分子动力学模拟揭示亨廷顿蛋白外显子1中依赖于多聚谷氨酰胺长度的构象变化：对易聚集状态的环境共溶剂调控的启示

J. Geddes-Nelson、X. Liu、K.T. Yong*悉尼大学工程学院生物医学工程学院，澳大利亚新南威尔士州悉尼 *2006**通讯作者：ken.yong@sydney.edu.au

摘要

亨廷顿病（HD）由*HTT*基因中的 CAG 重复扩增所致，该基因编码多聚谷氨酰胺（polyQ）片段，进而促使亨廷顿蛋白（HTT）错误折叠和聚集。尽管实验研究已阐明依赖于多聚谷氨酰胺长度的聚集特征，但聚集前的构象动力学在原子级仍知之甚少。本文报告了 HTT 外显子1构建体的全原子分子动力学（MD）模拟，研究对象包括 N17 结构域、三种具有临床意义长度的多聚谷氨酰胺片段（Q21，野生型；Q40，成年发病阈值；Q70，青少年发病）以及多聚脯氨酸（polyP）区域。采用 OPLS-AA 力场，在显式 SPC/E 水中进行多拷贝（四链）模拟，时长为 100 ns。我们对回转半径（Rg）、溶剂可及表面积（SASA）、均方根偏差（RMSD）和蛋白内氢键进行定量分析，将其作为聚集倾向的构象代理指标。随着多聚谷氨酰胺片段延长，Rg（Q21 为 32.1 ± 0.7 Å，Q70 为 40.2 ± 0.5 Å）和 SASA 均逐渐增加，这与病理变体呈现伸展且溶剂暴露的构象相符。在 0.5–1.0 M 浓度下加入有机助溶剂（甲醇、己烷、三氯乙烯 [TCE]），进一步改变了这些构象景观：TCE 暴露使 Q21（Rg = 43.9 ± 1.3 Å）和 Q40（Rg = 39.5 ± 0.4 Å）变体均出现显著的结构扩展，而甲醇则使 Q21 轻微紧缩（Rg = 28.5 ± 0.5 Å）。据我们所知，本研究首次利用 MD 考察了助溶剂效应对 HTT 外显子 1 构象动力学的影响。尽管单条轨迹的局限性使我们无法得出确切的机制性结论，但观察到的趋势表明，疏水共溶剂可能使 HTT 外显子 1 的构象系综向易聚集状态偏移，为研究 HD 中的基因—环境相互作用提供了计算框架。

关键词：亨廷顿病；多聚谷氨酰胺；分子动力学；亨廷顿蛋白外显子1；助溶剂；聚集；蛋白质构象动力学；环境暴露

1. 引言

亨廷顿病（HD）是一种进行性常染色体显性神经退行性疾病，由亨廷顿蛋白（*HTT*）基因外显子1中的不稳定 CAG 三核苷酸重复扩增所致 [1,2]。这种扩增会使亨廷顿蛋白 N 端区域内的多聚谷氨酰胺（polyQ）片段延长；重复长度超过约 36 个谷氨酰胺时与疾病病理相关，且片段越长，发病年龄越早、病情越严重 [3,4]。由此产生的突变型亨廷顿蛋白（mHTT）会发生错误折叠和聚集，形成神经元内包涵体，这是 HD 的病理标志 [5,6]。

由 HTT 外显子1编码的 N 端片段对于 体外和细胞模型中的聚集体形成既是必要的，也是充分的 [7,8]。该片段由三个功能各异的结构域组成：介导膜结合并调节聚集动力学的两亲性 N17 区域 [9,10]；其长度决定聚集倾向的多聚谷氨酰胺片段；以及被认为可作为内在聚集调节因子的 C 端多聚脯氨酸（polyP）区域 [11,12]。下游的 HEAT（亨廷顿蛋白、延伸因子 3、蛋白磷酸酶 2A、TOR1）重复结构域还会影响外显子1片段的结构环境，但其在聚集中的确切作用仍存在争议 [13,14]。

分子动力学（MD）模拟已成为一种强有力的工具，可在原子级分辨率下探究含多聚谷氨酰胺蛋白的构象动力学。以往的计算研究发现，多聚谷氨酰胺片段会形成异质构象集合，包括塌缩线团、β 发夹和伸展构象；随着片段长度增加，具备聚集能力的状态所占比例也会随之上升 [15–17]。针对 HTT 外显子 1 片段进行的粗粒化和全原子模拟表明，N17 结构域可形成 α 螺旋构象，并促成分子间接触

[18,19], 而 polyP 区域可能会对多聚谷氨酰胺的塌缩造成空间位阻 [11]。近期的副本交换和多微秒模拟为多聚谷氨酰胺片段长度依赖性的结构域间相互作用及构象采样变化提供了证据 [20–22]。尽管人们在理解多聚谷氨酰胺内在的聚集决定因素方面已取得显著进展，但对 HTT 构象动力学的环境调节因子关注仍相对较少。流行病学证据表明，环境暴露——包括三氯乙烯（TCE）等有机溶剂——可能影响神经退行性疾病的发病或进展 [23,24]。三氯乙烯与帕金森病风险升高有关，其作用机制涉及线粒体复合体 I 抑制和氧化应激 [25,26]。然而，尚未有研究采用分子动力学模拟系统考察有机助溶剂是否能够直接调节突变型 HTT 等易聚集蛋白质的构象景观。

在本研究中，我们采用全原子分子动力学模拟，表征 HTT 外显子1构建体对以下因素的构象响应：(i) 三种临床相关重复长度（Q21、Q40 和 Q70）下的 polyQ 区段长度变化；以及 (ii) 暴露于一组具有不同理化性质的有机助溶剂（甲醇、己烷、丙酮和三氯乙烯）。采用 OPLS-AA 力场，在显式 SPC/E 水中进行多链模拟（每个模拟盒含四条蛋白质链），生产运行时间为 100 ns。我们计算了包括回转半径（Rg）、溶剂可及表面积（SASA）、均方根偏差（RMSD）和蛋白质内氢键数量在内的构象代理指标，以评估结构紧密度、溶剂暴露和内部稳定作用。

我们强调，此处采用的指标——回转半径、溶剂可及表面积、均方根偏差和氢键数量——是反映结构特性的构象代理指标，与聚集有关，但并不等同于聚集。直接观察聚集事件需要

大幅延长模拟时间尺度（从多微秒到毫秒）和/或采用超出本研究范围的增强采样方法[20,27]。此外，所有模拟均为单条轨迹，且没有独立重复，因此限制了统计推断。尽管如此，观察到的聚谷氨酰胺长度依赖性趋势与实验观察及既往计算研究一致，而助溶剂数据则为探索与亨廷顿病发病机制相关的基因—环境相互作用提供了初步计算框架。

2. 方法

2.1. 蛋白质结构生成

使用 ColabFold v1.5.5 [28],（一种加速版 AlphaFold2 [29]，采用基于 MMseqs2 的多序列比对）生成 HTT 外显子1构建体。共建立了六种构建体模型：三种含有第一个 HEAT 重复结构域（含 HEAT），三种不含该结构域（不含 HEAT）；每种分别具有 Q21（野生型）、Q40（成人/晚发型阈值）和 Q70（少年型/早发型）多聚谷氨酰胺长度。输入序列由 N17 两亲性区域、多聚谷氨酰胺片段、多聚脯氨酸（polyP）区域，以及适用情况下的第一个 HEAT 重复结构域组成。ColabFold 使用默认参数运行（补充表 S1），并使用 pLDDT 置信度分数评估预测结构；所有构建体在结构化结构域（N17、HEAT 重复结构域）中均显示出预期的高置信度，而在内在无序的多聚谷氨酰胺和多聚脯氨酸区域中置信度较低。

2.2. 系统制备与模拟方案

所有分子动力学模拟均使用 GROMACS（2021.4 和 2025.2 版本）进行 [30]。每个模拟体系包含四个目标蛋白构建体副本，置于立方周期性盒子中，蛋白质与盒边缘之间的最小缓冲距离为 1.5 nm。蛋白质采用 OPLS-AA 力场进行参数化 [31], ，体系则用 SPC/E 水分子溶剂化 [32]。通过添加 Na^+ 和 Cl^- 抗衡离子使体系电荷达到中性。

使用最速下降算法进行能量最小化（离子定位时进行 5000 步，收敛阈值为 1000 kJ mol^{-1} nm^{-1}；随后进行 500000 步主最小化，阈值为 500 kJ mol^{-1} nm^{-1}）。之后，系统分两个阶段进行平衡：（i）在目标温度下使用 V-rescale 恒温器（τ_T = 0.1 ps）进行 1000 ps 的 NVT 平衡，并对所有蛋白质重原子施加位置约束；（ii）使用 Berendsen 压力耦合器进行 1000 ps 的 NPT 平衡（各向同性耦合，τ_P = 1.0 ps，可压缩性 = 4.5×10^{-5} bar^{-1}），同时保持位置约束。

生产模拟运行了 100 ns，积分时间步长为 2 fs。生产运行中，Parrinello–Rahman 压力耦合器 [33] 取代了 Berendsen 压力耦合器，以确保正确的系综采样。温度耦合采用 V-rescale 恒温器，并分别对蛋白质原子和非蛋白质原子设置耦合组。长程静电作用采用粒子网格Ewald（PME）方法 [34]（四阶插值，傅里叶间距为 0.1 nm）。短程范德华作用和库仑相互作用采用 1.4 nm 截断距离，并使用 Verlet邻居列表方案。所有涉及氢原子的键长均使用 LINCS算法 [35] 进行约束（四阶展开，单次迭代）。同时应用了能量和压力色散校正。完整的模拟参数见补充表S2。

2.3. 共溶剂模拟

为研究有机助溶剂对 HTT 外显子 1 构象动力学的影响，对选定的蛋白质构建体（Q21 和 Q40，无HEAT变体）在甲醇（0.5–1.0 M）、己烷（1.0 M）、丙酮（1.0 M）或三氯乙烯（TCE，1.0 M）存在下进行了模拟。助溶剂分子使用 LigParGen [36], 进行参数化，该工具提供与 OPLS-AA 兼容的参数。助溶剂分子与 SPC/E 水一同分布在模拟盒中。所有其他模拟参数与上述水相对照模拟保持一致。

2.4. 轨迹分析

轨迹分析使用 MDAnalysis [37,38] 和 GROMACS 内置工具进行。对每次模拟的完整 100 ns 轨迹计算了以下构象指标：

均方根偏差 (RMSD)：以能量最小化的初始结构为参照，通过对初始帧进行最小二乘拟合计算主链均方根偏差。RMSD 反映了相对于初始构象的整体结构漂移。

均方根涨落 (RMSF)：在生产阶段轨迹中计算逐残基均方根涨落，并对每次模拟中的所有四个蛋白质拷贝取平均值。均方根涨落可识别构象柔性较高的区域。

回转半径 (Rg)：Rg 计算为所有原子到蛋白质质心的质量加权均方根距离。Rg 可用于衡量整体结构紧密度；Rg 增大表示结构扩展。

溶剂可及表面积 (SASA)：使用半径为 1.4 Å 的球形探针计算 SASA，该半径约等于水分子的半径。SASA 用于量化暴露于溶剂中的蛋白质表面积；其增大可能表明发生解折叠或三级结构堆积丧失。

蛋白内氢键：依据标准几何判据，对每条蛋白质链内的氢键进行统计（供体–受体距离 ≤ 3.0 Å 、供体 – 氢 – 受体角 ≥150°，以及距离 ≤ 1.2 Å ）。氢键数量减少可能表明二级/三级结构丧失。

为进行定量比较，我们计算了每条轨迹最后 20 ns（80–100 ns）内的时间平均值和标准差，以聚焦于模拟的平衡阶段或准平衡阶段。需要指出的是，100 ns 可能不足以使所有指标完全收敛，尤其是对于较大的 Q70 构建体；这一局限性将在第 5 节讨论。

2.5. 分析范围与局限性

所有模拟均为单条轨迹（每个条件 n = 1），未进行独立重复。因此，所报告的标准差反映的是单条轨迹内的时间波动，而非独立运行之间的统计不确定性。每个模拟盒中的四个蛋白质拷贝处于相同的溶剂环境中，并经历相同的盒尺度波动，因此在统计学上不构成独立重复。这种设计限制了条件间正式统计比较的开展，也意味着应将观察到的趋势视为初步观察，并通过重复模拟或增强采样方法加以验证。

3. 结果

3.1. 水相对照中构象指标对多聚谷氨酰胺长度的依赖性为建立基线构象行为，我们首先在纯水溶剂中考察了 HTT 外显子 1 的三种多聚谷氨酰胺长度变体（不含 HEAT）。图 1 和表 1 总结了每条 100 ns 轨迹最后 20 ns（80–100 ns）内的关键构象指标平均值。

观察到回转半径呈现明显的多聚谷氨酰胺长度依赖性趋势：野生型Q21 构建体保持相对紧凑的结构（Rg = 32.06 ± 0.69 Å），Q40 晚发型变体的回转半径略有增加（Rg = 33.38 ± 0.63 Å），而 Q70 早发型变体则呈现明显更扩展的构象（Rg = 40.17 ± 0.50 Å），较 Q21 增加了 25%。溶剂可及表面积呈现相似趋势，从 Q21 的 25,705 ± 332 $Å^2$ 增至 Q40 的 29,906 ± 344 $Å^2$，再增至 Q70 的 37,844 ± 369 $Å^2$，表明随着多聚谷氨酰胺长度增加，暴露于溶剂中的蛋白质表面逐渐增多。

有趣的是，蛋白质内氢键数量也随多聚谷氨酰胺长度增加而上升：从 24.0± 4.2（Q21）升至 31.2 ± 4.5（Q40），再升至 50.3 ± 5.8（Q70）。这可能反映了较长的多聚谷氨酰胺片段中潜在氢键供体和受体数量更多，同时扩展的多聚谷氨酰胺区域在采样伸展构象和类似转角的构象时形成了瞬时链内接触。均方根偏差值分别为 42.42 ± 0.44 Å（Q21）、33.40 ± 0.67 Å（Q40）和 44.15 ± 1.50 Å（Q70），表明在所有情况下，相较于 AlphaFold2 预测的起始结构都发生了显著的结构重排，这与内在无序区的预期一致。

这些结果与主流模型一致：扩展的多聚谷氨酰胺片段会采取愈发伸展、暴露于溶剂的构象，并可能成为聚集易感状态的前体 [15–17]。然而，我们强调，在这些模拟的 100 ns 时间尺度内，并未观察到直接的聚集事件。

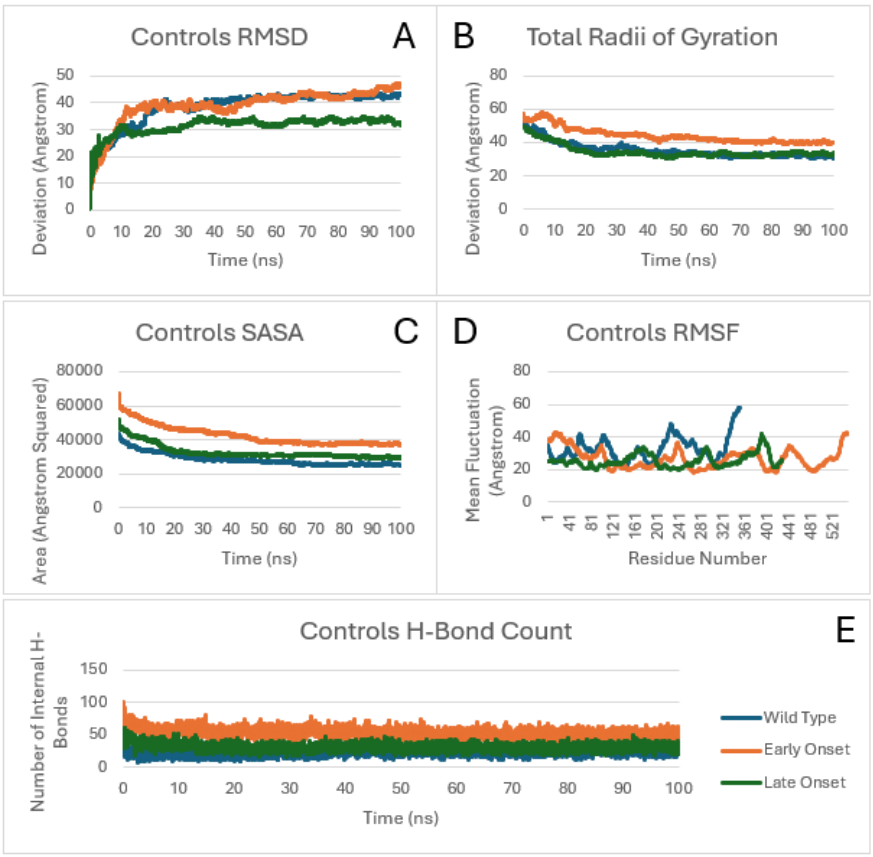


图 *1.* 非*HEAT*对照的分析汇总图，展示了 *A*）均方根偏差、*B*）总回转半径、*C*）总溶剂可及表面积、*D*）均方根涨落和 *E*）蛋白质氢键总数。

3.2. 助溶剂对野生型（Q21）HTT 外显子 1 的效应

我们在三种有机助溶剂存在的条件下对野生型Q21构建体进行了模拟，以评估环境分子是否会扰动 HTT 外显子 1 的构象景观。结果见表2和图 2。

甲醇（0.5 M）使 Q21 构建体略微紧缩，回转半径从 32.06± 0.69 Å（水相对照）降至 28.53 ± 0.50 Å，同时溶剂可及表面积也略有下降（ 25,283 ± 558 Å^2 对比 25, 705 ± 332 Å^2）。这与低浓度甲醇已知的类结构促成行为一致，该行为可稳定紧凑蛋白状态 [39]。蛋白内氢键降至 17.8 ± 3.4，这可能反映了溶剂与蛋白质之间氢键竞争的变化。

相比之下，己烷（1.0 M）促使结构扩展：回转半径增至 37.10 ± 0.85 Å（相较水相对照增加 +16%），溶剂可及表面积升至 31,331 ± 578 Å^2 （+22%）。氢键减少至 18.7 ± 4.1，表明内部稳定作用有所丧失。作为一种非极性疏水性溶剂，己烷可能优先与暴露的疏水残基结合，从而促进结构去折叠。

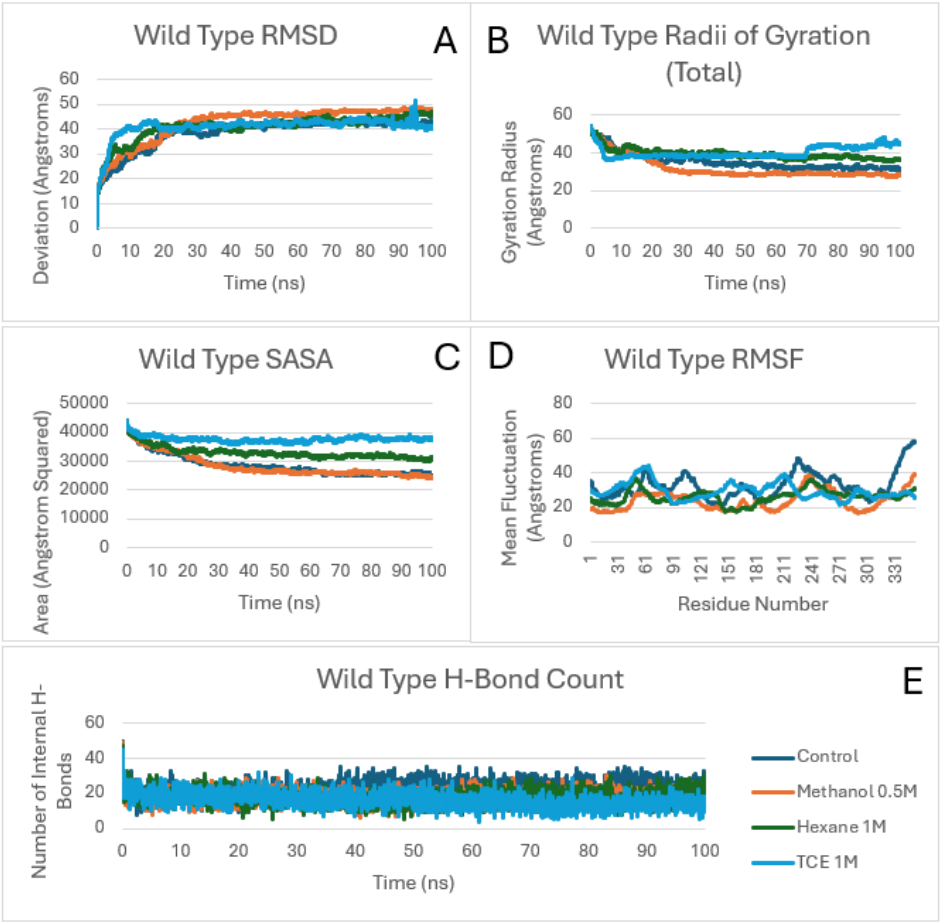


图 *2.* 非*HEAT* 野生型*HTT* 溶剂比较的分析汇总图，分别展示 *A*）均方根偏差、*B*）总回转半径、*C*）总溶剂可及表面积、*D*）均方根涨落和 *E*）蛋白质氢键总数。

在 TCE（1.0 M）作用下观察到的效应最为显著：Rg 达到 43.86 ± 1.26 Å（相对于对照为 +37%），同时 SASA 大幅增加至 37,729 ± 346 Å^2（+47%），氢键则显著减少至 13.8 ± 3.4（–42%）。TCE 存在时这种显著的结构扩展以及内部氢键的丧失值得注意，

因为流行病学研究已将 TCE 这一环境毒物与神经退行性疾病联系起来[25,26]。Q21 + TCE 的构象特征（Rg = 43.86 Å）接近纯水中病理性 Q70 构建体的构象特征（Rg = 40.17 Å），这提出了一种可能性——尽管现阶段仍属推测——即某些环境暴露可能会使野生型 HTT 的构象系综转向类似于疾病相关突变体的状态。

3.3. 共溶剂对晚发型（Q40）HTT 外显子 1 的影响

研究将共溶剂模拟扩展至 Q40 晚发型变体，以考察多聚谷氨酰胺扩增是否会改变其对共溶剂诱导的构象变化的易感性。表 3 和图 3 总结了结果。

甲醇（1.0 M）对 Q40 回转半径的影响极小（32.05 ± 0.48 Å，对照组为 33.38 ± 0.63 Å），其溶剂可及表面积（29,879 ± 302 $Å^2$）和氢键（33.5 ± 4.9）也与对照组相近。这与 Q21 中观察到的紧缩现象形成对比，说明扩展的多聚谷氨酰胺链段可能具有缓冲作用，减轻甲醇诱导的结构扰动。

己烷（1.0 M）使 Q40 明显扩展，回转半径增至 40.65 ± 1.02 Å（较对照组增加 +22%），溶剂可及表面积增至 37,678 ± 413 $Å^2$ （+26%），而氢键相对稳定（31.0 ± 4.8，对照组为 31.2 ± 4.5）。这种扩展模式与 Q21 +hexane 中观察到的相似，表明疏水溶剂介导的解折叠机制具有一致性。

丙酮（1.0 M）使 Q40 的回转半径略微增大至 34.01 ± 0.86 Å（对照组为 33.38 ± 0.63 Å），氢键数量则显著增加至 47.6 ± 6.2，这可能表明丙酮的双重亲水/疏水特性促进了新链内接触的形成。

TCE（1.0 M）使 Q40 的回转半径增至 39.50 ± 0.35 Å（+18%），其中最显著的变化体现在溶剂可及表面积上，其增至 43,023 ± 481 $Å^2$ （相较对照组增加 +44%）。氢键数量增至 42.0 ± 5.6，这可能反映了 TCE 诱导的重组，使

多聚谷氨酰胺片段暴露，同时促进形成新的主链接触。这种复杂的响应模式不同于 Q21 + TCE 中观察到的较简单的扩张并伴随氢键减少，这表明较长的多聚谷氨酰胺片段会调节蛋白质对疏水共溶剂的响应。

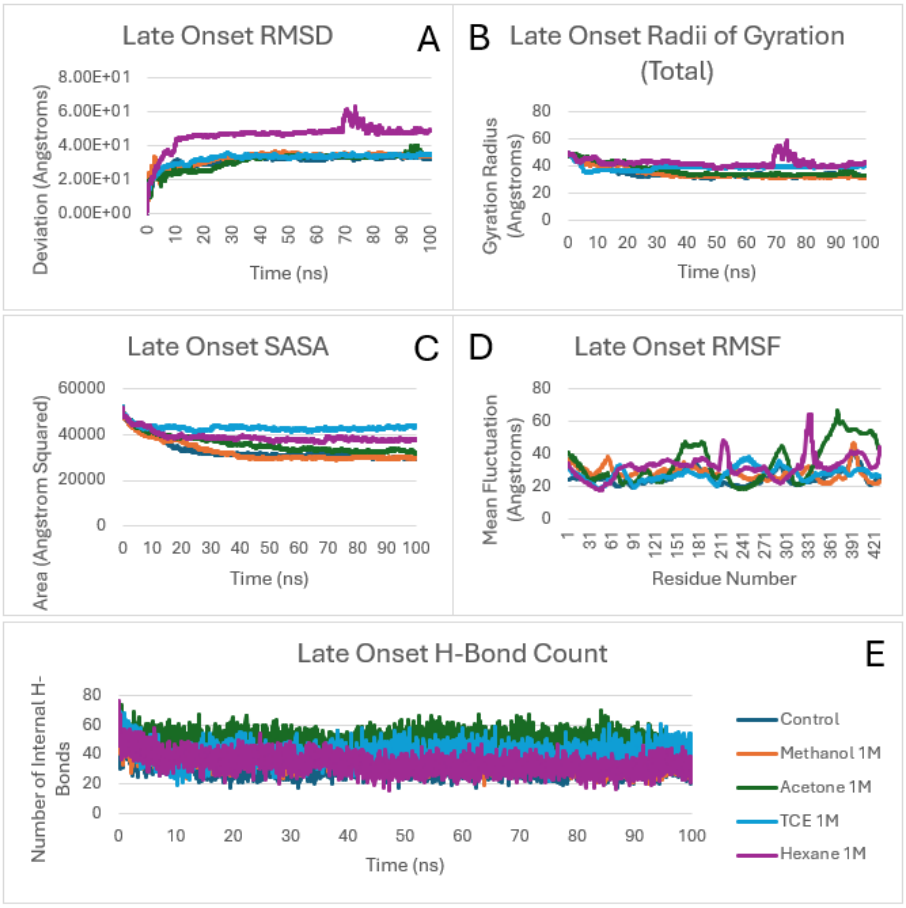


图 *3.* 非*HEAT lo-HTT* 溶剂比较的分析汇总图，展示了 *A)* 均方根偏差、*B)* 总回转半径、*C)* 总溶剂可及表面积、*D)* 均方根涨落和 *E)* 蛋白质氢键总数。

3.4. 构象趋势总结

在所考察的所有条件下，均呈现出两种主要趋势：(i) 在水相条件下，随着多聚谷氨酰胺长度增加，结构逐渐扩展，这可由回转半径和溶剂可及表面积的变化衡量；(ii) 疏水共溶剂，尤其是三氯乙烯和己烷，会加剧这种扩展，其影响程度取决于助溶剂的种类和多聚谷氨酰胺长度。甲醇则表现出不同的定性特征，在 Q21 中诱导紧缩，

但对 Q40 的影响很小。这些观察结果以示意图形式汇总于表4。需要指出的是，所有趋势均基于单轨迹模拟，仍需通过重复研究加以验证。

表 1. 水相对照模拟（无HEAT变体）的构象指标，取最后 20 ns 的平均值。

构建体	多聚谷氨酰胺长度	均方根偏差（Å）	回转半径（Å）	溶剂可及表面积 (Å2)	氢键
野生型	Q21	42.42 ± 0.44	32.06 ± 0.69	25,705 ± 332	24.0 ± 4.2
晚发型	Q40	33.40 ± 0.67	33.38 ± 0.63	29,906 ± 344	31.2 ± 4.5
早发型	Q70	44.15 ± 1.50	40.17 ± 0.50	37,844 ± 369	50.3 ± 5.8

表2. 野生型（Q21）HTT 外显子 1构象指标的助溶剂效应（无HEAT变体）。

条件	均方根偏差 (Å)	回转半径 (Å)	Δ回转半径 (%)	溶剂可及表面积 (Å2)	氢键
水相对照	42.42 ± 0.44	32.06 ± 0.69	—	25,705 ± 332	24.0 ± 4.2
+ 甲醇 (0.5 M)	47.61 ± 0.43	28.53 ± 0.50	−11%	25,283 ± 558	17.8 ± 3.4
+ 己烷 (1.0 M)	45.03 ± 0.95	37.10 ± 0.85	+16%	31,331 ± 578	18.7 ± 4.1
+ 三氯乙烯 (1.0 M)	42.24 ± 1.43	43.86 ± 1.26	+37%	37，729 ± 346	13.8 ± 3.4

表 3. 助溶剂对晚发型（Q40）HTT 外显子 1 构象指标的影响（无HEAT变体）。

条件	均方根偏差（Å）	回转半径（Å）	Δ回转半径 (%)	溶剂可及表面积（Å2）	氢键
水相对照	33.40 ± 0.67	33.38 ± 0.63	—	29,906 ± 344	31.2 ± 4.5
+ 甲醇 (1.0 M)	34.60 ± 0.70	32.05 ± 0.48	−4%	29,879 ± 302	33.5 ± 4.9
+ 己烷 (1.0 M)	48.42 ± 0.87	40.65 ± 1.02	+22%	37,678 ± 413	31.0 ± 4.8
+ 丙酮（1.0 M)	34.76 ± 1.55	34.01 ± 0.86	+2%	32,756 ± 451	47.6 ± 6.2
+ 三氯乙烯（1.0 M)	34.51 ± 0.51	39.50 ± 0.35	+18%	43,023 ± 481	42.0 ± 5.6

4. 讨论

4.1. 依赖于多聚谷氨酰胺长度的构象扩展

在我们的水相对照模拟中观察到的 Rg 和 SASA 随多聚谷氨酰胺长度逐渐增加，这总体上与关于 HD 中多聚谷氨酰胺扩增的生物物理学认识一致。采用荧光相关光谱 [4]，小角 X 射线散射（SAXS） [40], 和原子力显微镜（AFM） [6] 等技术进行的实验研究表明，较长的多聚谷氨酰胺片段会采取更伸展的构象，并表现出更强的聚集倾向。我们的模拟重现了这种依赖长度的扩展趋势，其中 Q70 的 Rg 比 Q21 高 25%。

观察到的趋势也与先前的分子动力学研究一致。Jakubek 等人 [15] 报告称，随着多聚谷氨酰胺链长度增加，其结构异质性和形成伸展状态的倾向也逐渐增强。Kang 等人 [22] 采用温度副本交换分子动力学（T- REMD）证明，扩展的多聚谷氨酰胺片段会采样更广阔的构象景观，且 β 折叠样结构的占比增加。Mohanty 等人 [20,21] 对 HTT 外显子 1 进行了多微秒聚集模拟，观察到依赖于多聚谷氨酰胺长度的结构域间相互作用和紧密程度变化。尽管我们的 100 ns 模拟远短于这些基准研究，但在依赖于多聚谷氨酰胺长度的扩展趋势方面，两者在定性上相符，这使我们有信心认为，我们的模拟捕捉到了构象响应的基本特征。

我们注意到，绝对均方根偏差值（33–44 Å）相较于典型球状蛋白质模拟而言较大，这反映了 HTT 外显子 1 中 polyQ 和 polyP 区域的内在无序特性。当起始结构是内在无序蛋白质的 AlphaFold2 预测结构时，均方根偏差值较大是意料之中的，因为模拟会迅速偏离静态预测构象，以探索可及的构象系综。因此，不应将均方根偏差解读为不稳定性的度量，而应将其视为对以下事实的确认：

模拟中的蛋白质会采样广阔的构象空间，这与其无序特性相符。

4.2. 共溶剂对构象景观的调节

据我们所知，助溶剂结果首次系统地考察了有机共溶剂对 HTT 外显子 1 构象动力学的影响。不同助溶剂和多聚谷氨酰胺长度下观察到的差异性响应表明，蛋白质–共溶剂相互作用受疏水效应、氢键竞争和优先溶剂化的复杂协同作用所支配。

甲醇使 Q21 略微紧缩（回转半径 Rg 降低 11%），但对 Q40 的影响可以忽略不计。这与低浓度短链醇的已知行为一致：它们可作为促结构剂，优先水合蛋白质表面并稳定紧凑态 [39]。Q40 的响应减弱可能反映了扩展的多聚谷氨酰胺链段具有更高的构象熵，因而不易紧缩。

己烷持续促进 Q21（回转半径 Rg 为 +16%）和 Q40（回转半径 Rg 为 +22%）发生结构扩展，并伴随溶剂可及表面积（SASA）增加。作为一种纯疏水性溶剂，己烷可能会破坏维持紧凑构象的疏水核心相互作用，但无法与蛋白质主链竞争形成氢键。Q21 + 己烷中观察到的蛋白内氢键减少（18.7，而对照组为 24.0）支持了这种由内部堆积破坏所致的机制。

在两种多聚谷氨酰胺长度下，TCE 均产生了最显著的影响。在 Q21 中，TCE 使回转半径增加了 37%，溶剂可及表面积增加了 47%，氢键减少了 42%，表明结构紧密度和内部稳定作用显著丧失。在 Q40 中，TCE 使回转半径增加了 18%，溶剂可及表面积增加了 44%，但氢键不减反增。这种氢键响应的差异颇为引人关注，可能反映了依赖于多聚谷氨酰胺长度的

TCE 相互作用模式差异：在 Q21 中，TCE 可能主要破坏已有接触；而在 Q40 中，较长多聚谷氨酰胺片段的重组可能创造新的链内氢键形成机会。

观察发现，TCE 暴露会使野生型Q21的构象特征向与病理性 Q70 构建体相似的指标偏移，这一结果颇具启发性。不过，我们强调，这一比较仅基于单轨迹模拟中的一项构象指标（回转半径）。TCE 暴露与神经退行性疾病风险之间的流行病学关联 [25,26] 使这一发现具有环境相关性，但要确立因果分子机制，还需要开展范围大得多的计算和实验研究。

4.3 机制考量

根据观察到的趋势，我们对助溶剂对 HTT 外显子 1 构象动力学的影响提出如下初步分类：

极性质子溶剂（甲醇）：在低浓度下，可能通过优先水合稳定紧凑态，但对扩展的多聚谷氨酰胺片段的作用较弱。

非极性溶剂（己烷）：通过破坏疏水堆积促进结构扩展，且对不同多聚谷氨酰胺长度的影响一致。

氯代溶剂（三氯乙烯）：引发最显著的扩展和构象重组，这可能是疏水作用受扰与其和多聚谷氨酰胺骨架发生特异性相互作用共同作用的结果。三氯乙烯效应的强度可能反映了其中等极性，以及其渗入蛋白质疏水区和极性区的能力。**极性非质子溶剂（丙酮）：**对全局指标的影响较温和，但可能促进链内氢键形成，表明其相互作用模式有所不同。

我们强调，这些分类只是暂定的，且基于未进行重复模拟的单轨迹模拟。应通过以下方法检验所提出的机制：(i) 使用独立初始条件进行重复模拟；(ii) 分析优先溶剂化系数；(iii) 按残基分解共溶剂接触；以及 (iv) 进行自由能计算。

4.4. 与文献比较

我们的水相对照结果与更广泛的 HTT 分子动力学模拟文献在定性上相符。Moldovean 和 Chiș [16] 回顾了 HTT 聚集的计算研究，并强调了多聚谷氨酰胺长度依赖的构象采样的重要性。Kang 等人[22] 报告称，在温度副本交换分子动力学模拟中，较长的多聚谷氨酰胺片段表现出更强的 β-折叠倾向，而这一结构特征与聚集成核有关。尽管我们在 100 ns 时间尺度上进行的标准分子动力学模拟无法直接捕捉 β-折叠形成的动态过程，但观察到的扩张和溶剂暴露增加，与群体向更不紧凑、更易聚集的状态转变相符。

共溶剂模拟将现有研究拓展到了新的领域。尽管共溶剂分子动力学已广泛用于蛋白质工程和结合位点定位 [39], ，但将其应用于神经退行性疾病相关蛋白尚属新颖。最相近的研究是关于渗透溶质对多聚谷氨酰胺肽影响的研究 [17], ，这些研究表明，氧化三甲胺等保护性渗透溶质能够稳定紧凑多聚谷氨酰胺构象。我们观察到疏水性助溶剂会使紧凑态不稳定，这一发现从机制上与上述研究互为补充，说明多聚谷氨酰胺蛋白的构象平衡对周围溶剂环境的化学性质十分敏感。

5. 局限性

必须承认，本研究存在若干重要局限。

单轨迹设计：所有模拟均为单条轨迹（每种条件 n = 1 ）。报告的标准差反映的是单条轨迹内的时间波动，而非独立重复模拟之间的统计不确定性。这从根本上限制了条件之间的正式统计比较，也意味着所观察到的趋势应视为初步观察，而非经统计验证的结论。未来研究应针对每种条件至少进行三至五次独立重复模拟，并采用不同的初始速度分配。

模拟时间尺度有限：100 ns 的生产运行，相较于聚谷氨酰胺聚集所需的时间尺度（微秒至毫秒），甚至相较于内在无序蛋白质实现充分构象采样所需的时间尺度，都很短。本文报告的指标反映了可用模拟时间窗口内的早期构象动力学和准平衡性质，但可能无法代表真正的热力学平衡系综。通过目视检查时间序列图（补充图）评估了回转半径和溶剂可及表面积的收敛情况。尽管大多数指标在分析窗口内似乎趋于平稳，但并未进行正式的收敛性检验（如分块平均或自相关分析）。

四拷贝系统设计： 每个模拟盒中使用四个蛋白质拷贝，旨在同时采样多条构象轨迹。然而，这些拷贝共享相同的溶剂环境和周期性边界条件，因此彼此并非独立。任何拷贝间相互作用（例如瞬时接触）都会使单链构象动力学与分子间缔合相混淆，不过目视检查轨迹时未观察到持久的拷贝间接触。

力场的局限性： OPLS-AA 力场虽然已得到广泛认可，但用于内在无序蛋白质时存在已知局限，可能会高估二级结构倾向，或低估对伸展构象的采样 [31]。近期开发的力场（例如 CHARMM36m、a99SB-disp）采用了改进的内在无序蛋白质参数化方法，可能会产生定量上不同的结果。力场的选择是多聚谷氨酰胺模拟中系统性不确定性的重要来源 [16]。

AlphaFold2 预测的起始结构：起始结构由 ColabFold/AlphaFold2 生成，该方法预测的是单一静态结构，对内在无序区的预测准确性有限。尽管模拟很快便偏离了这些起始结构（大均方根偏差值可证明这一点），但由于动力学陷阱的存在，初始构象的选择可能会影响 100 ns 轨迹中可及的构象景观。

助溶剂参数化：有机助溶剂采用 LigParGen 和 OPLS-AA 参数进行参数化。尽管这确保了与蛋白质力场的兼容性，但这些参数在所用浓度下对液体混合物模拟的准确性尚未经过独立验证。所用浓度（0.5–1.0 M）高于典型的环境暴露水平，因此应从构象敏感性的角度解读结果，而非将其视为具有生理相关性。

分析范围： 采用的构象指标（Rg、SASA、均方根偏差、氢键）是衡量蛋白质结构性质的成熟替代指标，但无法直接反映聚集。本研究未进行链间接触分析、二级结构指派或自由能计算。要直接评估聚集倾向，需分析分子间接触和成核事件，并采用显著更长的模拟时间尺度。

GROMACS 版本一致性： 模拟分别使用 GROMACS 2021.4 和 2025.2 进行。虽然不同版本的核心算法一致，但实现中的细微数值差异可能导致不同模拟组之间出现小幅差异。

早发型助溶剂数据不完整：本研究未完成 Q70（早发型）变体的助溶剂模拟，因此无法评估助溶剂效应在临床相关多聚谷氨酰胺长度全范围内的表现。

6. 未来方向

此处报告的初步观察提示，未来有若干值得深入探索的方向。首先，每种条件下进行重复模拟（$n \geq 3$–5），并采用独立生成的初始速度，对于确认观察到的趋势具有统计可信度、开展正式的假设检验至关重要。其次，采用复制交换分子动力学（REMD）、元动力学或自适应采样算法等增强采样方法，可在计算可行的时间尺度内更全面地探索构象自由能景观。第三，将助溶剂模拟扩展至 Q70 早发型变体，将补全多聚谷氨酰胺长度序列，并检验 Q21 中观察到的显著 TCE 诱导扩展是否会因多聚谷氨酰胺进一步扩增而增强或减弱。

对现有轨迹进行进一步分析，包括逐残基均方根涨落分解、二级结构指派（DSSP）、主成分分析（PCA）、链间接触频率图以及优先溶剂化系数计算，将有助于深入揭示其机制，这些分析计划在后续发表的论文中报告。自由能微扰或热力学积分计算则可量化助溶剂效应影响多聚谷氨酰胺构象平衡的能量基础。

最后，通过圆二色谱（CD）、动态光散射（DLS）或单分子荧光共振能量转移等生物物理技术对计算预测进行实验验证，将增强本研究的转化相关性，并明确计算机模拟观察结果是否反映了可通过实验测量的现象。

7. 结论

我们展示了亨廷顿蛋白外显子1构建体的全原子分子动力学模拟，这些构建体在水溶液中及有机助溶剂存在下，涵盖了三种具有临床相关性的多聚谷氨酰胺长度（Q21、Q40、Q70）。模拟显示，随着多聚谷氨酰胺长度增加，结构扩展和溶剂暴露增加的趋势十分明显，这与亨廷顿病中多聚谷氨酰胺驱动聚集的生物物理学基础一致。加入有机助溶剂，尤其是三氯乙烯（TCE）和己烷，会加剧这种结构扩展；而甲醇则使野生型Q21变体发生紧缩。

这些发现为研究环境助溶剂如何调节易聚集蛋白质的构象动力学，提供了初步的计算框架。观察结果表明，TCE 会使野生型 HTT 外显子 1 的构象指标向病理性多聚谷氨酰胺扩增变体的相似数值偏移。尽管这一推测仍需验证，但它引出了有关亨廷顿病发病机制中潜在基因—环境相互作用的有趣问题。

我们强调，这些结果源自时间尺度有限的单轨迹模拟，应将其视为用于提出假设的观察结果，而非确定性的机制结论。所采用的构象指标是聚集倾向的替代指标，并非对聚集本身的直接测量。尽管如此，不同多聚谷氨酰胺长度及助溶剂条件下呈现的一致趋势，为进一步开展更严谨的计算和实验研究奠定了基础，以探究环境调节因子对HTT错误折叠的影响。

参考文献

[1] MacDonald ME 等. 一种含有三核苷酸重复序列的新基因，该重复序列在亨廷顿病染色体中扩增且不稳定. Cell. 1993;72(6):971-983.[2] Bates GP 等. 亨廷顿病. Nat Rev Dis Primers. 2015;1:15005.[3] Andrew SE 等. 三核苷酸（CAG）重复长度与亨廷顿病临床特征之间的关系. Nat Genet. 1993;4(4):398-403.[4] Duyao M 等. 亨廷顿病中的三核苷酸重复长度不稳定性与发病年龄. Nat Genet. 1993;4(4):387-392.[5] DiFiglia M 等. 脑中神经元核内包涵体及营养不良性神经突内亨廷顿蛋白的聚集. Science. 1997;277(5334):1990-1993.[6] Scherzinger E 等. 亨廷顿蛋白编码的多聚谷氨酰胺扩展在体外和体内形成类淀粉样蛋白聚集体. Cell. 1997;90(3):549-558.[7] Mangiarini L 等. 带有扩增CAG重复序列的HD基因外显子1足以在转基因小鼠中引起进行性神经学表型. Cell. 1996;87(3):493- 506.[8] Wetzel R. 多聚谷氨酰胺的物理化学：单调序列中引人入胜的故事. J Mol Biol. 2012;421(4-5):466-490.[9] Thakur AK 等. 亨廷顿蛋白外显子1的N端被多聚谷氨酰胺扰动后，会触发复杂的聚集机制. Nat Struct Mol Biol. 2009;16(4):380-389.[10] Jayaraman M 等. 含短多聚谷氨酰胺的亨廷顿蛋白片段通过富含α螺旋的寡聚中间体缓慢发生淀粉样蛋白成核. J Mol Biol. 2012;415(5):881-899.[11] Bhattacharyya A 等. 多聚谷氨酰胺聚集成核：高度不利的蛋白质折叠反应的热力学. Proc Natl Acad Sci USA. 2005;102(43):15400-15405.[12] Darnell G 等. 侧翼多聚脯氨酸序列通过诱导PPII样螺旋结构抑制多聚谷氨酰胺片段中的β折叠结构. J Mol Biol. 2007;374(3):688-704.[13] Palidwor GA 等. 利用神经网络检测α杆状蛋白重复序列并将其应用于亨廷顿蛋白. PLoS Comput Biol. 2009;5(3):e1000304.[14] Guo Q 等. 亨廷顿蛋白的冷冻电子显微镜结构. Nature. 2018;555(7694):117-120.[15] Jakubek RS 等. 多聚谷氨酰胺的溶液态结构倾向性取决于重复长度. J Phys Chem B. 2019;123(22):4193-4203.[16] Moldovean SN, Chis V. 分子动力学模拟在亨廷顿蛋白结构与动力学转变研究中的应用：综述. ACS Chem Neurosci. 2020;11(2):105-120.[17] Escobedo A 等. 侧链与主链之间的氢键可稳定转录因子中的多聚谷氨酰胺螺旋. Nat Commun. 2019;10(1):2034.

[18] Haaga J 等. 多聚谷氨酰胺聚集早期阶段的计算研究. 《蛋白质》. 2018;86(9):1001-1009.[19] Falk AS 等. 亨廷顿蛋白外显子 1 原纤维中富脯氨酸结构域的结构模型. 《生物物理学杂志》. 2020;119(10):2019-2028.[20] Mohanty D 等. 亨廷顿蛋白外显子1中的结构域间相互作用：分子动力学模拟带来的见解. J Phys Chem B. 2024;128(22):5349-5362.[21] Mohanty D 等. 亨廷顿蛋白外显子1的构象动力学：一项多微秒研究. 《蛋白质科学》. 2025;34(6):e70188.[22] Kang H 等. 超紧凑亨廷顿蛋白外显子1突变体结构中出现的富含 β 折叠的构象. 《美国化学会志》. 2017;139(26):8820-8827.[23] Goldman SM. 环境毒素与帕金森病. 《药理学与毒理学年度评论》. 2014;54:141-164.[24] Ilieva NM 等. 帕金森病中的环境暴露与 LRRK2. 《NPJ 帕金森病》. 2024;10(1):103.[25] De Miranda BR, Greenamyre JT. 三氯乙烯：一种普遍存在的环境污染物与帕金森病风险. 《环境科学：过程与影响》. 2020;22(3):543-554.[26] Dorsey ER 等. 三氯乙烯：帕金森病的一种隐形病因？《帕金森病杂志》. 2023;13(2):203-218.[27] Zheng W 等. 通过微秒级原子级模拟探究蛋白质凝聚体的分子细节. J Phys Chem B. 2020;124(51):11671-11679.[28] Mirdita M 等. ColabFold：让所有人都能进行蛋白质折叠. 《自然·方法》. 2022;19(6):679-682.[29] Jumper J 等. 使用 AlphaFold 进行高精度蛋白质结构预测. Nature. 2021;596(7873):583-589.[30] Abraham MJ 等. GROMACS：通过从笔记本电脑到超级计算机的多级并行实现高性能分子模拟. 《SoftwareX》. 2015;1-2:19-25.[31] Jorgensen WL 等. OPLS 全原子力场在构象能量学和有机液体性质方面的开发与测试. 《美国化学会志》. 1996;118(45):11225-11236.[32] Berendsen HJC 等. 有效对势中缺失的项. J Phys Chem. 1987;91(24):6269-6271.[33] Parrinello M, Rahman A. 单晶中的多晶型转变：一种新的分子动力学方法. J Appl Phys. 1981;52(12):7182-7190.[34] Essmann U 等. 一种平滑的粒子网格 Ewald 方法. J Chem Phys. 1995;103(19):8577-8593.[35] Hess B 等 . LINCS：用于分子模拟的线性约束求解器. 《计算化学杂志》. 1997;18(12):1463-1472.

[36] Dodda LS 等. LigParGen 网络服务器：用于有机配体的自动 OPLS-AA 参数生成器. 《核酸研究》. 2017;45(W1):W331-W336.[37] Michaud-Agrawal N 等. MDAnalysis：用于分析分子动力学模拟的工具包. 《计算化学杂志》. 2011;32(10):2319-2327.[38] Gowers RJ 等. MDAnalysis：用于快速分析分子动力学模拟的 Python 软件包. 第15届 Python 科学会议论文集. 2016;98-105.[39] Ren P 等. 生物分子静电学与溶剂化：计算视角. 《生物物理学评论季刊》. 2012;45(4):427-491.[40] Vitalis A, Pappu RV. 评估寡聚体和单体的异质分布对聚集热力学的贡献. 《生物物理学杂志》. 2011;101(2):494-503.

图注

图 1. 不含 HEAT 的 HTT 外显子 1 变体在水相对照模拟中的构象指标。Q21（野生型）、Q40（晚发型）和 Q70（早发型）构建体的主链均方根偏差（Å）、均方根涨落（按残基计算，对四条链取平均）、回转半径（Rg，Å）、溶剂可及表面积（Å^2）及蛋白质内氢键数量的时间序列图（0–100 ns）。

图 2. 助溶剂对野生型（Q21）HTT 外显子 1 构象动力学的影响。图示为 Q21 在甲醇（0.5 M）、己烷（1.0 M）和 TCE（1.0 M）存在时的时间序列图，指标同图 1。

图 3. 助溶剂对晚发型（Q40）HTT 外显子 1 构象动力学的影响。图示为 Q40 在甲醇（1.0 M）、己烷（1.0 M）、丙酮（1.0 M）和 TCE（1.0 M）存在时的时间序列图，指标同图 1。

补充图 S1–S6。ColabFold 预测的全部六种 HTT 外显子1构建体（含 HEAT 和无 HEAT 变体，分别为 Q21、Q40、Q70）的结构及 pLDDT 置信度分数。