或晶体，那么衍射信号将包含对强度求和的项，而不再与其复结构因子的单一相干和成正比。其次，群体 α_i（t和结构因子$\mathbf{F}_i$）均假设在每个时间点保持不变，尽管每次测量都在有限的时间分辨率内进行平均；该时间分辨率由泵浦-探测抖动以及脉冲或曝光时长等因素决定，而在此期间群体仍会持续演化。最后，如果结构因子是从多个晶体中估算得到的，例如在串行晶体学实验中，那么上述公式假设每个晶体的制备方式完全相同，每种单独构象都具有相同的群体。上述假设成立的程度，与其说取决于数据分析，不如说取决于实验本身，例如晶体尺寸、泵浦通量和时间分辨率等选择。我们在附录B中考察晶体之间占有率方差的影响。然而，在这些可以通过实验实现的假设下，该模型为分析时间分辨数据提供了一个有效框架，本文将始终采用这一框架。

在该模型下，我们要解决的问题是：给定在不同时间点 t 测得的一组振幅 $F_{\mathrm{mix}}(t)$，恢复种群振幅 α_i 以及中间态复结构因子 $\mathbf{F}_i$。理想情况下，应测量许多时间点，并且不同观测值的数量$F_{\mathrm{mix}}(t)$远多于所关注的中间态数量。其列为$\mathbf{F}_{\mathrm{mix}}(t)$的矩阵，其列数将远大于其秩；该秩等于线性无关中间态的数量。在这种理想情形下，可以使用低秩方法推断潜在的结构因子和种群（Schmidt *et al.*，2003）。遗憾的是，在当前实践中，获取 TRX 时间点的成本极高，因此几乎不可能在时间上进行过采样。当可实现的时间点数量与所关注的结构中间态数量大致相当或更少时，一种务实的方法是选择时间点 t，使得——除去任何仍未触发的蛋白质——其中一个中间态占据压倒性优势，而其他中间态的贡献可以忽略。本文将专门关注这种更为常见的数据受限区间。

如果满足这一条件，且在特定时间 t 某个单一中间态占绝对优势，则观测到的结构因子近似为

$$\mathbf{F}_{\mathrm{mix}} \approx \alpha_0 \mathbf{F_0} + \alpha_1 \mathbf{F_1} \tag{2}$$

其中，$\alpha_0\mathbf{F_0}$ 是仍处于参考态的未触发蛋白质分子所产生的残余贡献，而 $\alpha_1\mathbf{F_1}$ 对应于随时间演化的群体产生的信号，也就是目标信号。正如我们接下来将要讨论的，遗憾的是，分离这一信号是一个欠定问题。

有些量是已知的。如上所述，振幅 $F_{\mathrm{mix}}(t)$ 可直接通过时间分辨测量获得。通过在触发样品之前获取参考结构，可以

确定振幅 F_0 以及相应的 $\mathbf{F_0}$ 相位。假设存在两个状态，则给出$\alpha_0(t) + \alpha_1(t) = 1$，从而只剩下一个布居参数 $\alpha_1(t)$——通常称为占有率，简称为 α。因此，剩余的未知量是占有率α、$\mathbf{F}_{\text{mix}}$ 的相位，以及复结构因子 $\mathbf{F_1}$。

根据公式 2可知，若能估算出 α 以及$\mathbf{F}_{\text{mix}}$的相位，就有可能求得$\mathbf{F_1}$。在估算出$\mathbf{F}_{\text{mix}}$后，估算合适 α的任务通常称为占有率测定。然而，事实证明这非常困难，在实践中很少这样做；我们将说明，大多数研究人员实际上采用了一个更简单但仍然有效的问题，即我们所说的外推因子确定。目前，大多数 TRX 实验都是通过外推因子确定来区分状态混合物。由于这个问题本质上欠定，任何解决方案都必须引入额外假设，以确定 α 以及$\mathbf{F}_{\text{mix}}$ 的相位——或者直接确定$\mathbf{F_1}$ 。

近期有一种方法完全绕过占有率估算。该方法沿着基于物理规律的反应坐标枚举候选变化，并根据每个候选值计算出的差值图与观测到的差值图的吻合程度为其评分（Fong等，2026）。这种方法仅适用于有限区域，且必须预先指定该区域及其反应坐标。更通用的途径，也是本文的重点，是直接估算外推因子。已有多种外推因子估算方法，每种方法都采用务实的物理假设来解决数据固有的模糊性（Pearce 等，2017；Coquelle 等，2018；Tripathi 等，2012年）。然而，迄今为止，这些假设及其适用范围尚未得到充分评估。因此，用户无从判断哪种方法最适合自身情况，也不清楚可以期待何种精度。

鉴于目前缺乏对外推因子确定方法的全面综述与分析，同时我们也希望运用其中一种或多种方法来解析自己的 TRX 结构，因此系统地研究了迄今发表的所有方法。在此，我们介绍研究结果以及一种我们提出的新方法。首先，我们详细阐述占有率估算问题，包括几乎所有现有方法所依据的恒相位近似。随后，我们采用统一的符号体系，简要回顾目前使用的各类方法，并明确说明其假设。此外，我们在一个模拟数据集上对这些方法的表现进行基准测试，并展示我们的方法在两个各具代表性、差异显著的 TRX 数据集上的表现。

这些结果启发我们开发了一种新方法，称为真空能级匹配。该方法以一项物理约束为基础进行估算：电子密度不能低于真空密度。这种方法是基于非负性的占有率估算技术的演进，而这类技术如今已在使用中（Schmidt，2023）。真空能级匹配在单个体素的层面上运行，用数据驱动的截断值取代某些现有方法所需的手动拟合选择，并自动报告质量控制指标，以标记数据不足以支持可靠预测的情况。我们将真空能级匹配作为一个小型、可复用的软件包提供，以便直接应用于新的数据集。这些结果应对所有分析 TRX数据的人有所助益。

2 结构因子外推及相位的作用

上文已说明，将触发态和参考结构的结构因子振幅分离是一个欠定问题。自然的做法是从混合态中减去某种参考信号，只留下我们感兴趣的触发信号。本文将详细回顾这一思路。

2.1 具有完整相位信息的外推

为清晰起见，我们从最简单的情况开始。假设除了通常测量的振幅之外，我们还准确估计了相位，因而获得了参考态 $\mathbf{F_0}$ 和混合态 $\mathbf{F}_{\mathrm{mix}}$的复结构因子。混合态结构因子为

$$\mathbf{F}_{\mathrm{mix}} = (1-\alpha)\mathbf{F_0} + \alpha\mathbf{F_1}, \tag{3}$$

将其重新整理，可得到触发态结构因子，它是参考态结构因子和复差结构因子 $\Delta\mathbf{F} = \mathbf{F}_{\mathrm{mix}} - \mathbf{F_0}$,的线性组合

$$\mathbf{F_1} = \mathbf{F_0} + \Delta\mathbf{F}/\alpha. \tag{4}$$

将其视为一个外推过程，通过对后者按 $1/\alpha$进行缩放，便可根据参考态和差分结构因子恢复 $\mathbf{F_1}$ 。

如果通过某种合适的过程，我们能够推断出一个 外推因子 χ，使得 $\chi \approx \alpha$，那么就可以计算出复外推结构因子

$$\mathbf{F}_\chi = \mathbf{F_0} + \chi^{-1}\,\Delta\mathbf{F} \tag{5}$$

作为$\mathbf{F_1}$的最优、最低误差估计。在获得完整复数信息的情况下， χ简单地等于 α。

公式 3 继承了引言中讨论的一般线性模型的假设，并增加了一条：它将随时间演变的中间态 $\mathbf{F_1}$ 视为单一纯态，尽管在一般时间延迟下，系统可能处于态的混合中。此外，TRX 实验通常以串行方式进行，因此合并晶体之间的占有率可能有所不同；当这种方差较大时，应将其纳入考虑，具体如附录B所述。

2.2 仅使用参考态相位的外推

虽然在上一节中我们假设已知 $\mathbf{F}_{\mathrm{mix}}$的相位，但实际上这些相位无法测量。 公式 5 因此无法直接使用。几乎所有当前方法所依据的实用替代方案，是用已知的 $\Delta\mathbf{F}$ 的相位替代未知的 $\mathbf{F_0}$,

$$\Delta\mathbf{F} \;\rightarrow\; \Delta|F|\, e^{i\varphi_0}, \tag{6}$$

其中 $\Delta|F| = F_{\mathrm{mix}} - F_0$ 是观测振幅的同晶型差值，而 φ_0 是参考态相位。在此参考相位近似下，方程 5变为振幅之间的关系，

$$F_\chi = F_0 + \chi^{-1}\,\Delta|F|. \tag{7}$$

Genick（2007）详细分析了这种用于结构因子外推的标量近似所存在的误差和局限性。

这种近似的准确程度如何，在这种情况下 χ 的最佳选择是什么？代入 公式6 将测得的振幅差与参考态相位结合起来。对于复数差值相对于参考振幅较小的反射， $|\Delta\mathbf{F}| \ll F_0$，一阶近似给出（Luzzati，1952；Henderson 和 Moffat，1971）

$$\Delta|F|\, e^{i\varphi_0} \approx \frac{1}{2}\Delta\mathbf{F} + \frac{1}{2}\Delta F\, e^{i2\varphi_0 - i\varphi_\Delta}, \tag{8}$$

其中 φ_Δ 是 $\Delta\mathbf{F}$ 的真实相位（有关包含晶体间方差的推导，见 附录B ）。因此，同晶型差值包含了以一半尺度表示的真实差值，以及携带相位 $e^{i2\varphi_0 - i\varphi_\Delta}$的复共轭项。为了解释这一分解，我们进一步近似认为，共轭贡献在整个晶胞中是弥散的，并且在包含结构变化的区域内可以忽略。该行为并不能由分解本身保证：参考相位与差值相位之间的相关性可能会留下结构化残差。

Su将 公式8 代入 公式7，再结合 ΔF的定义，可得 ds

$$F_\chi\, e^{i\varphi_0} \approx \frac{\chi^{-1}\,\alpha}{2}\mathbf{F_1} + \left(1 - \frac{\chi^{-1}\,\alpha}{2}\right)\mathbf{F_0} + \frac{\chi^{-1}}{2}\,\Delta\mathbf{F}^* e^{i2\varphi_0}. \tag{9}$$

最后一项是上文讨论的共轭贡献。在假设该贡献是弥散的，并且在包含结构变化的区域中可以忽略不计的情况下， $\chi = \alpha/2$ 可将相干差值信号恢复至其完整振幅。在这里，我们所说的“最优”是指恢复局部结构特征的振幅，而不一定是使整个晶胞范围内的总重构误差最小化。后一个标准还会惩罚被放大的共轭贡献，因此可能倾向于采用不同的外推因子。

2.3 结构因子加权的影响

在当前的 TRX 分析中，通常会对差分结构因子应用标量权重 W_{SF} ，以减小离群值的影响并提高所得映射中的信噪比（Ursby & Bourgeois，1997；Ren等，2001；De Zitter等，2022）。对于每个反射， W_{SF} 作为 ΔF上的额外前因子，其通常取决于测量误差以及 F_0 和F_{mix}的绝对值，因此 方程 9变为

$$F_\chi\, e^{i\varphi_0} = \frac{\chi^{-1}\,W_{\mathrm{SF}}\,\alpha}{2}\mathbf{F_1} + \left(1 - \frac{\chi^{-1}\,W_{\mathrm{SF}}\,\alpha}{2}\right)\mathbf{F_0} + \frac{\chi^{-1}\,W_{\mathrm{SF}}}{2}\,\Delta\mathbf{F}^* e^{i2\varphi_0}. \tag{10}$$

我们将该因子 $1/2$ 与 W_{SF} 分开，以区分参考相位近似与加权方案的影响，而在公式12中，两者则合并在一起。此时，最优外推因子为

$$\chi = \frac{1}{2}\,\alpha\,W_{\mathrm{SF}}. \tag{11}$$

这使得直接确定占有率变得更加复杂，主要体现在两个方面。首先，最优外推因子现在与所采用的具体加权方案相关。其次，χ 与 α 之间的关系不再是简单的正比关系，因为 W_{SF} 会在倒易空间中发生变化。

2.4 相位修正

如上文所述，真实的结构因子振幅 ΔF 是未知的，而实际测量的只有绝对振幅 $\Delta|F|$ 的差值。施加全变差去噪等约束的实空间精修技术（Fadini 等，2025）必然会同时改变振幅和相位。此外，实空间操作不一定能够保持晶体学对称性，从而可能在 F_{mix} 或F_0中对称性禁戒反射处引入非零结构因子。这些影响进一步加剧了估计结构因子与真实 $\Delta\mathbf{F}$之间关系的不确定性，也使 χ与 α之间的联系更加模糊。

一种补救方法是在实空间约束（例如全变差去噪）与已知结构因子 F_{mix} 和 $\mathbf{F_0}$ 之间进行迭代投影（Fadini 等，2025）。这可以改善差分结构因子的振幅和相位估计，但改善程度未知。因此，实际上得到的是带有复数权重参数的 $\chi = \alpha\mathbf{W}_{\mathrm{IP}}$。尽管迭代相位精修可能改善相位估计，但它对振幅和相位的影响都可能增加 χ 与 α之间关系的不确定性。

2.5 外推因子与占有率的关系

前文各节表明，最佳外推因子不仅取决于潜在的物理占有率，还取决于差值图的相位精度和后处理。因此，我们建议将外推因子和物理占有率视为彼此相关但不同的量，并在报告外推因子时同时说明处理流程的具体细节。基于此，我们更倾向于使用 外推因子确定 而非更常见的 占有率测定，因为这些方法测得的是外推因子，只有在第2.1小节 所述的理想化条件下，它才与物理占有率一致。

一般而言，外推结构因子可写为

$$\mathbf{F}_\chi = \left(\chi^{-1}\alpha\mathbf{W}\right)\mathbf{F_1} + \left(1 - \chi^{-1}\alpha\mathbf{W}\right)\mathbf{F_0} + \left(\chi^{-1}\alpha\mathbf{W}\right)\mathbf{N}_{\Delta F}, \tag{12}$$

其中，$\mathbf{W}$ 充当一个通用的傅里叶权重，用于捕获所应用的数据修改，包括来自参考相位近似的 $1/2$ 项。$\mathbf{N}_{\Delta F}$ 项则捕获差值图中的噪声和误差，这些噪声和误差既可能来自潜在的参考相位近似，也可能来自结构因子不正确或相位精修不完善所导致的其他误差。

不同于方程 9 和 方程 10，此处我们将 α 从该项中提取出来，并吸收一个补偿项$1/\alpha$ 到 $\mathbf{N}_{\Delta F}$的定义中，从而使 方程 16 中的实空间形式更加简洁。因此，必须了解并报告差值图计算流程的影响，才能对 χ 与 α之间的关系作出实质性判断。

3 估计外推因子

由于占有率测定是欠定的，每种估计占有率的方法都必须引入额外假设才能得到解。这些假设界定了各方法的适用范围及其失效模式。在本节中，我们将综述截至目前已发表的方法

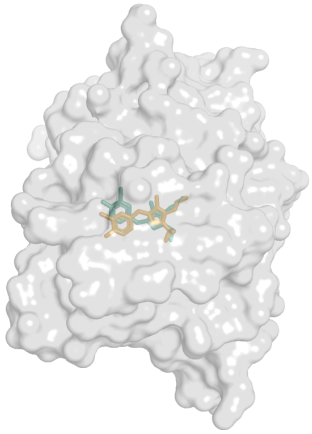

图1：**建模变化示意图**。这里展示了基于 PDB 条目 8A6G 的 rsEGFP2 从*trans* 参考态（绿色）变为*cis* 触发态（橙色）的过程。

，先分别按照各自的定义介绍这些方法，然后在 第4节中对它们进行实证基准测试。在整个讨论过程中，我们假设，除了复数参考结构因子 $\mathbf{F_0}$，之外，还可以获得差值图相位 $\mathbf{\Delta F}$的某种估计，从而能够确定实空间对应量 ρ_0 和 $\Delta\rho$ 以及类似对象。尽管我们将这些密度写成实空间坐标 $\mathbf{r}$的连续函数，但在实践中，每个密度都会在离散网格上采样，我们将网格中的单个采样点称为体素。只有在指代特定点处的值（例如 $\Delta\rho(\mathbf{r})$）时，我们才明确写出其对 $\mathbf{r}$ 的依赖；除此之外则将其省略。

我们首先介绍一些早期 TRX 研究中出现过的方法（第3.1节），然后转向近期工作中最广泛使用的方法：PanDDA（第3.2小节）、差值图匹配（第3.3节）和负和爆炸（第3.4小节）。后者对我们尤其重要，因为它建立在一个简单明晰的物理约束之上。在第3.5小节中，我们直接在此基础上引入真空能级匹配作为我们自己的改进：它保留了相同的负密度原理，但不再对整个图谱进行聚合，而是逐体素求解估计值。在下一节中，我们将在模拟数据集上测试我们的方法以及其他较新的方法。通过这些模拟，我们选取其中一次运行结果，并以此为例说明每种方法。为实现逼真的模拟，我们根据先前一次时间分辨串行飞秒晶体学（TR-SFX）实验所确定的 rsEGFP2 的*cis* 和*trans* 发色团结构计算信号（Fadini等人，2023年），使用的模拟占有率为 $\alpha_{sim} = 0.4$，并采用参考相位近似（目标外推因子 $\chi = 0.2$）。在本节中，我们将展示一个具有中等噪声的代表性示例。有关模拟设置的更多细节，请参见第7.1小节。模型及其变化见图1。

3.1 基于模型和体积的方法

有三种方法早于我们将详细进行基准测试的方法，现列于下文以求完整。每一种方法都已成功应用于基础性时间分辨晶体学研究，并且各自体现了

一种为外推因子估计提供依据的独特方式。我们不将它们纳入定量比较。B 因子匹配和距离分析需要一个在多个候选外推因子下进行精修的触发态模型，因此无法像其他方法一样，在完全基于图谱的统一标准下进行评估；而积分差电子密度则需要手动确定积分体积，而这一体积无法推广到我们的测试数据集。

3.1.1 积分差值电子密度

估计外推因子的一种方法是，在 $\Delta\rho$ 中识别出一片对应于完整且边界清晰的密度特征的密度区域，例如配体，在 ρ_{mix}中。该物种的理论电子数 N_{e} 可根据其化学组成确定。在将图谱置于绝对标度后，随后对该区域 V内的差值密度进行积分。选择外推因子，使外推差值密度 $\chi^{-1}\,\Delta\rho$ 的积分等于完整的电子数 N_{e}，即

$$\chi = \frac{\left|\int_V \Delta\rho \,\mathrm{d}\mathbf{r}\right|}{N_{\mathrm{e}}}, \tag{13}$$

从而外推图包含完整的物种。无需对参考态相位进行显式校正，因为相干信号损失一半（第2.2小节）已经反映在测得的积分值中。值得注意的应用包括测定肌红蛋白中的一氧化碳光解产率（Srajer 等，1996）以及估计光活性黄色蛋白中的光引发水平（Schmidt 等，2004）。

主要困难在于界定精确的体积边界。这需要找到一个没有重叠信号的区域，而这种区域很难可靠地识别，尤其是在占有率较低或噪声较高的情况下。即使差异特征明显且界定清晰，目前仍必须手动绘制边界，并依赖人工判断。

3.1.2 B 因子匹配

该方法利用了晶体学 B 因子与结构精修过程中外推因子之间的强相关性。它假设，触发后发生移动的原子的 B 因子（作为局部无序程度的度量）与其相邻的未发生变化的原子的 B 因子不会存在显著差异。由于占有率和 B 因子都会影响原子散射贡献的大小，因此，所赋予占有率中的误差通常会被人工引入的 B 因子偏移所吸收。为了确定正确外推因子，需要针对不同外推水平下的数据对模型进行精修。最优外推因子通过寻找这样一个点来确定：扰动后发生移动的原子的 B 因子与周围静态蛋白质环境中的 B 因子相匹配。

Nango 等（2016）成功采用了这一方法来确定细菌视紫红质中的结构变化。在他们的实现中，他们指出，由于参数耦合，在当时的分辨率下同时精修 B 因子和外推因子并不稳定。因此，他们在约束 B 因子的同时，对触发态构象的坐标和外推因子进行精修。为得出每个时间点的最终共识外推因子，他们根据表现出最大结构位移的分组残基计算加权平均。

3.1.3 距离分析

距离分析的核心原理由 Genick 等（1997）首次提出。在精修过程中，错误假定的外推因子会在精修模型中产生结构伪影。当假定的 χ^{-1} 过低时，外推无法完全去除参考态贡献，精修后的原子位置仍会偏向参考态。当假定的 χ^{-1} 达到或超过真实值时，原子位置预计会趋于稳定。Xtrapol8（De Zitter 等，2022年）将这一观察转化为系统性流程，并称之为“距离分析”。

在此工作流程中，使用一系列假定外推因子生成的外推结构因子来精修模型。随后，将相对于参考模型的原子间距离变化拟合为关于 χ^{-1}的 S 形函数。对于每个通过筛选的拟合，提取距离变化达到其拟合平台值 99 % 时对应的 χ^{-1} 值。汇总这些成对估计值，得到共识估计值。该判据用于确定进一步外推几乎不会带来额外结构位移的范围。与此处介绍的其他方法不同，距离分析要求在每个测试的外推因子下，分别在倒易空间和实空间中精修模型。

3.2 比较局部与全局相关性

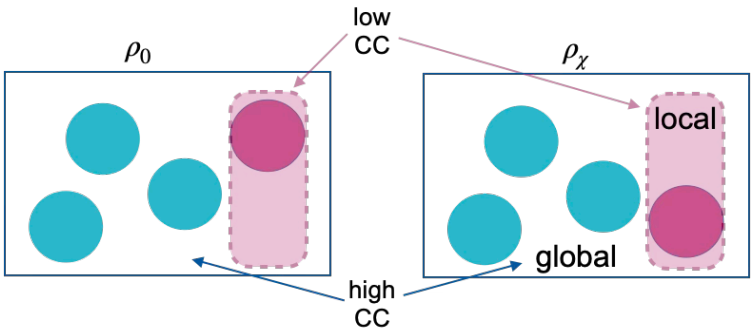

图 2：**PanDDA 方法的原理**。 我们展示了参考态 ρ_0 和外推态 ρ_χ中的一个含有四个原子的模型分子。三个原子（青绿色）保持原位，而第四个原子（深红色）表示局部结构变化。该方法在感兴趣区域内计算 ρ_0 和 ρ_χ之间的全局和局部互相关（CC），并遍历许多不同的 χ值。在最优外推因子下，全局 CC 应较高，而局部区域中的 CC（虚线）应较低。该方法将两种 CC 差值最大的 χ 值选为最优值。

PanDDA（Pearce 等，2017年）是一款广泛用于基于片段的药物发现的软件工具，其用途是在晶体学数据集中检测弱结合配体。该工具中的一个程序用于确定背景密度分量，在实际操作中，这相当于 $1-\chi$。尽管 PanDDA 及其用于确定背景分量的程序都并非为时间分辨数据而设计，但同一程序——俗称“PanDDA 方法”——后来已应用于 TRX 研究中（Fadini 等，2023年）。

核心思想是最大限度地提高外推图在实空间中的对比度，使得

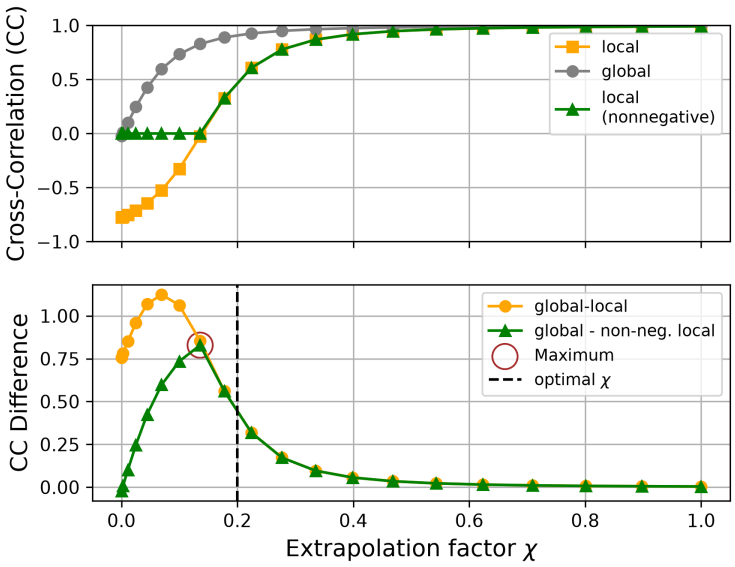

图3：**“PanDDA 方法”的实际应用示例**。上图显示了全局（灰色）、局部（橙色）和非负局部（绿色）相关性随 χ的变化，其中仅纳入满足 $|\Delta\rho(\mathbf{r})| > 3\sigma_\Delta$ 的体素。下图显示了全局相关性与局部（橙色）或非负局部（绿色）相关性之间的差异。结果基于单个模拟数据集，其预期 $\chi = 0.2$（使用 $\alpha_{sim} = 0.4$和参考相位近似进行建模），信噪比为 15，详见第7.1小节。

变化区域的密度差尽可能大。该程序首先识别发生变化的局部感兴趣区域。整个晶胞被视为全局区域。然后，对于一系列 χ，分别计算局部区域和全局区域中外推图 $\rho_\chi = \rho_0 + \chi^{-1}\,\Delta\rho$ 与参考图 ρ_0 的互相关。局部感兴趣区域可以通过多种方式确定。一种简单的方法是对绝对差值图进行阈值筛选，选取数值超过差异图均方根偏差三倍的所有体素（即 $|\Delta\rho(\mathbf{r})| \geq 3\,\sigma_\Delta$）。

两个互相关的行为随 χ变化而改变，符合直觉。当 χ较小时， ρ_χ 主要由$\Delta\rho$决定，两个互相关都趋近于零。当 χ较大时， ρ_χ 与 ρ_0十分相似，全局和局部互相关都接近于 1。由于差值图在局部区域内“差异最大”，局部互相关只有在 χ 大于全局互相关时才会上升（见图2）。

全局互相关与局部互相关之间的差异最大值被提出作为正确外推因子的估计量（Pearce 等，2017）。该方法的应用示例见图3。

适用于时间分辨数据 将该程序应用于 TRX 而非配体结合数据时，会出现一个复杂问题。在 PanDDA 所针对的配体结合场景中，局部差异信号主要为正，这是因为配体原子占据了先前为空的空间。相比之下，在 TRX 场景中，只要原子离开其参考态位置，差异信号就会出现明显的负特征。因此，在较小的 χ 下，局部互相关不仅不存在，反而为负，而局部差异与全局差异之间的轨迹也相对于该启发式方法最初校准时所处的状态发生了偏移。

一种简单的修改方法是将互相关限制在非负体素值范围内，从而抑制反转的起始点。我们将其称为非负 *PanDDA* 变体。

PanDDA 程序的优点在于，它直接处理图谱之间易于计算的互相关。与其他方法相比，该方法有两点值得注意。最大差异准则是一种启发式方法。PanDDA 的原始论文指出，该准则与最终的占有率估计相关，但存在不可忽略的方差（Pearce等，2017）。这种关系是经验性的，并非源于对外推图的独立物理约束。对于原始片段筛选领域之外的数据，该准则的准确性尚未得到系统表征。

此外，结果对局部区域掩膜十分敏感。不同的过滤选择会产生不同的掩膜，进而导致不同的局部互相关，这可能改变最大值的位置，从而影响估计的 χ。可以确定的是，最佳外推因子位于全局与局部差异明显非零的范围内：当全局和局部互相关相近时，这表明外推不足（两者都较高）或外推过度（两者都较低）。但无法定量保证报告的最大值比该范围内的任何其他点更接近真实估计。

我们测试了标准版本和非负版本，但在第4节中仅展示非负版本。如前所述，标准版本由于局部区域中的负相关，往往会产生过度外推。

3.3 差值图匹配

差值图匹配方法最早由 Coquelle 等人（2018年）提出，修订后的版本已在软件包 Xtrapol8（De Zitter 等人，2022年）中实现。对于每个候选外推因子，该方法会对处理并归一化后的观测差值图和外推差异图进行比较。使两幅图谱达到最接近一致的数值被确定为最优外推因子。

在 Xtrapol8 中，该技术有两种形式，我们根据各自发现信号的方法，将其分别称为“峰积分方法”和“CC方法”。这两种方法都已应用于近期的多项 TRX 研究（Hadjidemetriou等人，2022年；Christou 等人，2023年；Rios- Santacruz等人，2026年）。

此方法的核心在于两个量：给定的外推结构与计算得到的参考态之间的差异，即 $(\Delta\rho)_\chi = W_\chi(\rho_\chi - \rho_0^{\text{calc}})$，以及实验“$F_o - F_o$”图谱，即观测到的差异 $(\Delta\rho)_{\text{obs}} = W_{\text{obs}}(\rho_{mix} - \rho_0)$，其中 W_i 表示对每个差异密度图所应用的线性后处理 i。两种差值图匹配的目标都是找到使 $(\Delta\rho)_\chi$ 最接近 $(\Delta\rho)_{\text{obs}}$的 χ 。两种方法都量化了 $(\Delta\rho)_\chi$ 与 $(\Delta\rho)_{\text{obs}}$之间的差异。为了解所发现的差异，我们将 $(\Delta\rho)_\chi$ 改写为 $(\Delta\rho)_{\text{obs}}$的函数，

$$(\Delta\rho)_\chi = \chi^{-1}(\Delta\rho)_{\text{obs}} + \chi^{-1}\delta_\Delta + \delta_0, \tag{14}$$

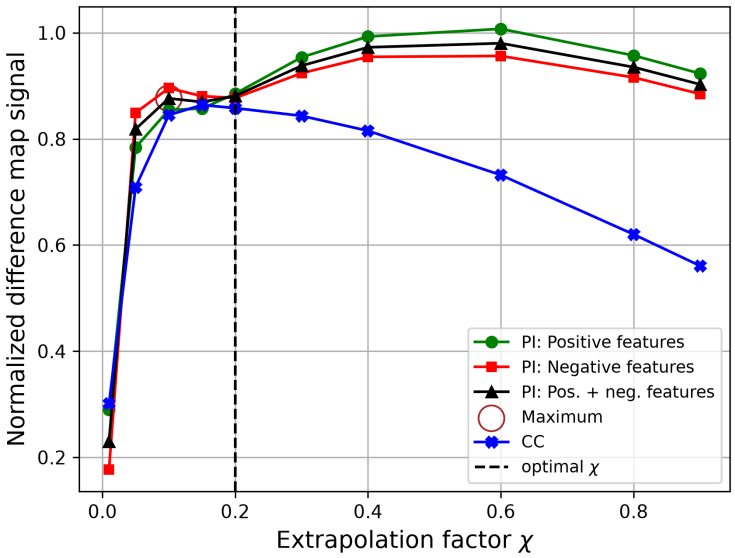

图4：**“差值图匹配”的应用示例**。 将差值图匹配应用于模拟的 rsEGFP2 数据集（目标 $\chi = 0.2$）。曲线显示峰积分（PI）方法在正、负及组合特征上评估得到的归一化差值图信号随 χ 变化的情况，同时还显示了互相关（CC）方法的结果。峰积分比在较大的 χ 处也会上升（不足外推），但要求外推差值图与观测差异图在绝对值最大的体素处符号一致，即可排除这一虚假的最大值（$\Delta\rho$）$_{\text{obs}}$（第3.3小节）。对于该数据集，峰积分方法（所有特征）得到 $\chi = 0.10$，CC方法得到 $\chi = 0.15$。

其中， $\delta_0 = W_\chi(\rho_0^{\mathrm{obs}} - \rho_0^{\mathrm{calc}})$反映观测参考图谱与计算参考图谱之间的失配——即参考态的传统“$F_o - F_c$”图谱。表达式 $\delta_\Delta = (W_\chi - W_{\mathrm{obs}})(\rho_{mix} - \rho_0)$则涵盖了分别执行“后处理”步骤所造成的差异。对于固定的后处理权重， $\chi^{-1}\delta_\Delta$ 会随外推程度增加，而 δ_0 与之无关；候选因子依赖的后处理可能会使这两个失配项都产生额外变化。这两种方法都依赖于不可忽略的失配项，因为正是其依赖于 χ的贡献相对于 $\chi^{-1}(\Delta\rho)_{\mathrm{obs}}$ 的大小，形成了不同候选外推因子之间的对比。

峰积分方法。 该方法通过比较观测差异信号最显著的位置，检验外推差异图能否良好地再现观测差异信号。首先，从观测差异图（$\Delta\rho$）$_{\mathrm{obs}}$ 中选出一部分显著体素$\mathbf{r}^{\mathrm{signal}}$ ，方法是选择最大峰，也就是绝对值较大的区域（默认$> 3.5\sigma_\Delta$ ）。然后，分别用各自的均方根偏差对（$\Delta\rho$）$_{\mathrm{obs}}$ 和（$\Delta\rho$）$_\chi$ 进行归一化。这一步至关重要：在绝对标度下，即使发生过度外推，$\mathbf{r}^{\mathrm{signal}}$ 处的积分密度仍会线性增长，因此原始总和无法提供有效信息。由于过度外推会提高所有体素处的密度，而不仅仅是$\mathbf{r}^{\mathrm{signal}}$处的密度，因此归一化积分在最优值两侧都会下降。接下来，在该固定掩模上对（$\Delta\rho$）$_\chi$ 进行积分，分别计算正峰和负峰（根据（$\Delta\rho$）$_\chi$的符号指定），并计算二者合并后的结果。每个通道（$\pm$）的得分为

$$\mathrm{S}(\pm) = \frac{\sum_{\mathbf{r}_{\mathrm{signal}}} \left|(\Delta\rho)_\chi(\mathbf{r}_{\mathrm{signal}}, \pm)\right|}{\sum_{\mathbf{r}_{\mathrm{signal}}} \left|(\Delta\rho)_{\mathrm{obs}}(\mathbf{r}_{\mathrm{signal}}, \pm)\right|}. \tag{15}$$

所有三个符号通道 $\pm$（仅正值、仅负值和合并）均会显示，但最优 χ是根据合并后的（正值加负值）绝对值总和选出的。

根据构造，在（$\Delta\rho$）$_{\mathrm{obs}}$ 中，大部分密度集中在 $\mathbf{r}^{\mathrm{signal}}$，而差值图处理 $\delta_\Delta \propto W_\chi - W_{\mathrm{obs}}$ 中的差异以及观测参考图与计算参考图 $\delta_0 \propto \rho_0^{\mathrm{obs}} - \rho_0^{\mathrm{calc}}$之间的差异则分布在整个图谱中。此指标用于寻找在 $\mathbf{r}^{\mathrm{signal}}$处浓度最高的外推差异图。在过度外推和欠外推区间，相较于 $\mathbf{r}^{\mathrm{signal}}$ 内部，预计会有更多密度累积在其外部。由于符号通道仅根据（$\Delta\rho$）$_\chi$ 指定，而不参考（$\Delta\rho$）$_{\mathrm{obs}}$，因此可能出现伪第二极大值，通常位于较大的 χ处，此时掩模内的密度特征已发生反转（图4）。为排除该极大值，我们检查在（$\Delta\rho$）$_{\mathrm{obs}}$中绝对值最大的体素处，（$\Delta\rho$）$_\chi$ 与（$\Delta\rho$）$_{\mathrm{obs}}$ 是否符号相同；只有符合这一条件时，才接受该极大值。

互相关方法。 第二个程序将外推差异图（$\Delta\rho$）$_\chi$ 与观测差异图（$\Delta\rho$）$_{\mathrm{obs}}$ 直接进行比较，计算二者的皮尔逊互相关。该相关性衡量两幅图谱的均值中心化空间模式有多么一致；失配项的作用推导见 附录 C。由于它考察的是全局，因此不会将信号峰与噪声峰区分开来。因此，噪声峰也会对相关性产生影响，所以溶剂或其他变化较小区域中的随机差异可能会使相关性产生偏差，尽管这些差异不包含任何结构信息。

峰积分方法在传统差异图中只有少量

彼此分离良好且十分显著的峰时效果最佳，因为此时失配项会分散到图谱的其余部分，并与$\mathbf{r}^{\mathrm{signal}}$处的信号形成鲜明对比。当密度弥散，或信号峰与噪声峰的显著程度相近时，该方法的表现就会受限。CC方法对整个图谱进行汇总，不受峰选择方案影响；不过，最重要的信息包含在极值中，而峰积分方法能够捕捉这些信息。

当两种子方法产生分歧时，正如我们的示例（图4）所示，由于两者都没有量化其置信度，因此没有原则性依据来选择其中之一。另一个实际成本是，每个采样的 χ 按照目前的形式，都需要分别进行后处理（针对每个$\Delta\rho$）x，与所考察的其他方法相比，这使精细采样的成本更高。

我们在第4节中评估峰值积分和CC方法。

3.4 负和爆炸

接下来，我们讨论一种最早由Tripathi等（2012）采用、并在Schmidt（2023）中进一步完善的技术。据我们所知，这项技术尚未被赋予名称，因此我们将其称为“负和爆炸”方法（NSE）。NSE已用于近年的多篇出版物中（Pandey等，2020；Baxter等，2024；Maestre-Reyna等，2025）。该方法基于一条物理约束，即任何电子密度图都应主要为非负值，只有少量负信号源于实验误差和傅里叶截断。相比之下，过度外推（$\chi < \chi_{\mathrm{opt}}$）会在外推图$\rho_\chi(\mathbf{r}) = \rho_0(\mathbf{r}) + \chi^{-1}\Delta\rho(\mathbf{r})$中 $\Delta\rho(\mathbf{r})$为负的区域产生负密度（参见图（5c））。不足外推（$\chi > \chi_{\mathrm{opt}}$）意味着，差值图中的负密度被参考图中的正密度完全抵消后仍有余量，因此$\rho_\chi(\mathbf{r})$中的负密度水平保持较低，即使 χ^{-1} 略微增加，这种情况也不会改变（参见图（5a））。因此，估计外推因子的一种方法是确定$\rho_\chi(\mathbf{r})$中开始积累大量负密度的点（即图（5b））。

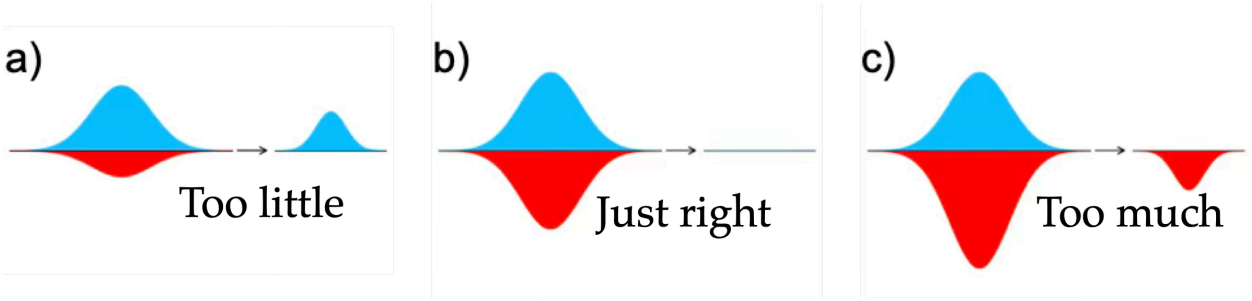

图5：**NSE 原理示意图**。此图（参考 Schmidt（2023）的图8）展示了箭头左侧的正参考密度（蓝色）与负差值密度（红色）的三种示意叠加，以及箭头右侧由此得到的外推密度。a）展示了不足外推（“过少”）的示例，此时外推密度仍为正值。b）展示了正确外推（“恰到好处”）的示例，此时外推密度接近零。c）展示了过度外推（“过多”）的示例，此时外推密度变为负值。

具体而言，可针对一系列 χ值计算 $\rho_\chi(\mathbf{r})$，并绘制所有负体素之和

相对于 χ^{-1} 的图。在欠外推区间（χ较大且 χ^{-1} 较小）中，大多数体素满足$\rho_0(\mathbf{r}) > -\chi^{-1}\Delta\rho(\mathbf{r})$，因此 $\rho_\chi(\mathbf{r})$保持为正，负和基本保持平坦。在过度外推区间（χ 较小且 χ^{-1} 较大）中，大多数具有负 $\Delta\rho(\mathbf{r})$ 的体素满足$\rho_0(\mathbf{r}) < -\chi^{-1}\Delta\rho(\mathbf{r})$，负和以斜率 $\sum_{\mathbf{r}:\ \Delta\rho(\mathbf{r})<0} \Delta\rho(\mathbf{r})$ 线性增长。因此，可以预期存在两个近似线性区间，其间由一个很短的过渡区分隔。对每个区间分别拟合线性回归，并将它们的交点估计为外推因子。

以下两点值得注意：

1. **零频项**。该方法的相关文献未提及一个重要问题，即对缺失的零频项 F_{000}进行校正。这个“直流分量”是图谱中每个体素的恒定加性密度贡献。结构因子通常在不包含 F_{000} 的情况下报告，因此重建的实空间图谱的均值为零，而不是以真空为参考的真实电子密度均值。这会使 $\rho_0(\mathbf{r})$ 被一个全局常数人为降低，导致 $\rho_\chi(\mathbf{r})$ 的表观零交叉点发生偏移，并造成对 χ^{-1} 的系统性低估。不过，现有工具可以在给定原子模型的情况下估计 F_{000} （Lang 等，2014），其中包括蛋白质和溶剂贡献。在分析前应用这些工具可以避免这一偏移。

2. **聚焦于感兴趣区域**。一些 $\Delta\rho(\mathbf{r})$ 值因伪影（噪声、系统误差）而为负，而另一些则因真实的结构变化而为负。NSE 依赖于后者占主导地位。然而，通过聚焦于感兴趣区域，我们可以减轻伪影的影响，因为结构变化通常局限于空间中的特定区域（Schmidt，2023年）。

图6 展示了校正这两个因素对我们模拟数据集的影响（详情见第7.1小节）。将求和范围限制为满足 $\Delta\rho(\mathbf{r}) < -3\,\sigma_\Delta$ 且应用 F_{000} 校正的体素，能显著缩小过渡区域，而两者的交点则得到了一个接近真实值的估计。

NSE 建立在一个明确的物理假设之上，但其可靠性取决于不足外推区间与过度外推区间之间是否存在清晰的过渡。通过谨慎选择 ρ_χ 的输入体素，可以使这一过渡更加清晰——聚焦于感兴趣区域，并剔除 $\Delta\rho$ 中数值较小的部分；随着差值图质量下降，这种筛选也愈发重要。

3.5 真空能级匹配

这些观察促使我们开发了真空能级匹配，这是对 NSE 方法的一种改进。目标仍然是确定负差值密度抵消参考密度的位置。我们不再对整张图谱进行汇总，而是逐个体素评估这种抵消情况。这样便可以评估每个逐体素估计的可靠性，并仅合并符合下文所述筛选标准的估计。

该方法依赖于参考态中至少有一小部分密度被真空取代。对于蛋白质内部发生的变化，这是一个合理的假设，因为溶剂无法进入，或被固定在某个位置。如果配体结合时没有取代参考结构中的任何原子，则此假设不适用。在蛋白质表面及其外部，空出的区域通常会被体相溶剂重新填充，因此密度会降至溶剂能级，而不是降至真空。不过，

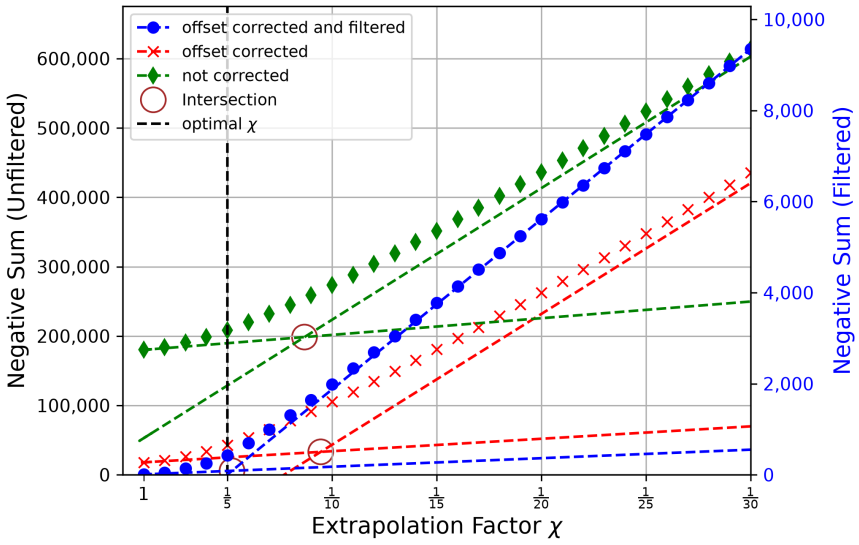

图6：**“负和爆炸”实例演示。** 负和ρ_χ 随反转后的外推因子变化的情况，数据来自一个模拟数据集（参见 第7.1小节），其真实值为 $\chi = 0.2$，展示了三种情形。圆点表示计算值，虚线表示不足外推和过度外推的拟合。第一种情形（绿色）对 ρ_χ 整个晶胞求和，且未进行 F_{000} 校正。第二种情形（红色）对整个晶胞求和，并应用了F_{000} 校正。第三种情形（蓝色）应用 F_{000} 校正，并将求和范围限制在 $\Delta\rho\,(\mathbf{r}) < -3\,\sigma_\Delta$的体素内。所得数值显著较小，因此使用单独的刻度（右轴）绘制。线性拟合在固定范围内进行 $1 \le \chi^{-1} \le 2$ （某些点未绘出，因为它们呈单调递增），且 $100 < \chi^{-1} < 200$，这些范围选得远离任何合理的 χ，以避免在选择过渡区域时产生偏差。交点给出的 χ 值分别为 0。12、0。11 和 0。19，对应上述三种情形。

只要蛋白质内部也发生了一些变化，该方法仍能确定最优外推因子，正如我们在小节3.5.1中所论述的。

3.5.1 局部估计

要找出发生过度外推的位置，应聚焦于这样一些体素：它们对应的位置在参考态中存在原子，而在触发态中已不再存在。因此，我们的首要目标是得出逐体素估计，确定在哪些外推因子下参考密度会被完全消除。

为此，我们从公式12中的外推图一般表达式入手。现在，我们改用实空间密度相加的形式 ρ，而非复结构因子$\mathbf{F}$，得到

$$\rho_\chi(\mathbf{r}) = \chi^{-1}\alpha w\,(\rho_1(\mathbf{r}) - \rho_0(\mathbf{r}) + n_\Delta(\mathbf{r})) + \rho_0(\mathbf{r}), \tag{16}$$

其中，我们将处理的影响建模为两部分：一个全局标量重缩放因子 w （主要来自差值图处理，包括参考相位校正前因子）以及由噪声或误差 $n_\Delta(\mathbf{r})$导致的逐体素偏差。

从上面的表达式中可以看出，当 $\chi^{-1}\alpha w = 1$ 时，我们可以消除参考密度的贡献。对于每个体素而言，该表达式都变成线性表达式，可以轻松求解。与 NSE 的情况一样，我们希望关注尽可能多地消除参考密度的情况。因此，我们要求 $\rho_\chi(\mathbf{r}) = 0$，现在即可得到每个体素的外推因子估计值。

$$\tilde{\chi}(\mathbf{r}) = \alpha\, w \left(1 - \frac{\rho_1(\mathbf{r})}{\rho_0(\mathbf{r})} - \frac{n_\Delta(\mathbf{r})}{\rho_0(\mathbf{r})}\right) = \frac{-\Delta\rho(\mathbf{r})}{\rho_0(\mathbf{r})}. \tag{17}$$

大多数单点估计都有偏差。只有在参考密度同时大于两个混杂项的体素处，单个体素的估计才可靠，也就是说，在 $\rho_1(\mathbf{r}) \ll \rho_0(\mathbf{r})$ 和 $|n_\Delta(\mathbf{r})| \ll \rho_0(\mathbf{r})$ 的情况下。如果不满足这一条件，估计就会产生系统性偏差：对于 $\rho_1(\mathbf{r})$ 会偏向过度外推（图 7a），而噪声造成的偏差则可能朝任一方向。

体相溶剂回填是这种偏差的一种特殊情况。空置体素的 $\rho_1(\mathbf{r}) = 0$，而由溶剂回填的体素则有 $\rho_1(\mathbf{r}) = \rho_{solvent}$，这会使其估计值降低一个因子（$1 - \rho_{solvent}/\rho_0(\mathbf{r})$）。关键是，回填体素的差值密度也没那么负，因为 $\rho_0(\mathbf{r}) - \rho_{solvent} < \rho_0(\mathbf{r})$。由于体素按 $\Delta\rho(\mathbf{r})$ 排序，最负者排在最前（小节 3.5.2），空置体素会最先进入估计，并占据接受窗口的主导地位。较早进入的回填体素只能处于较大的 $\rho_0(\mathbf{r})$ 值，此时 $\rho_{solvent}/\rho_0(\mathbf{r})$ 较小，因此其偏差也很小。任何不完全或不充分的回填都介于这两个极限之间，因此即使溶剂模型并不完全准确，估计仍大致正确。

3.5.2 朝着单一估计及其可靠性迈进

公式（17）使我们能够计算出一个外推因子，使外推图与真空水平重合——该量会因体素而异。接下来，我们要找出一组体素，以便据此得到一致的全局外推因子估计值。该组体素应包含差值图中的一个区域，其中有真实的负信号，且其强度高于任何背景噪声。

在图 7中，我们绘制了 $\tilde{\chi}$（$\mathbf{r}$）随 $\Delta\rho$ 变化的曲线，涵盖多种理论情形，以展示逐体素估计的外推因子如何随差值图中该体素的负向程度而变化。总体而言，对于负值足够大的差值图体素，不同的估计值 $\tilde{\chi}$（$\mathbf{r}$）会围绕单一值形成窄分布。使用参考相位会使最佳外推因子减半，并引入噪声。参考振幅中的恒定误差，其相对强度会随着 $-\Delta\rho$减小而增大。此外，我们在公式 17中证明，较大的参考密度会带来较低的混杂因子。

因此，选择哪些体素纳入用于定义全局外推因子的集合，涉及内在的偏差-方差权衡。纳入更多体素可以抑制离群值的影响，但公式 17表明，纳入具有显著 $\rho_1(\mathbf{r})$ 的体素会引入一种偏向过度外推的系统偏差。

为平衡这些考量，我们采用以下启发式程序。我们按照$\Delta\rho(\mathbf{r})$对体素排序，从最负值开始，并沿此顺序计算$\tilde{\chi}(\mathbf{r})$的累积均值和标准差。我们将参考密度截断值 ρ_0^{cut}设为体溶剂密度，由此确定逐体素估计图中的对角边界 $-\Delta\rho(\mathbf{r})/\rho_0^{\text{cut}}$。当累积

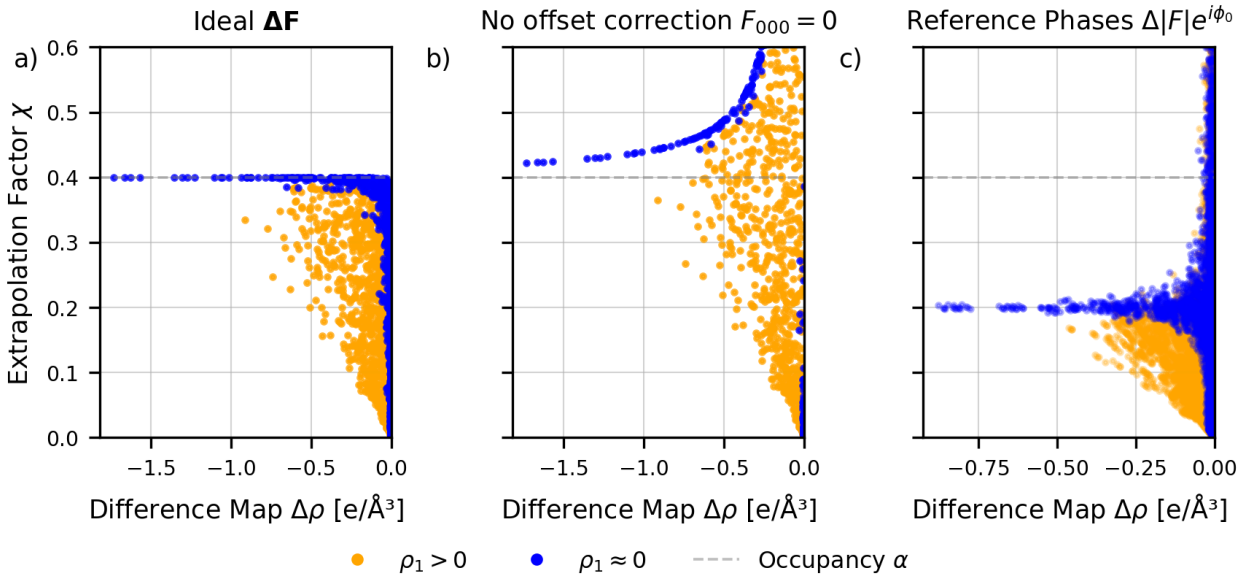

图 7：**在不同条件下 $\tilde{\chi}(\mathbf{r})$ 的理想化示意图**。逐体素估计$\tilde{\chi}(\mathbf{r}) = -\Delta\rho(\mathbf{r})/\rho_0(\mathbf{r})$相对于 $\Delta\rho(\mathbf{r})$作图，所用模拟数据集基于 rsEGFP2（PDB 8A6G，采用 $\alpha_{sim} = 0.4$建模）。(a) 经校正的 F_{000}，使用精确的差值图相位。估计值集中在真实值处及其下方，橙色体素构成低估尾部（ $\rho_1(\mathbf{r})$的影响）。(b) 与(a)相同，但未进行 F_{000} 校正，因此估计值偏移至真实值上方。(c) 经校正的 F_{000}，差值图仅使用参考态相位计算。参考相位前因子使真实 χ减半，而差值图噪声则使估计值分散在较宽的范围内。

当 r 均值加上三倍运行标准差首次超过此边界时 y,

$$-\Delta\rho(\mathbf{r})/\rho_0^{\text{cut}} < \langle\tilde{\chi}(\mathbf{r})\rangle + 3\sigma_\chi. \tag{18}$$

外推因子的最终估计值取接受的体素集合中运行均值的中位数。正如前文所述，这一过程是一种启发式方法，我们尚未证明它在任何概率模型下都是最优的。不过，正如我们将要展示的，实证测试表明，该方法稳健且有效。

图8 展示了该方法应用于我们的 rsEGFP2 示例案例，其中运行均值返回的结果为 0.192，与最优值 $\chi_{\text{opt}} = 0.2$ 相差 4%。实现细节见第 7.3 小节；在下一节中，我们将进一步量化各方法在模拟数据上的准确性。

NSE 与这种方法的相似之处，不止在于消除了参考密度这一理念。事实上，正如我们在附录A中所示，NSE 的交集方法在一些宽松假设下，是对逐体素量 $\chi^{-1}(\mathbf{r}) = {}^{\rho_0(\mathbf{r})}{}_{-\Delta\rho(\mathbf{r})}$按 $(-\Delta\rho(\mathbf{r}))$ 加权求得的加权均值。因此，NSE 与真空能级匹配基于相同的逐体素估计，但在汇总这些估计时采用了两种我们的方法未采用的方式：它对逆量 χ^{-1} 而非 χ求平均，并对所提供的任意体素按 $(-\Delta\rho(\mathbf{r}))$ 加权求均值，而不是依据可靠性

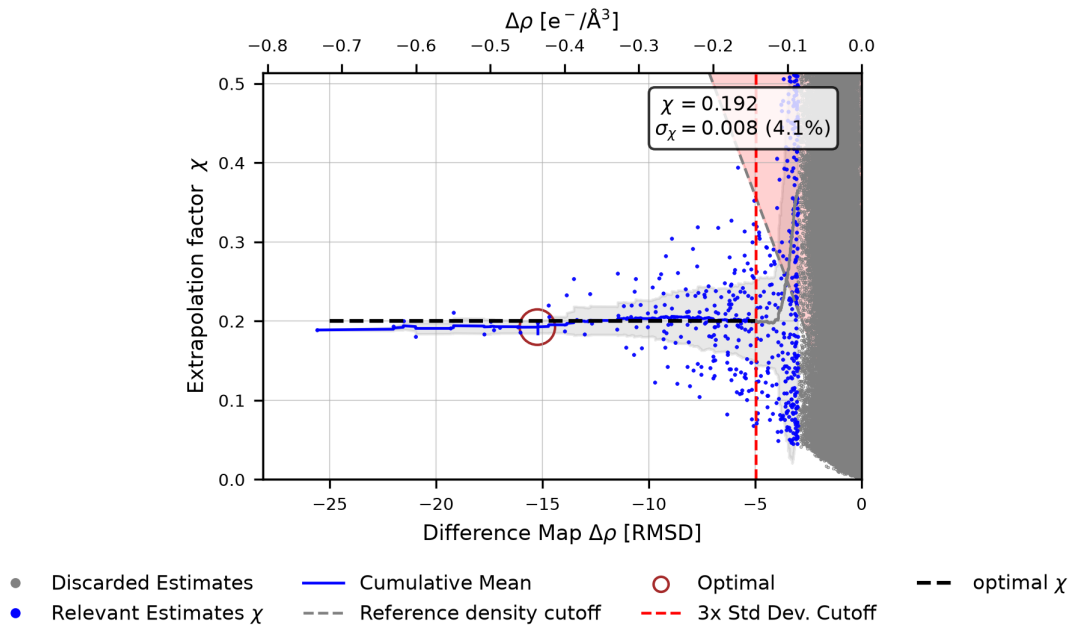

图8：**“真空能级匹配”的应用示例**。与前述方法的示例所用数据集和噪声水平相同（rsEGFP2, PDB 8A6G, $\alpha_{sim} = 0.4$）。每个蓝点都是逐体素估计值 $\tilde{\chi}(\mathbf{r}) = -\Delta\rho(\mathbf{r})/\rho_0(\mathbf{r})$，以 $\Delta\rho(\mathbf{r})$ 为横坐标绘制，单位分别为σ_Δ （下轴）和绝对单位（上轴）。差值图滤波器剔除的估计值以灰点表示。参考密度截断值（灰色虚线）是由$-\Delta\rho(\mathbf{r})/\rho_0^{\text{cut}}$描绘的线，其中 ρ_0^{cut}设为溶剂能级。高于该线（红色背景）的估计值均来自小分母 $\rho_0(\mathbf{r})$，因此可靠性较低。蓝线为累积运行均值，灰色带表示其运行标准差；二者均从最负的 $\Delta\rho(\mathbf{r})$ 开始向零方向累积。红色虚线标示动态截断值，超过该值的体素将从估计中排除，其后的累积均值以灰色显示。报告的估计值（圈出并以蓝色竖线标示）是接受窗口内累积运行均值的中位数。

标准来选择体素。

4 比较模拟数据上的 χ估计方法

为了系统地比较我们在第3节中介绍的方法，我们使用不同占有率和噪声水平下 rsEGFP2 的模拟参考数据集和混合态数据集。模拟设置的详细信息见第7.1小节。图9 展示了针对测试的四种占有率，预测外推因子随信噪比 SNR 的变化。每个数据点的各方法估计值均取自十个独立噪声样本的平均值，并展示均值和标准差（阴影区域）。

差值图使用参考态相位，结构因子未以其他方式处理。因此，最优外推因子是真实占有率的一半，$\alpha/2$，这是一个所有

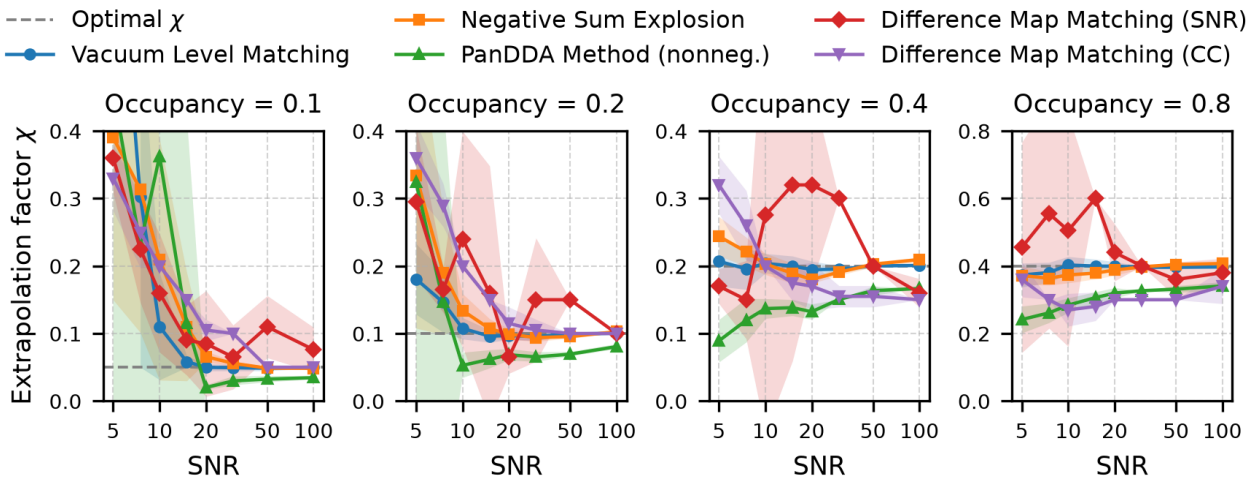

图9：**使用模拟 TR-SFX 信号对外推因子估计值进行基准测试的比较**。四种真实占有率下，估计外推因子随信噪比 SNR 的变化。曲线和标记表示十个独立噪声样本的均值，阴影带表示 ± 一个标准差。灰色虚线表示最优 $\chi_{\mathrm{opt}} = \alpha/2$，即参考相位近似下的最优外推因子（见第2节）。

方法在低噪声下大致能够恢复（信噪比 ≥ 50；见第2节）。在信噪比极低时，没有任何方法是可靠的。同样地，当占有率较低时，较小的信号意味着，在给定信噪比下，差值图的噪声更大，因此所有方法都会遇到困难。因此，最好在中等噪声下区分这些方法，此时它们会产生差异，方法之间的选择也很重要。

差值图匹配的峰积分变体方差最大，因为，正如第3.3小节所述，在找不到良好解时，它经常会报告一个非常高的 χ 。CC方法在高信噪比和低占有率下能够恢复正确外推因子，但在较高占有率下会持续过度外推。对于两种差值图匹配程序，这可能源于它们对参考模型失配的依赖 δ_0，而参考模型失配会随占有率变化。

非负 PanDDA 变体在类TRX数据中表现出系统性的过度外推偏差（未展示的标准版本偏差甚至更大）。其方差下降后，估计值仍低于 χ_{opt}，并且只有随着信噪比升高才逐渐接近它。这种系统性偏移表明，可以重新校准该启发式方法以提高准确性。

负和爆炸与真空能级匹配基于相同的物理原理，即最小化外推图中的负密度。在高信噪比下，它能够可靠地恢复真实的外推因子。然而，在噪声较大的情况下，其方差更高，因为它对整个图谱求和，无法降低逐体素估计最不可靠的体素的权重。

真空能级匹配在高信噪比极限下收敛到相同结果，并在 NSE 能够恢复真实值的情况下恢复该真实值。它的优势体现在噪声较大的情况下。通过逐体素解析估计值并进行自适应滤波，该方法能够使均值更接近真实值，同时保持更小的离散程度；随着噪声增加，这一优势也会进一步扩大。

在这项模拟测试中，与所测试的其他方法相比，真空能级匹配能够在更高的噪声水平下持续且准确地恢复正确外推因子。不过，模拟并不能完美代表实验，因此接下来我们将转向两个真实数据测试案例。

5 真空能级匹配：实验数据

在真实数据上评估外推因子方法，与在模拟数据上评估有着根本区别。这里没有真实值，而任何可靠的准确性衡量标准本身都会构成一种方法。任何特定方法所固有的假设都应与实验现实相符。对于 NSE 和真空能级匹配，这意味着最大的变化会形成这样一种结构：部分电子密度发生移动，并在原处留下一个没有任何电子密度的区域。

首先，我们来看 Fadini 的 rsEGFP2 数据集等人（2023年），该数据集已作为上一节模拟研究的基础。由于真空能级匹配的表现最佳，因此本部分仅考察这一方法。在这一案例研究之后，在第5.2小节中，Christou 等人提供的多时间点光裂合酶数据等人（2023年）将用于检验估计结果在不同预处理流程之间是否保持一致，以及内置质量指标能否正确区分支持可靠预测的数据和不支持可靠预测的数据。

5.1 单时间点应用：rsEGFP2

rsEGFP2 是一种可逆开关荧光蛋白，其发色团在光照下发生 cis–$trans$ 异构化。结构转变局限于发色团及其邻近的少数残基。Fadini等（2023年）的数据集捕捉了低触发态占有率下的这一转变，并用于构建第4节中的模拟基准。

我们首先将参考结构因子和混合态结构因子缩放至模型，并计算相应的平均密度（或零频项）（Lang 等，2014年）。利用参考模型的相位，我们可以通过简单的傅里叶变换计算参考密度 ρ_0 。随后，我们使用参考结构相位并应用迭代全变差去噪（IT-TV），计算两者之间的差值图。这样可以获得具有更精细振幅和相位的差值密度 $\Delta\rho$ 。有关更多细节，请参阅第 7.2 小节。

我们展示了所得的外推因子估计值 $\tilde{\chi}$（$\mathbf{r}$）见图 10。正如预期，对于最负的 $\Delta\rho$（$\mathbf{r}$），存在一个窄分布区间。截断条件能够识别出该区间，并按照预测，在稳定区间内选取外推因子。与图8中的模拟示例相比，窄分布区间具有更高的方差。虽然这种差异在视觉上很明显，但为了量化此类差异，我们认为相对离散度 σ_χ/χ是最具信息量的度量。在超过 30% 时，其数值相对较大，这反映了该数据集固有的高噪声水平。出于同样的原因，截断条件在更负的 $\Delta\rho$（$\mathbf{r}$）处就已被激活。因此，接受窗口的范围比模拟示例中的更短。如上文所述，所报告的 $\chi = 0.34$并不一定对应于实际占有率，但应能给出外推密度的最佳估计。

对于实际测量而言，检验外推因子估计是否实用，关键在于所得映射能否解析出实验旨在捕捉的结构转变。图-

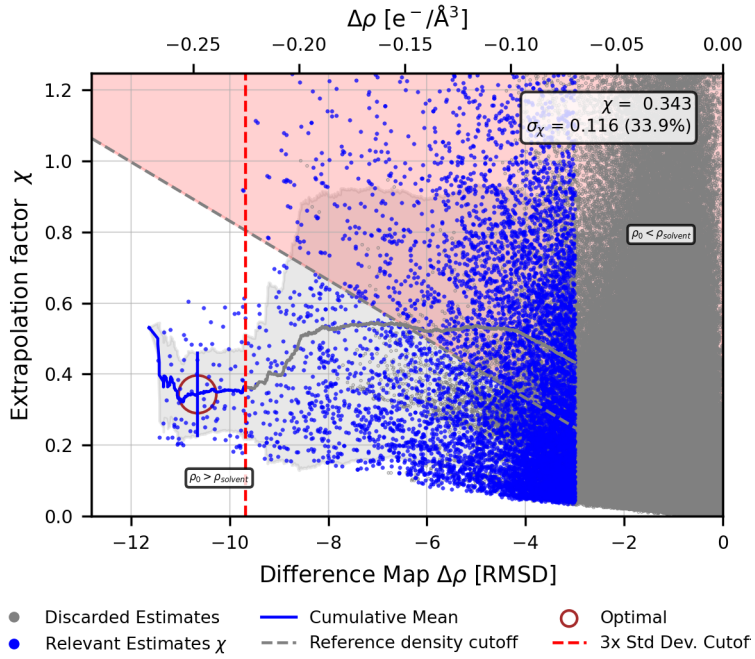

图 10：**对实验数据进行真空能级匹配**。对 Fadini等人 (2023年) 的 rsEGFP2 数据集进行迭代 TV 去噪后，计算得到差值图。逐体素估计 $\tilde{\chi}(\mathbf{r}) = -\Delta\rho(\mathbf{r})/\rho_0(\mathbf{r})$ 从最负的 $\Delta\rho(\mathbf{r})$ 开始向零累积（下方坐标轴以 σ_Δ 为单位，上方坐标轴为绝对单位）。累积均值以蓝色绘制，阴影带表示 $\pm 1\sigma_\chi$ 区域。超过红色虚线（即首次满足截断条件$\langle\tilde{\chi}(\mathbf{r})\rangle + 3\sigma_\chi > -\Delta\rho(\mathbf{r})/\rho_0^{\mathrm{cut}}$的位置）后，累积均值以灰色显示。红色阴影区域表示 $\rho_0(\mathbf{r}) < \rho_0^{\mathrm{cut}}$，此处小分母会放大逐体素估计中的噪声。

11 显示了在发色团位点处的外推电子密度，该电子密度使用 χ中报告的图 10计算得到。该图谱符合顺式发色团模型，而非反式参考状态，从而证实所报告的 χ 在没有显著过度外推的情况下去除了参考贡献。

5.2 光裂合酶：不同预处理下的多时间点一致性

光裂合酶是一种修复紫外线诱导的环丁烷嘧啶二聚体（CPD）损伤的黄素酶。Christou 等人 (2023年) 展示了十个时间点，时间范围从皮秒到数百微秒。在最初的几个纳秒内，环丁烷环逐键断裂，经过约 1 ns 时的单键中间体，并在数纳秒内达到完全断裂状态。在微秒时间尺度上，断裂后的胸腺嘧啶翻转出结合口袋，从而完成产物释放。

我们使用这些数据对我们的方法进行回顾性测试。我们展示了四个时间点：1 ns、3 ns、30 ns 和 1 µs；每个时间点均采用两种不同的差异图处理流程进行处理。第一种处理流程使用*i*加权的差值图，这些差值图采用结构因子加权计算，正如 Christou*et al.*（2023年）发表的那样。另一种处理流程是我们认为目前最先进的方法——迭代 TV

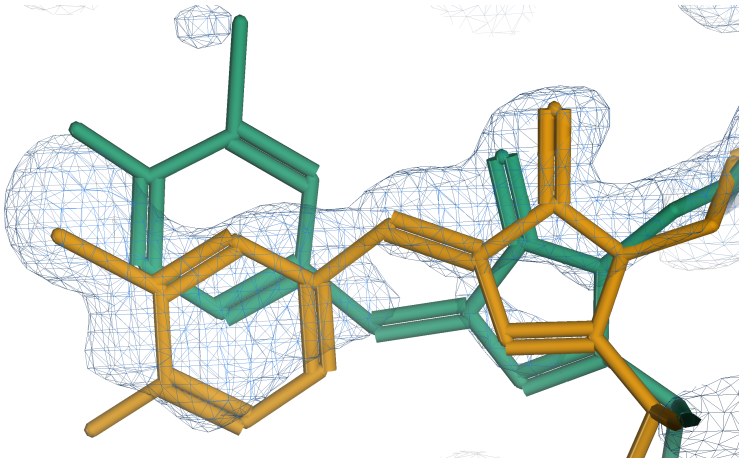

图 11：**外推电子密度示例**。图中显示了 rsEGFP2 发色团及使用最优 χ（由图 10给出）计算的外推密度。图中以绿色显示反式（参考态）发色团模型，以橙色显示已发表结构中的顺式（触发态）发色团模型。外推图（网格）以均方根偏差 $= 1$为等值面，在发色团区域周围绘制。

去噪（IT-TV）。该方法在实空间中调整图谱，并以迭代方式使这些变化与测量得到的结构因子保持一致，从而改善相位估计。

图 12展示了两种处理流程下这四个时间点的逐体素分布和累积统计，而图 13展示了相应的外推图。我们将两者结合起来讨论，因为只有结合其生成的图谱，才能理解某个时间点所得到的估计值的意义。

在 3 ns、30 ns 和 1µs 时，差值图信号呈现出相似的重排。因此，两种处理流程下的真空能级匹配都能找到稳定的运行均值（图 12）。在这三个时间点，我们观察到大量差值图体素具有相似的 χ。与第5.1小节中的 rsEGFP2 数据集相比（其信号占主导的区域较短），光裂合酶图在更大的 $\Delta\rho$范围内都保持狭窄。这增强了我们对已识别出可信外推因子的信心。

绝对 χ 值在k加权方法与 IT-TV 之间于每个时间点都存在差异，而且两种处理流程在不同时间点的分歧程度也各不相同。然而，图13 中的相应图谱都解析出了相同的裂解区域结构。时间点之间的差异可能具有物理原因，反映了泵浦激光功率或该状态的实际占有率。但在单个时间点，k加权方法与 IT-TV 之间的差异源于相同的底层数据和相同的结构状态，因此只能是处理引起的。由此可见，仅处理过程就可能扰乱物理占有率与外推因子之间的关系。这与第2节 的分析一致，并进一步支持这样的建议：任何报告的 χ 都应注明生成它的处理流程。

我们的自动防护机制在 1ns 时间点发出警告。它们检测到，在截断规则触发前纳入的估计值数量很少。因此，估计值过少，

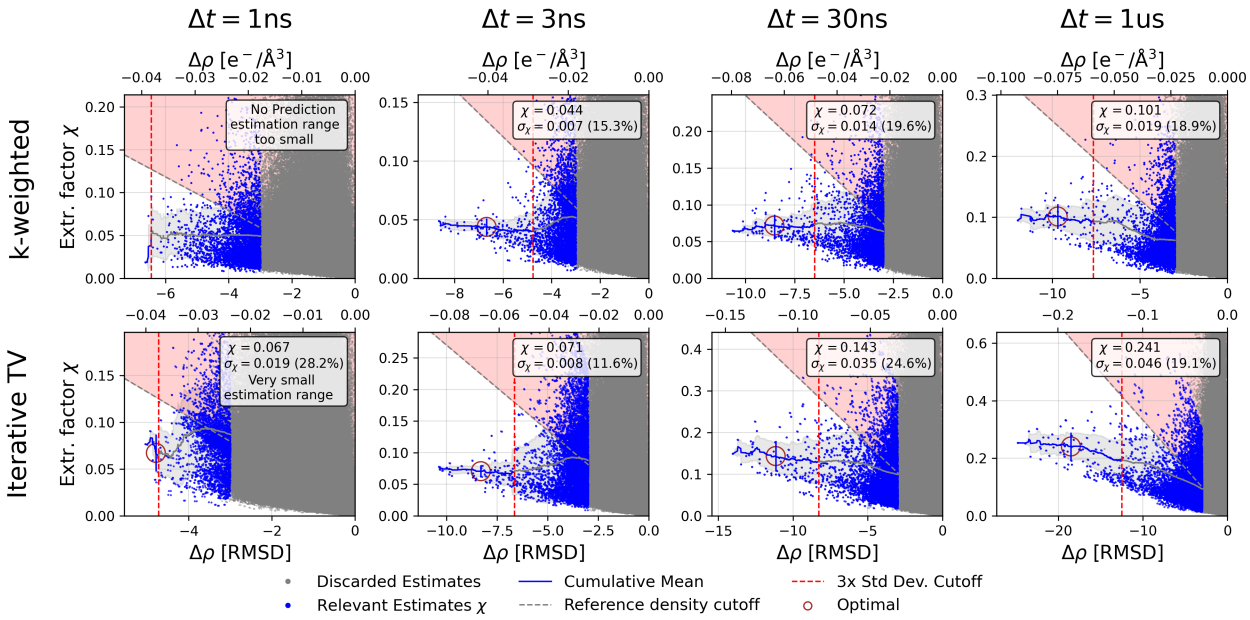

图12： **将真空能级匹配应用于一系列差值图**。使用 Christou 等人（2023年）的光裂合酶数据，我们展示了采用两种差值图处理流程的四个时间点。上排展示了 Christou 等人在 2023 年发表的 k 加权差值图，下排展示了第 7.2 小节所述的迭代 TV 去噪差值图。每个面板均显示逐体素估计（蓝点）、累积运行均值（蓝色实线）、运行标准差（灰色区域）、动态截断值（红色虚线），以及数值 χ预测或警告标志。

无法以自动化方式继续处理数据点。

这里之所以会出现这种情况，是因为差值图整体较弱，而且其中最负的体素并不像后续时间点那样位于两个主要裂解位点，而是在该区域周围形成一系列小幅偏移。因此，逐体素估计的分布相应地很分散，所以在默认设置下，两种处理流程都会触发警告，不过严重程度不同，因为它们抑制噪声的方式不同。对于 IT-TV，由于接受窗口较短，覆盖范围不到负值范围的 20%，因此会发出警告。对于k加权图，落在足够小的标准差范围内的数据点非常少，因此完全不会报告估计值。在这两种情况下，该方法都表明自身无法保证估计可靠，必须进行目视检查。

在这种情况下，检查该图仍能得到一个可用的数值。对于相当一部分最负的体素，累积运行均值大致保持平坦，因此可以读出一个有依据的估值，但需注意前述限制。我们选取了 $\chi = 0.05$ 作为k加权图谱的参数，并为 IT-TV 选取了 0.067；在该值附近，均值一直徘徊到约 $4.2\,\sigma_{\Delta}$。两者都产生了合理的外推图（图13），其中 IT-TV 图谱对裂解区域密度的解析略为详细，这与 IT-TV 流程所采用的实空间锐化相一致。

我们在两种不同的预处理流程下，获得了全部四个时间点的 χ 估计值，并且

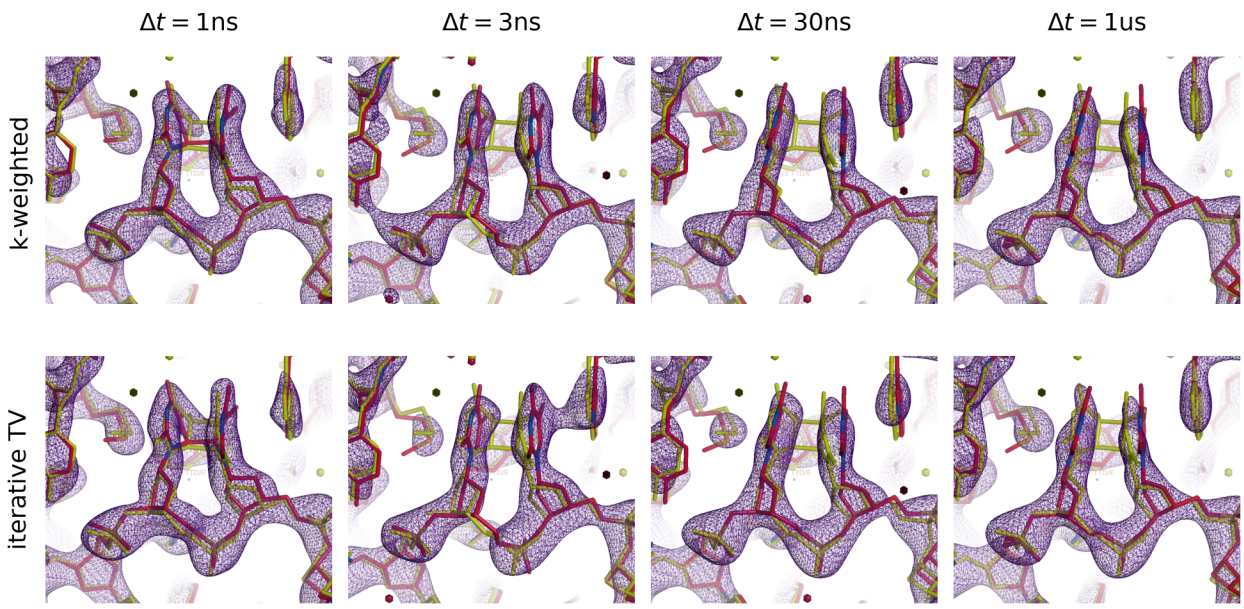

图13：**外推电子密度图谱**。这里，我们展示使用 χ 在图12中估算的外推密度所产生的主要变化。上排：来自 Christou k加权差值图谱的图谱 等人 （2023年）。下排：来自迭代 TV 去噪差值图的图谱。两种图谱的等高线均设为均方根偏差 $= 2$。同时显示参考模型（黄色）和已发表的外推模型（红色）。

由此得到的外推图谱能够解析此前需要人工干预才能识别的结构变化。综合来看，光裂合酶的结果表明，真空能级匹配在同一处理流程内以及结构相似的时间点之间产生的估计值具有内部一致性；绝对值会按预期受到预处理流程的影响（见第2节）；此外，方法会根据数据质量自动给出结果，或适时交由人工审查。

6 结论

TRX 正在迅速发展，但准确区分混合结构状态仍是一个主要瓶颈。Xtrapol8 等工具如今已能完成从测得的结构因子到首个精修模型的所有步骤，使非专业人士也能进行 TRX 分析。因此，包括外推因子估计在内的每个步骤都可靠且明确说明其假设，这一点愈发重要。我们提出并通过一个实验实例证明，外推因子估计不能简单地等同于占有率，因为差值图处理对其造成的影响通常无法逆转。在光裂合酶数据中，同一物理状态经两种处理流程处理后会得到不同的外推因子，但二者都能产生可解释的结果。我们调查了当前使用的方法；Tripathi 等人（2012年）提出的负密度假设启发我们引入了真空能级匹配。这是一种外推因子估计方法，其透明度和准确性都优于

现有方法，正如我们在第4节中所示。

透明性是该方法的核心设计目标。该方法不仅会生成一个估计值，还会说明该估计值的可靠程度，呈现逐体素证据，并标记数据不足以支持有把握预测的情况。其输出分为三种诊断状态。未触发任何警告时，估计值会被报告为可信。当触发警告时，估计值会被标记为需要人工评估。正如光裂合酶的结果所示，在这种情况下，通常仍可从逐体素图中读出一个有依据的数值。当没有任何估计值满足筛选标准时，该方法会返回误差，而不是返回一个数值。这些状态基于本研究中识别并表征的若干失效模式（第3.5.2小节）。未经测试的数据条件可能会揭示更多失效模式。这些保守的诊断标准旨在标记未预料到的失效，或阻止返回估计值，从而降低将不可靠的估计值报告为可信结果的风险。

该方法的适用范围受到若干限制。真空能级匹配建立在双态近似之上。它会针对图谱中的主导负密度信号估计单一外推因子；但不会确定有多少种不同的结构变化共同促成该信号。因此，该估计值描述的是主导状态，可能不适用于所有发生变化的区域。当多个变化可以在空间上分离时，可使用局部掩膜分别分析各个区域。当促成信号的状态在空间上重叠时，则需要专门的多态分离，这是单图谱方法所无法提供的。此外，该方法还要求存在与现有密度位移相关的大量负差值密度。因此，它对配体结合的适用性仅限于配体置换现有密度、而非仅占据空白空间的情况。

两个实验数据集以及基于其中一个数据集进行的模拟，都涉及空间上局部的小幅变化：rsEGFP2中的发色团异构化，以及光裂合酶中的键断裂。这些测试案例并未涉及大规模构象重排、膜蛋白，也未涉及差值信号分散于整个晶胞的系统。下一步可以研究该方法如何处理大尺度变化，并将其应用于更广泛的时间分辨数据集。

另一个后续步骤是改进用于定位 $\tilde{\chi}(\mathbf{r})$ 的窄分布区间并从中确定 χ的方法（见第3.5.2小小节）。尽管这些启发式方法在我们的测试中表现有效，但缺乏正式的统计学依据。一种途径是估计差值图的噪声水平，从而得到真实空置体素的逐体素估计的预期分布。这样可以更有依据地对合理的真空态数值进行分类，并确定窄分布区间中的方差有多大程度仅由噪声造成。类似地，匹配基线也可以设为局部基线。目前，该方法只要求至少有一些体素被置换为空置状态，而由溶剂回填的体素实际上会被其余体素的影响所压过（第3.5.1小小节）。相反，将每个体素与其预期的局部背景进行匹配——蛋白质内部为真空，其表面处为溶剂能级——可以让溶剂回填体素参与估计，而不仅仅是容忍它们，从而增加信息性体素的数量。

除了该方法本身之外，我们不建议使用差值图直接确定占有率，尤其是在无法量化处理对占有率–外推因子关系的影响时。相反，我们建议首先利用外推因子获得 $\mathbf{F}\chi$ 的最佳估计值。根据该估计值精修的模型计算出的结构因子

随后可以代替 $\mathbf{F_1}$ ，代入 公式 3。在只剩一个未知数的情况下，可以估计占有率 α，使其能够最佳地解释测得的振幅 $F_{\mathrm{mix}} \approx \left|(1-\alpha)\,\mathbf{F_0} + \alpha\,\mathbf{F}_\chi\right|$，其中 $\mathbf{F_0}$ 和$\mathbf{F}\chi$ 保持不变。

7 方法

7.1 rsEGFP2 的模拟

结构因子的生成。 为了将第3节中的方法与已知真实值进行基准测试，我们生成了rsEGFP2的模拟参考态和混合态数据集，其中发色团在光照下发生 $trans$到cis 异构化。我们以PDB 条目 8A6G为基础，该条目包含具有部分占有率的 $trans$ 和cis 态。参考态建模为完全占有率下的$trans$ 态，混合态则建模为指定占有率下的cis 态 α_{sim} 和$trans$ 态（占有率为 $1-\alpha_{sim}$），发色团之外的所有原子均保持在其参考位置。使用$gemmi$软件包（Wojdyr，2022年）计算两种状态的复结构因子，其中包括体溶剂贡献，分辨率为1。6 ˚Å。

噪声模拟。 我们根据计数项与强度成比例项的组合，对结构因子振幅中的两种误差来源进行建模；这种组合通常用于标度误差模型，其依据是参考文献（Evans，2006年）中使用的误差模型（见附录A3）。首先，串行数据主要受乘性误差影响，例如部分性、脉冲强度和晶体尺寸的脉冲间变化，以及不完善标度；这些因素会按每个强度的大小成比例地扰动强度（Kirian 等，2010年；White 等，2013年）。传播到振幅后，会产生一个与 $|F|$成比例的噪声项。其次，光子计数服从泊松分布，因此$\sigma_I = \sqrt{I}$。通过 $\sigma_{|F|} \approx \sigma_I/(2|F|)$传播这种不确定性后，会在振幅上产生近似恒定¯的噪声水平，对最弱反射的影响最为显著。我们用最外层、最高分辨率反射环上的平均振幅 $\langle F_{\mathrm{clean}}\rangle_{\mathcal{S}}$ 表示这一恒定项 $\mathcal{S}$，该反射环定义为分辨率截止值1.6 ˚A以内的所有$\mathbf{q}$ 0.1 ˚。我们对每个模型的结构因子按如下方式施加噪声：

$$F_{\mathrm{noisy}}(\mathbf{q}) = F_{\mathrm{clean}}(\mathbf{q}) + \frac{1}{\mathrm{SNR}}\left(\varepsilon_1 F_{\mathrm{clean}}(\mathbf{q}) + \varepsilon_2 \langle F_{\mathrm{clean}}\rangle_{\mathcal{S}}\right), \tag{19}$$

其中， ε_1 和 ε_2 是从均值为零、标准差为单位值的高斯分布中抽取的随机数。我们将此缩放参数称为 SNR。由于加性项固定为最高分辨率环的平均振幅，因此除最弱的反射外，SNR 都是所有反射信噪比的上限。

外推因子估算方法的配置。我们使用 Xtrapol8 的 `fextr` 和 `fofo` 选项进行差值图匹配，这些选项不对结构因子进行重加权；在整个过程中，我们在 χ 介于 0.01 和 0.9 之间的 15 个取值处评估差值图匹配。我们同时报告峰积分版本（结合正峰和负峰）以及 CC 版本。对于峰积分方法，我们报告在 $(\Delta\rho)\chi$ 和 $(\Delta\rho)_{\mathrm{obs}}$在 $(\Delta\rho)_{\mathrm{obs}}$ (第3.3小节) 中绝对值最大的体素处符号一致时所对应的最大值，这排除了在较大 χ下不足外推图谱的虚假高分。CC方法无需进行此类检查，因为符号反转会直接降低皮尔逊相关。

我们采用 NSE 变体，将差值图过滤至 $3\sigma_{\Delta}$ 并纳入 F_{000} 偏移量，具体讨论见 第3.4小节。为拟合欠外推区间，我们在 1 到 2 之间的五个等距点计算 ρ_{χ}，其中 χ^{-1}；为拟合过度外推区间，我们采用 100 到 200 之间的五个等距点。图6所示的拟合也基于这些计算结果。

对于 PanDDA，我们采用第3.2小节中的非负形式。真空能级匹配按照第3.5小节中的说明进行。

模拟设计。对于第4节的基准测试，我们在一组共同的占有率网格（$\alpha_{sim} \in \{0.1, 0.2, 0.4, 0.8\}$）和噪声水平（信噪比 $\in \{5, 7.5, 10, 15, 20, 30, 50,100\}$）上评估了每种方法。每种（占有率、噪声）组合均使用十个独立噪声样本重复测试；对于每个样本，所有方法都应用于同一张差值图，因此结果的离散程度反映的是方法的差异，而非数据的差异。贯穿第3节展示的示例数据集使用 $\alpha_{sim} = 0.4$，由此得到最佳 $\chi = 0.2$，信噪比为 $= 15$。

7.2 实验数据处理

确定常数分量。由于零频项 F_{000} 通常无法测量，重建图缺少常数偏移（见第3.4小节）。我们按照 Lang 等人 (2014) 的方法恢复这一偏移。参考图的绝对密度由蛋白质贡献和体溶剂贡献构成。首先，使用$gemmi$计算模型图的结构因子。然后，使用掩模识别蛋白质区域和溶剂区域。我们通过有界一维搜索估算体溶剂密度 $\rho_{solvent}$。对于每个候选值$\rho_{solvent}$，我们在模型图的溶剂掩模区域上添加均匀密度。将所得模型缩放至实验数据，并计算平均绝对振幅差；计算范围仅限于分辨率粗于 4 $^{\circ}$A 的反射，此处体相溶剂占主导地位。选取使该差值最小的 $\rho_{solvent}$。蛋白质贡献是模型图在蛋白质区域内的体积加权平均密度，溶剂贡献则是溶剂掩模区域内的体相溶剂水平 $\rho_{solvent}$。

差值图。 在第5节中，我们重复使用k加权差值图，该图由 Christou等人（2023年）报道，使用 Xtrapol8 计算 $k = 1$。我们使用的第二种方法是迭代全变差去噪（Fadini 等人，2025年）（ meteor 0.5.0 版，可从以下地址获取：`https://github.com/rs-station/meteor`）。

完整外推流程。 从实验测定的结构因子 F_0和 F_{mix} 参考模型开始，我们首先将 F_0 和 F_{mix} 缩放至模型。随后，采用 k 加权方法或迭代 TV 去噪方法计算差值图，并确定参考结构的常数分量 F_{000}。接下来，我们估计最佳外推因子 χ (第3节)。最后，按照公式 5进行外推；为方便起见，现将其重述如下：

$$\mathbf{F}_{\chi} = \mathbf{F_0} + \chi^{-1}\,\Delta\mathbf{F}.$$

7.3 真空能级匹配实现细节

采样率。 实空间差值图在采样因子为五的网格上进行采样，即 $2.5\times$ 奈奎斯特–香农最低采样率，以便为逐体素分析提供足够的空间分辨率。更精细的采样不会增加信息量，而是以更高密度对同一图谱进行采样，从而增加进入运行统计量的点数。因此，当累积均值不稳定时，更精细的采样很有用；当统计量不受影响时，则可以降低采样率。

差值图滤波。 我们将分析限制在自身显著为负、 $-\Delta\rho(\mathbf{r}) \geq 3\,\sigma_{\Delta}$，且连接成体积至少为 $V_{\min} = 1.5\ \text{Å}^3$ 的体素上。这样可以去除孤立的负异常值，因为它们很可能是噪声，而非真实的结构信号。阈值 $V_{\min}$ 设得较为宽松。当图谱中存在大量负信号时，可能更适合采用限制性更强的 $V_{\min}$ 。

参考密度截断值。 正文（第3.5.2小节）使用对角项 $-\Delta\rho(\mathbf{r})/\rho_0^{\text{cut}}$来终止累积。我们将 $\rho_0^{\text{cut}} = \rho_{solvent}$设为体溶剂水平，该值在预处理过程中通过估算 F_{000}确定。此选择基于两个理由。首先，如第3.5.2小节所述，较小的 $\rho_0(\mathbf{r})$ 会放大 $\tilde{\chi}(\mathbf{r})$ 分子中的噪声，而 $\rho_{solvent}$ 标记了这种放大效应占主导地位的密度下限。其次，若某体素的 $\rho_0(\mathbf{r})$ 小于或等于 $\rho_{solvent}$，则无法将其与体相溶剂区分开来，因此其中的负 $\Delta\rho(\mathbf{r})$ 既可能归因于溶剂涨落，也可能归因于有序蛋白质密度的真实变化。这类体素无论其分母为何，都不包含可靠的结构信息。

警告标准。 有三项检查会降低自动估计值的置信度。在每种情况下，数据都违反了该方法所依据的某项假设。第一项检查会在接受窗口跨度小于负和-$\Delta\rho$ 范围的 20% 时发出警告，这表明窄分布区间覆盖的范围过小，无法进行可靠估计。第二项检查会在运行均值在整个窗口内不稳定时发出警告，其定义为：运行均值的最小值（不包括最初的几个体素）低于其最大值的三分之二。第三项检查会在运行标准差相对于报告的估计值较大时发出警告。满足任一条件时，未经检查不应信任该估计值。

真空水平匹配与NSE之间的联系

本附录表明，在两个温和的假设下，负和爆炸等价于定义真空能级匹配的各体素零交叉点的加权平均。NSE 估计值是两条回归直线的交点：一条拟合于欠外推区间（较小的 χ^{-1}），另一条拟合于过度外推区间（较大的 χ^{-1}）。对于每个 χ^{-1}，负和会汇总所有满足 $\rho_\chi(\mathbf{r}) < 0$ 的体素。为了将每条回归直线表示为线性方程，我们首先确定在每个区间内哪些体素会参与计算。

我们只考虑在物理上可能穿过零点的体素。参考密度为 $\rho_0(\mathbf{r}) \leq 0$ 的体素会被排除，因为低于真空水平的参考密度不符合物理规律；这类体素主要由缺失的常数偏移导致，一旦恢复 F_{000}，几乎所有这类体素都会被去除（第3.4小节）。同样也会排除 $\Delta\rho(\mathbf{r}) \geq 0$的体素：鉴于 $\rho_0(\mathbf{r}) > 0$，其外推密度 $\rho_0(\mathbf{r}) + \chi^{-1}\Delta\rho(\mathbf{r})$ 只有在 $\chi^{-1} < 0$时才可能变为负值，也就是说，只有通过减去差值而非加上差值才会如此。与正文一样，我们进一步将分析限定于足够负的差值密度，由此得到体素集合

$$\mathbf{r}_{neg} = \{\mathbf{r} : \Delta\rho(\mathbf{r}) \leq -c,\ \rho_0(\mathbf{r}) > 0\}, \qquad c \geq 0. \tag{20}$$

设定 $c = 0$ 包含所有负差值体素，而 $c > 0$ （正文中为 $c = 3\,\sigma_\Delta$）会排除那些小 $|\Delta\rho(\mathbf{r})|$会使其零交叉点延迟至非常大的 χ^{-1}的体素。

第一个假设涉及欠外推区间。对于足够小的 χ^{-1}，每个体素都满足 $\rho_0(\mathbf{r}) > -\chi^{-1}\Delta\rho(\mathbf{r})$，因此 $\rho_\chi(\mathbf{r})$ 保持为正，负和恒为零。因此，其回归线为零，而两条直线的交点位于过度外推直线的根处。

第二个假设涉及过度外推区间。对于足够大的 χ^{-1}，在 $\mathbf{r}^{neg}$ 的每个体素中，差值项都占主导地位并超过参考值，因此所有体素都已越过零点。此时负和仅遍历固定集合$\mathbf{r}^{neg}$，与 χ^{-1} 无关，并且关于 χ^{-1} 呈线性，

$$\sum_{\mathbf{r}\in\mathbf{r}_{neg}} \rho_\chi(\mathbf{r}) = \sum_{\mathbf{r}\in\mathbf{r}_{neg}} \rho_0(\mathbf{r}) + \chi^{-1} \sum_{\mathbf{r}\in\mathbf{r}_{neg}} \Delta\rho(\mathbf{r}), \tag{21}$$

因此回归可以精确恢复这条直线。我们通过用截距除以斜率来求其根，

$$\chi^{-1} = \frac{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} \rho_0(\mathbf{r})}{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} -\Delta\rho(\mathbf{r})}. \tag{22}$$

获得全局估计后，我们将注意力转向逐体素估计。将相同的构造应用于单个体素，即可在每个点获得一个估计值，

$$\tilde{\chi^{-1}}(\mathbf{r}) = \frac{\rho_0(\mathbf{r})}{-\Delta\rho(\mathbf{r})}. \tag{23}$$

取 对所有逐体素交叉点求平均，并以 $-\Delta\rho(\mathbf{r})$ 加权，我们得到 n

$$\begin{aligned}\chi^{-1} &= \frac{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} \tilde{\chi^{-1}}(\mathbf{r})(-\Delta\rho(\mathbf{r}))}{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} -\Delta\rho(\mathbf{r})} \\ &= \frac{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} \frac{\rho_0(\mathbf{r})}{-\Delta\rho(\mathbf{r})}(-\Delta\rho(\mathbf{r}))}{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} -\Delta\rho(\mathbf{r})} \\ &= \frac{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} \rho_0(\mathbf{r})}{\sum_{\mathbf{r}\in\mathbf{r}_{neg}} -\Delta\rho(\mathbf{r})}.\end{aligned} \tag{24}$$

这与 χ^{-1} 的全局估计表达式相同。因此，在这种情况下，NSE 是对真空能级匹配明确呈现的逐体素零交叉点进行加权平均，每个体素的权重为 $-\Delta\rho(\mathbf{r})$。如果任一假设不成立，例如低-χ^{-1} 区间在零值处并非平坦，或者 $\mathbf{r}^{neg}$ 中的某些体素在拟合范围内尚未穿过零点，那么这种等价关系就只是近似成立，两个估计值会略有不同。

由于 NSE 通过线性回归而非直接计算来获得这一平均值，因此拟合范围的选择具有任意性；而真空能级匹配则揭示了逐体素分布以及每种选择的影响。一旦分析仅限于强负值体素，按差值密度确定的权重本身就没有太大差异。差值图噪声作为加性项进入公式 17，即 $\alpha\, w\, n_\Delta/\rho_0(\mathbf{r})$，其幅度由 $\rho_0(\mathbf{r})$ 决定，且独立于该体素的真实 χ。在倒数映射 $\chi^{-1} = 1/\chi$下，作用于 χ 的加性扰动 δ 会在 χ^{-1} 上产生幅度为 δ/χ^2 的扰动，其大小取决于该体素处 χ 的值。因此，在 χ 中进行聚合，可在各体素间保持统一的加性噪声尺度；而在 χ^{-1} 中聚合（NSE 隐式采用这种方式）则无法做到这一点。

B 串行晶体学中的占有率方差

在串行飞秒晶体学（SFX）实验中，均匀且恒定的占有率这一假设几乎总是不成立。实际上，由于多种实验因素，占有率 α 会因晶体而异，其中包括单个晶体的大小、其在液体喷流中的位置、它相对于聚焦光学激光束的排列方式，以及晶格相对于激光偏振的取向。

由于 SFX 数据处理会合并数千个单独晶体的衍射图样，最终分析得到的是平均强度。为了解其影响，我们考察系综中占有率的方差如何影响最终的结构因子振幅。

为了直接呈现对 α 的依赖，我们引入$\mathbf{F_{1-0}} = \mathbf{F_1} - \mathbf{F_0}$, 触发态（状态 1）与参考态（状态 0）之间纯粹的复结构差异。正文中，我们主要讨论观测差分电子密度 $\mathbf{\Delta F} = \alpha\mathbf{F_{1-0}}$。然而，鉴于我们的目标是理解 α的作用，我们在以下推导中将其明确写出。若考虑一个具有特定占有率 α的单个晶体（或单次“曝光”），测得的强度

is

$$
\begin{aligned}
I_{mix}(\alpha) &= |\mathbf{F_0} + \alpha \mathbf{F_{1-0}}|^2 \\
&= |\mathbf{F_0}|^2 + 2\alpha \Re(\mathbf{F_0}^* \mathbf{F_{1-0}}) + \alpha^2 |\mathbf{F_{1-0}}|^2 .
\end{aligned} \tag{25}
$$

由于数据集是汇总结果，我们计算系综中的期望值（平均测量值）。利用统计恒等式 $\langle \alpha^2 \rangle = \langle \alpha \rangle^2 + \sigma_\alpha^2$，其中 $\langle \alpha \rangle$ 是平均占有率， σ_α^2 是其方差，平均强度为

$$
\begin{aligned}
\langle I_{mix} \rangle &= |\mathbf{F_0}|^2 + 2\langle \alpha \rangle \Re(\mathbf{F_0}^* \mathbf{F_{1-0}}) + (\langle \alpha \rangle^2 + \sigma_\alpha^2) |\mathbf{F_{1-0}}|^2 \\
&= |\mathbf{F_0} + \langle \alpha \rangle \mathbf{F_{1-0}}|^2 + \sigma_\alpha^2 |\mathbf{F_{1-0}}|^2 .
\end{aligned} \tag{26}
$$

这里的 σ_α 表示不同晶体间占有率的离散程度，与正文中使用的差值图和滚动估计标准差无关。

为得到观测到的混合态振幅 $|F_{\mathrm{mix}}|$，我们取平均强度的平方根。分别用 φ_0 和 φ_{1-0} 表示$\mathbf{F_0}$ 和 $\mathbf{F_{1-0}}$ 的相位，可得

$$
\begin{aligned}
|F_{mix}| &= \sqrt{\langle I_{mix} \rangle} \\
&= |\mathbf{F_0}| \Bigg[1 + 2\langle \alpha \rangle \frac{|\mathbf{F_{1-0}}|}{|\mathbf{F_0}|} \cos(\varphi_0 - \varphi_{1-0}) \\
&\qquad + (\langle \alpha \rangle^2 + \sigma_\alpha^2) \frac{|\mathbf{F_{1-0}}|^2}{|\mathbf{F_0}|^2} \Bigg]^{1/2} .
\end{aligned} \tag{27}
$$

对于占有率加权的结构差异相对于参考振幅较小的反射，我们将结构差异一致地展开至二阶。利用$\sqrt{1+x} = 1 + x/2 - x^2/8 + \cdots$，可得

$$
\begin{aligned}
|F_{mix}| \approx\ & |\mathbf{F_0}| + \langle \alpha \rangle |\mathbf{F_{1-0}}| \cos(\varphi_0 - \varphi_{1-0}) \\
& + \frac{(\langle \alpha \rangle^2 + \sigma_\alpha^2) |\mathbf{F_{1-0}}|^2}{2|\mathbf{F_0}|} \\
& - \frac{\langle \alpha \rangle^2 |\mathbf{F_{1-0}}|^2 \cos^2(\varphi_0 - \varphi_{1-0})}{2|\mathbf{F_0}|} .
\end{aligned} \tag{28}
$$

减去参考振幅并合并二次项，得到

$$
\begin{aligned}
\Delta|F| &= |F_{mix}| - |\mathbf{F_0}| \\
&\approx \langle \alpha \rangle |\mathbf{F_{1-0}}| \cos(\varphi_0 - \varphi_{1-0}) \\
&\quad + \frac{|\mathbf{F_{1-0}}|^2}{2|\mathbf{F_0}|} \left[\sigma_\alpha^2 + \langle \alpha \rangle^2 \sin^2(\varphi_0 - \varphi_{1-0}) \right] .
\end{aligned} \tag{29}
$$

因此，占有率方差会在二阶项中为差异振幅增加一个正项。即使不存在占有率方差，当结构差异具有垂直于复数平面中参考结构因子的分量时，仍会产生二阶修正

。
设定 $\sigma_\alpha = 0$ 并仅保留一阶项，即可得到第2.2小节中使用的单占据近似：

$$\Delta|F| \approx \langle\alpha\rangle|\mathbf{F_{1-0}}|\cos(\varphi_0 - \varphi_{1-0}).$$

指定参考相位并使用

$$\cos(\varphi_0 - \varphi_{1-0})\, e^{i\varphi_0} = \tfrac{1}{2}e^{i\varphi_{1-0}} + \tfrac{1}{2}e^{i(2\varphi_0 - \varphi_{1-0})}$$

即可得到公式8中的一阶分解，其中 $\alpha = \langle\alpha\rangle$。

C 差值图匹配评分的分解

我们推导公式14中的分解及其对 CC 方法的影响。我们考虑正的外推因子，并用以下形式表示后处理： W_{obs} 和 W_χ。
将外推图代入处理后的差值图，并加上再减去观测到的差值项，可得

$$\begin{aligned}(\Delta\rho)_\chi &= W_\chi\left[\rho_0^{\mathrm{obs}} + \chi^{-1}(\rho_{mix} - \rho_0) - \rho_0^{\mathrm{calc}}\right] \\ &= \chi^{-1}W_{\mathrm{obs}}(\rho_{mix} - \rho_0) + \chi^{-1}(W_\chi - W_{\mathrm{obs}})(\rho_{mix} - \rho_0) \\ &\quad + W_\chi(\rho_0^{\mathrm{obs}} - \rho_0^{\mathrm{calc}}) \\ &= \chi^{-1}(\Delta\rho)_{\mathrm{obs}} + \chi^{-1}\delta_\Delta + \delta_0, \end{aligned} \tag{30}$$

其中， $\delta_\Delta = (W_\chi - W_{\mathrm{obs}})(\rho_{mix} - \rho_0)$ 是两个处理步骤之间的失配，而 $\delta_0 = W_\chi(\rho_0^{\mathrm{obs}} - \rho_0^{\mathrm{calc}})$ 是参考模型失配。
皮尔逊相关不受正比例缩放的影响。因此，对于一组固定的体素，只要两个图谱的方差均不为零，

$$\mathrm{CC}((\Delta\rho)_{\mathrm{obs}}, (\Delta\rho)_\chi) = \mathrm{CC}\left((\Delta\rho)_{\mathrm{obs}}, (\Delta\rho)_{\mathrm{obs}} + \delta_\Delta + \chi\delta_0\right). \tag{31}$$

请注意，如果 $\delta_0 = 0$，则该分数与 χ 无关；如果两个失配项均为零，则所有候选值的分数均为 1。

因此，aCC 最大值反映的是失配项引入的模式，而非仅仅反映其绝对值之和最小。

资助
我们感谢德国电子同步加速器（DESY）——亥姆霍兹联合会成员——以及汉堡数据科学——亥姆霍兹物质结构研究生院（DASHH）提供的资金支持。本研究的部分工作由汉堡工业大学资助并在那里完成。本工作还得到了德国研究联合会（DFG）卓越集群“CUI：物质的先进成像”（EXC 2056，项目编号 390715994）的进一步支持。

利益冲突： 不存在利益冲突。

数据可用性： 用于重新生成图表的代码和数据可在以下地址找到：
https://github.com/cvazz/xtr-estimator/并根据 Apache 2 许可证提供。

参考文献

Banari, A., Samanta, A. K., Munke, A., Laugks, T., Bajt, S., Gr¨unewald, K., Marlovits, T. C., K¨upper, J., Maia, F. R., Chapman, H. N., Oberth¨ur, D. & Seuring, C. (2025). Nature Methods, 22(7), 1420–1435. Barends, T. R. M., Stauch, B., Cherezov, V. & Schlichting, I. (2022). Nature Reviews Methods Pri mers, 2(1), 59. Baxter, J. M., Hutchison, C. D. M., Fadini, A., Maghlaoui, K., Cordon-Preciado, V., Morgan, R. M. L., Agthe, M., Horrell, S., Tellkamp, F., Mehrabi, P., Pfeifer, Y., M¨uller-Werkmeister, H. M., von Stetten, D., Pearson, A. R. & van Thor, J. J. (2024). Journal of the American Chemical Society, 146(24), 16394–16403. Br¨and´en, G. & Neutze, R. (2021). Science, 373(6558), eaba0954. Christou, N.-E., Apostolopoulou, V., Melo, D. V. M., Ruppert, M., Fadini, A., Henkel, A., Sprenger, J., Oberthuer, D., G¨unther, S., Pateras, A., Rahmani Mashhour, A., Yefanov, O. M., Galchenkova, M., Reinke, P. Y. A., Kremling, V., Scheer, T. E. S., Lange, E. R., Middendorf, P., Schubert, R., De Zitter, E., Lumbao-Conradson, K., Herrmann, J., Rahighi, S., Kunavar, A., Beale, E. V., Beale, J. H., Cirelli, C., Johnson, P. J. M., Dworkowski, F., Ozerov, D., Bertrand, Q., Wranik, M., Bacellar, C., Bajt, S., Wakatsuki, S., Sellberg, J. A., Huse, N., Turk, D., Chapman, H. N. & Lane, T. J. (2023). Science, 382(6674), 1015–1020. Coquelle, N., Sliwa, M., Woodhouse, J., Schir`o, G., Adam, V., Aquila, A., Barends, T. R. M., Boutet, S., Byrdin, M., Carbajo, S., De La Mora, E., Doak, R. B., Feliks, M., Fieschi, F., Foucar, L., Guillon, V., Hilpert, M., Hunter, M. S., Jakobs, S., Koglin, J. E., Kovacsova, G., Lane, T. J., L´evy, B., Liang, M., Nass, K., Ridard, J., Robinson, J. S., Roome, C. M., Ruckebusch, C., Seaberg, M., Thepaut, M., Cammarata, M., Demachy, I., Field, M., Shoeman, R. L., Bourgeois, D., Colletier, J.-P., Schlichting, I. & Weik, M. (2018). Nature Chemistry, 10(1), 31–37. De Zitter, E., Coquelle, N., Oeser, P., Barends, T. R. M. & Colletier, J.-P. (2022). Communications Biology, 5(1), 640. Evans, P. (2006). Acta Crystallographica Section D: Biological Crystallography, 62(1), 72–82. Fadini, A., Apostolopoulou, V., Lane, T. J. & van Thor, J. J. (2025). Communications Biology, 8(1), 1649. Fadini, A., Hutchison, C. D., Morozov, D., Chang, J., Maghlaoui, K., Perrett, S., Luo, F., Kho, J. C., Romei, M. G., Morgan, R. M. L., Orr, C. M., Cordon-Preciado, V., Fujiwara, T., Nuemket, N., Tosha, T., Tanaka, R., Owada, S., Tono, K., Iwata, S., Boxer, S. G., Groenhof, G., Nango, E. & van Thor, J. J. (2023). J ournal of the American Chemical Society, 145(29), 15796–15808. Fong, M. I., Hosokawa, Y., Essen, L. O. & Maestre-Reyna, M. (2026). FEBS Open Bio, n/a(n/a). Genick, U. K. (2007). Acta Crystallographica Section D: Biological Crystallography, 63(10), 1029– 1041. ˇ Genick, U. K., Borgstahl, G. E. O., Ng, K., Ren, Z., Pradervand, C., Burke, P. M., Srajer, V., Teng, T.-Y., Schildkamp, W., McRee, D. E., Moffat, K. & Getzoff, E. D. (1997). Science, 275(5305), 1471–1475. Grosse-Kunstleve, R.W., Sauter, N. K., Moriarty, N. W.& Adams, P.D.(2002). Journal of Applied Crystallography, 35(1), 126–136. Hadjidemetriou, K., Coquelle, N., Barends, T. R. M., De Zitter, E., Schlichting, I., Colletier, J.-P. & Weik, M. (2022). Acta Crystallographica Section D: Structural Biology, 78(9), 1131–1142. Henderson, R. & Moffat, J. K. (1971). Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry, 27(7), 1414–1420.

Khusainov, G., Standfuss, J. & Weinert, T. (2024). Structural Dynamics, 11(2), 020901. Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., Spence, J. C. H., Hunter, M., Fromme, P., White, T., Chapman, H. N. & Holton, J. (2010). Optics Express, 18(6), 5713–5723. Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. (2014). 美国国家科学院院刊, 111(1), 237–242. Lee, B., White, K. I., Socolich, M., Klureza, M. A., Henning, R., Srajer, V., Ranganathan, R. & Hekstra, D. R. (2025). Cell, 188(1), 77–88. Luzzati, V. (1952). 晶体学报, 5(6), 802–810. Maestre-Reyna, M., Hosokawa, Y., Wang, P.-H., Saft, M., Caramello, N., Engilberge, S., Franz- Badur, S., Gusti NgurahPutu, E. P., Nakamura, M., Wu, W.-J., Wu, H.-Y., Lee, C.-C., Huang, W.-C., Huang, K.-F., Chang, Y.-K., Yang, C.-H., Fong, M.-I., Lin, W.-T., Yang, K.-C., Ban, Y., Imura, T., Kazuoka, A., Tanida, E., Owada, S., Joti, Y., Tanaka, R., Tanaka, T., Kang, J., Luo, F., Tono, K., Kiontke, S., Korf, L., Umena, Y., Tosha, T., Bessho, Y., Nango, E., Iwata, S., Royant, A., Tsai, M.-D., Yamamoto, J. & Essen, L.-O. (2025). Science Advances, 11(20), eadu7247.Nango, E., Royant, A., Kubo, M., Nakane, T., Wickstrand, C., Kimura, T., Tanaka, T., Tono, K., Song, C., Tanaka, R., Arima, T., Yamashita, A., Kobayashi, J., Hosaka, T., Mizohata, E., Nogly, P., Sugahara, M., Nam, D., Nomura, T., Shimamura, T., Im, D., Fujiwara, T., Ya- manaka, Y., Jeon, B., Nishizawa, T., Oda, K., Fukuda, M., Andersson, R., B˚ath, P., Dods, R., Davidsson, J., Matsuoka, S., Kawatake, S., Murata, M., Nureki, O., Owada, S., Kameshima, T., Hatsui, T., Joti, Y., Schertler, G., Yabashi, M., Bondar, A.-N., Standfuss, J., Neutze, R. & Iwata, S. (2016). Science, 354(6319), 1552–1557. Pandey, S., Bean, R., Sato, T., Poudyal, I., Bielecki, J., Cruz Villarreal, J., Yefanov, O., Mariani, V., White, T. A., Kupitz, C., Hunter, M., Abdellatif, M. H., Bajt, S., Bondar, V., Echelmeier, A., Doppler, D., Emons, M., Frank, M., Fromme, R., Gevorkov, Y., Giovanetti, G., Jiang, M., Kim, D., Kim, Y., Kirkwood, H., Klimovskaia, A., Knoska, J., Koua, F. H. M., Letrun, R., Lisova, S., Maia, L., Mazalova, V., Meza, D., Michelat, T., Ourmazd, A., Palmer, G., Ramilli, M., Schubert, R., Schwander, P., Silenzi, A., Sztuk-Dambietz, J., Tolstikova, A., Chapman, H. N., Ros, A., Barty, A., Fromme, P., Mancuso, A. P. & Schmidt, M. (2020). Nature Methods, 17(1), 73–78. Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M. & Von Delft, F. (2017). Nature Communications, 8(1), 15123. Pearson, A. R. & Mehrabi, P. (2020). Current Opinion in Structural Biology, 65, 168–174. ˇ Ren, Z., Perman, B., Srajer, V., Teng, T.-Y., Pradervand, C., Bourgeois, D., Schotte, F., Ursby, T., Kort, R., Wulff, M. & Moffat, K. (2001). Biochemistry, 40(46), 13788–13801. Rios-Santacruz, R., Poddar, H., Pounot, K., Heyes, D. J., Coquelle, N., Mackintosh, M. J., Johan- nissen, L. O., Schianchi, S., Jeffreys, L. N., De Zitter, E., Munro, R., Appleby, M., Axford, D., Beale, E. V., Cliff, M. J., D´avila-Miliani, M. C., Engilberge, S., Gotthard, G., Hadjidemetriou, K., Hardman, S. J. O., Horrell, S., Hub, J. S., Ishihara, K., Jaho, S., Karras, G., Kataoka, M., Kawakami, R., Mason, T., Okumura, H., Owada, S., Owen, R. L., Royant, A., Saaret, A., Sakuma, M., Shanmugam, M., Sugimoto, H., Tono, K., Zala, N., Beale, J. H., Tosha, T., Colletier, J.-P., Levantino, M., Hay, S., Kozlowski, P. M., Leys, D., Scrutton, N. S., Weik, M. & Schir`o, G. (2026). Nature, 650(8103), 1045–1052. Schmidt, M. (2023). Structural Dynamics, 10(4), 044303. Schmidt, M., Pahl, R., Srajer, V., Anderson, S., Ren, Z., Ihee, H., Rajagopal, S. & Moffat, K. (2004). 美国国家科学院院刊, 101(14), 4799–4804. Schmidt, M., Rajagopal, S., Ren, Z. & Moffat, K. (2003). Biophysical Journal, 84(3), 2112–2129. ˇSrajer, V., Teng, T.-y, Ursby, T., Pradervand, C., Ren, Z., Adachi, S.-i., Schildkamp, W., Bour- geois, D., Wulff, M. & Moffat, K. (1996). Science, 274(5293), 1726–1729. Tripathi, S., ˇSrajer, V., Purwar, N., Henning, R. & Schmidt, M. (2012). Biophysical Journal, 102(2), 325–332. Ursby, T. & Bourgeois, D. (1997). Acta Crystallographica Section A, 53(5), 564–575. White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. A., Zatsepin, N. A. & Chapman, H. N. (2013). Acta Crystallographica. Section D, Biological Crystallography, 69(Pt 7), 1231– 1240.

White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012年). *Journal of Applied Crystallography*, **45**(2),335–341. Wojdyr, M. (2022年). 开放源代码软件杂志, **7**(73),4200.