arXiv:2301.04424v1 [math.DG] 11 Jan 2023

黎曼几何与分子相似性 II：凯勒量子化

预印本

Rachael Pirie纽卡斯尔大学自然与环境科学学院 r.pirie2@ncl.ac.uk

Stuart J. Hall纽卡斯尔大学数学、统计学与物理学院 stuart.hall@ncl.ac.uk

Daniel J.Cole纽卡斯尔大学自然与环境科学学院 daniel.cole@ncl.ac.uk

Thomas Murphy
数学系 f Math
学，加利福尼亚州立大学富勒顿分校 tmurphy@fullerton.edu

2023年1月12日

摘要

分子之间的形状相似性是化学家用于虚拟筛选的一种工具，旨在降低药物发现项目的成本和所需时间。本文报告了一种全新的形状描述符，可作为此前所述 RGMolSA 描述符的替代方案 [1], ; 该描述符源自黎曼几何与凯勒量子化理论（KQMolSA）。将分子视为一系列相交球体，可使我们获得明确的黎曼度量，从而刻画曲面的几何结构，并进一步用于计算厄米矩阵 $\mathbb{M}$，作为可直接比较的曲面表示。我们通过一系列被认为形状相似的 PDE5抑制剂，展示了该方法的潜在效用。该方法在处理不同构象异构体方面展现出良好前景，并且与现有形状相似性方法相比表现良好。用于生成这些结果的代码和数据可在以下网址获取：https://github.com/RPirie96/KQMolSA。

关键词 黎曼几何 · 凯勒量子化 · 分子形状 · 基于配体的虚拟筛选

1 引言与第一部分总结

相似分子具有相同生物活性的概念广泛应用于药物发现 [2]。已知具有活性的分子可用作模板，从大型数据库中筛选其他潜在命中化合物。与单独进行实验筛选相比，这种方法效率更高，也能覆盖更广阔的化学空间 [3]。由于蛋白质与药物之间的形状互补性是实现强结合的必要条件，根据三维形状估算分子间的相似性日益受到关注。然而，形状并没有固定的定义。相反，比较依赖于对分子形状的数学近似，其依据包括分子体积、原子间距离的分布或曲面（通常视为范德华表面或溶剂可及表面） [4]。

在随附的论文 [1], 中，我们介绍了 RGMolSA 方法。该方法中开发的描述符使用一个简单的九元素向量来近似分子表面的形状，其中包含表面积以及常规拉普拉斯算子的前八个非零特征值的近似值。该描述符可视为对黎曼度量的近似；黎曼度量是描述曲面形状的基础数学对象。本文介绍一种截然不同的方法，借助凯勒量子化理论中的思想来近似黎曼度量；我们将这种方法称为分子表面近似的凯勒量子化（KQMolSA）。这套理论最初由数学家和弦论学家发展，用于明确表示物理理论中出现的四维对象（卡拉比–丘流形）的形状（有关率先将其用作数值技术的论文，请参见 [5]

）。简而言之，我们将一个称为凯勒势的函数与度量关联起来。接着，我们计算一个类似于该函数泰勒展开的表达式，其系数存储在厄米矩阵中。虽然这些矩阵本身确实取决于分子表面在三维空间中的具体位置和参数化方式 $\mathbb{R}^3$，但这种依赖关系很容易计算。因此，我们可以在厄米矩阵构成的“量子化”空间中进行计算，并据此为两个分子表面的形状赋予一个距离。最终得到的距离与分子的位置及其参数化方式的选择无关。

1.1 先前工作的总结

与随附论文 [1], 一样，我们的方法首先将分子视为一系列相交球体，其半径取组成原子的范德华半径。我们假设曲面的亏格为零，因此将任何环（例如苯）替换为半径为 2.25 埃的单个球体，以便于处理。随后，分子结构由球体的数量 N（每个环计为一个球体，且不包括任何氢原子）、每个球体的中心 c_i 和半径 r_i，以及描述球体相交情况的邻接矩阵 T 定义，其中

$$T_{ij} = \begin{cases} 1 & \text{if spheres } i \text{ and } j \text{ intersect} \\ 0 & \text{otherwise (or } i = j). \end{cases}$$

分子的表面积 $\mathcal{A}$ 按每个球体的面积减去两个球体相交时的“缺失部分”计算：

$$\mathcal{A} = 2\pi \sum_i \left(2r_i^2 - \left(r_i \sum_j T_{ij} |r_i - \lambda_{ij}| \right) \right). \tag{1}$$

利用该值对每个初始构型进行重新缩放，使分子的表面积等于单位球的表面积（或 4π），以解决黎曼几何将仅尺寸不同的两个物体视为具有相同形状这一问题。最终描述符中会对这种重新缩放进行一定加权，以免其主导相似度计算。

根据初始数据，构造一个映射，将曲面“展开”到复平面 $\mathbb{C}$ 上，这一过程称为分片立体投影。为构造该映射，需要选定一个原子作为起点，我们将其称为基球。具体而言，先求出分子的形心，再选取与该点欧几里得距离最小的原子作为基球。由映射 $\Phi_{ps}\colon \mathbb{C} \to \mathcal{S} \subset \mathbb{R}^3$ 诱导的黎曼度量 $g = \Phi_{ps}^*(g_{Euc})$形式如下

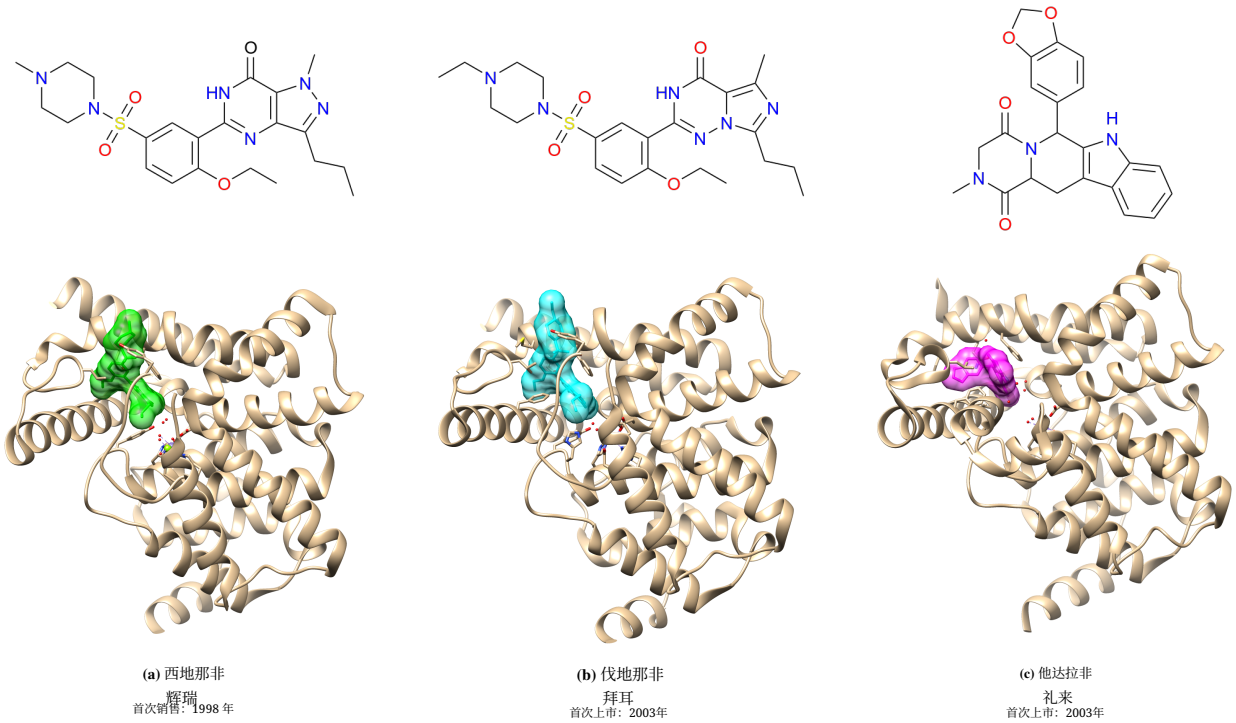
$$g = \begin{cases} \frac{4r_B^2}{(1+|z|^2)^2}(dx^2+dy^2) & \text{if } z \in \mathcal{C} \\ \frac{C_1}{(|z-A_1|^2+B_1)^2}(dx^2+dy^2) & \text{if } z \in \mathbb{D}(a_1, R_1), \\ \frac{C_2}{(|z-A_2|^2+B_2)^2}(dx^2+dy^2) & \text{if } z \in \mathbb{D}(a_2, R_2), \\ \vdots & \vdots \\ \frac{C_{N-1}}{(|z-A_{N-1}|^2+B_{N-1})^2}(dx^2+dy^2) & \text{if } z \in \mathbb{D}(a_{N-1}, R_{N-1}), \end{cases} \tag{2}$$

其中， r_B 为基球的半径，而

$$\mathcal{C} = \mathbb{C} \backslash \mathbb{D}(a_1, R_1) \cup \mathbb{D}(a_2, R_2) \cup \cdots \cup \mathbb{D}(a_{N-1}, R_{N-1}),$$

是.....的补集　　圆盘 $\mathbb{D}(a_1, R_1) \ldots, \mathbb{D}(a_{N-1}, R_{N-1})$的补集，对应于基球面上的点　　e.

RGMolSA描述符利用分段立体投影所提供的黎曼度量显式形式，来近似拉普拉斯算子的低阶特征函数 Δ。在 [1], 中，我们比较了西地那非、伐地那非和他达拉非的RGMolSA描述符。这几种PDE5抑制剂已知在其靶蛋白的结合口袋中占据相似的体积，因此形状也相似（图1） [6]。伐地那非是典型的“me-too”药物，仅对西地那非的结构做了少量小幅修改。由于这两种药物的化学结构高度相似，预计它们的形状也非常接近。另一方面，他达拉非在化学结构上与另外两种药物截然不同，但观察PDE5口袋中的分子可见，它们占据相似的结合构象，因此预计形状也相似。本文为便于与先前的工作 [1], 进行比较，再次以这三种分子为基础，研究新的形状描述符。



(a) 西地那非
辉瑞
首次销售：1998 年

(b) 伐地那非
拜耳
首次上市：2003年

(c) 他达拉非
礼来
首次上市：2003年

图1： 已知具有形状相似性的 PDE5抑制剂西地那非、伐地那非和他达拉非。他达拉非（化学结构不同、形状相似）是从首创药物西地那非进行骨架跃迁的一个例子，性能有了大幅提升；而伐地那非（一种“me-too”后续药物）仅有小幅改进。

虽然研究发现RGMolSA能够很好地描述形状，但它也存在一个潜在缺陷：结果取决于基球的选择，而基球又决定了用于计算积分、进而构建描述符的试探函数。基球附近的曲面几何特征描述得很好，但对于距离更远的原子，要准确描述曲面则需要更多特征值。分子越大，这一问题越突出；当分子足够大时，甚至会引入数值误差。我们通过忽略数值半径小于 10^{-9}的区域所产生的贡献来处理这类误差；然而，这会给形状描述符强加某种人为的“局域性”，意味着它可能只能准确捕捉基球附近的形状。

在下一节中，我们将概述 KQMolSA 描述符背后的理论。该理论同样利用黎曼度量来近似曲面的几何。由此得到的描述符位于流形 $GL(N,\ \mathbb{C})/U(N)$ 中，从而给出分子几何的全局描述符，并降低其对起始位置的依赖。图 2 以西地那非为例，总结了计算这些描述符的步骤。虽然描述符本身确实取决于所作的选择以及曲面在 $\mathbb{R}^3$ 中的位置，但在空间 $GL(N,\ \mathbb{C})/U(N)$ 中很容易对此加以处理。因此，计算形状描述符之间的“距离”尤其简单。

2 凯勒量子化的数学

2.1 理论概述

我们首先要说明，凯勒量子化理论过于高深，无法在本文中详加阐述。对于具备足够数学背景的读者，文献 [5] 对此有很好的介绍（也是最初将其用作数值技术的论述）。目前，两位作者正在撰写一份面向具有一般科学背景读者的数学理论介绍 [7]。

该理论研究复流形的几何（局部看起来像 $\mathbb{C}^n$ 的形状）；任何位于 $\mathbb{R}^3$ 中的曲面都是复流形，因为它在局部看起来像复数 $\mathbb{C}$ 的一个副本（即 $n = 1$）。更具体地说，我们将研究与 $\mathbb{S}^2$ 在拓扑上等价的曲面；在复流形的术语中，球面通常称为黎曼球面，记作 $\mathbb{CP}^1$。对曲面拓扑结构的这一限制是合理的，因为化学家认为，在大多数药物类分子中，环中心不会发生任何活动

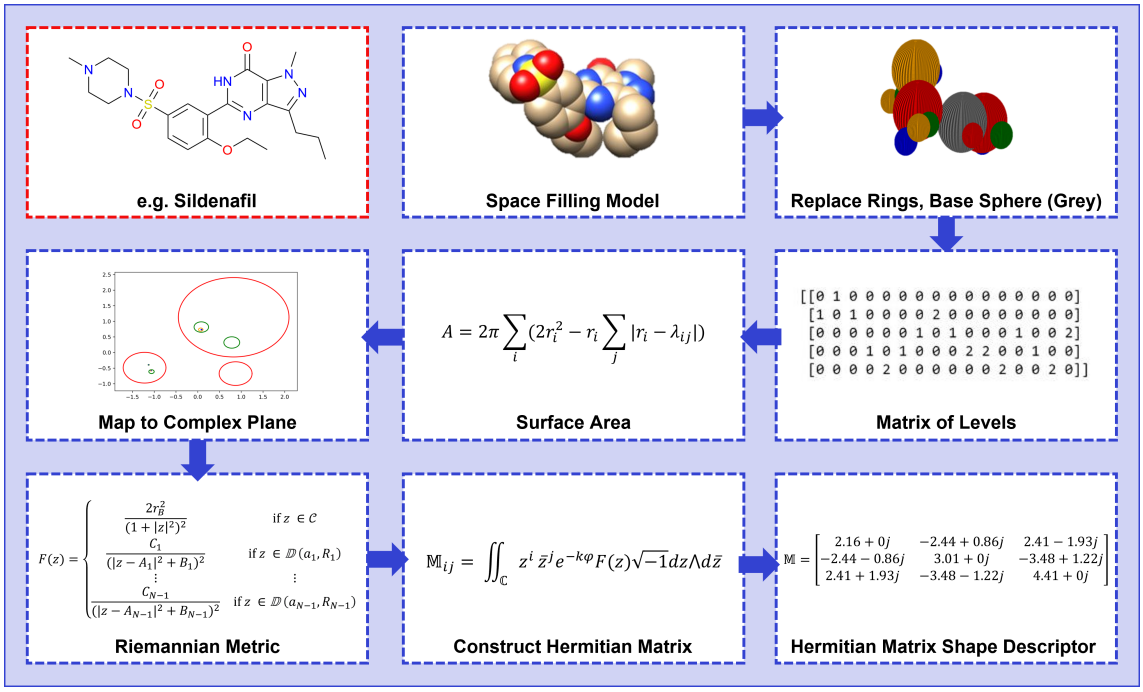


图 2： 计算西地那非（一种 PDE5抑制剂）的 KQMolSA 曲面描述符所涉及的关键步骤。

。例外情况是大环分子（含有超过 12 个原子的大环），其中环中心确实会发生活动。因此，这类分子不在所提出的两种方法的比较范围内。

处理复流形时，自然会用到的函数类是复可微函数，通常称为全纯函数。我们考虑一个一般的复流形 X；遗憾的是，如果流形 X 是紧致的，那么唯一的全纯函数 $f\colon X \to \mathbb{C}$ 都是常数。因此，我们不能指望仅通过研究 X上的全纯函数来理解 X。全纯函数概念的一种推广，是以 X为基底的全纯线丛 L 的截面。对于熟悉相关理论的读者来说，函数就是平凡丛的截面。如果线丛上存在具有正曲率的厄米度量 h，则称该线丛 L 为正的。小平 [8] 的一项基础性结果指出，如果线丛 L 为正，那么当 k 足够大时，其张量幂 L^k有许多全纯截面。事实上，所有此类截面组成的空间记作 $H^0(L^k)$，它是一个复向量空间，其维数的阶为 $O(k^n)$，当 $k \to \infty$时。

正曲率厄米度量 h 的曲率会产生一种称为 凯勒形式 的对象 ω，它进而会产生一个黎曼度量 g（数学中用来描述形状的对象）。事实证明，线丛 L 上所有正曲率厄米度量的集合，可以与满足以下条件的所有 $\bar{}$实值函数 φ 的集合等同： $X \to \mathbb{R}$ 在某个局部坐标 z中，它们满足 $\partial\bar{\partial}$-方程

$$\sqrt{-1}\partial\bar{\partial}\varphi = \omega - \omega_0$$

其中 ω 是该度量的凯勒形式， ω_0 是一个固定的参考凯勒形式。我们将在第2.3节详细介绍这些$\bar{}$微分算子 ∂ 和 $\bar{\partial}$；特别是，我们将说明，在分子表面这一情形中，$\partial\bar{\partial}$-方程实际上就是平面上熟悉的泊松方程。函数 φ 称为ω的凯勒势。相应的势并不唯一，但任意两个势只相差一个常数；这不会影响通过对势求两次导数构造出的度量。然而，我们将看到，在势中添加常数会使我们生成的、作为形状描述符的厄米矩阵按一个正实数缩放，因此在距离计算中，我们需要找到“最优”的缩放因子。

总而言之，对于一个正厄米线丛（L， h），我们有： $\to X$

- 一个凯勒形式 ω 和一个凯勒势 φ： $X \to \mathbb{R}$,
- 一个复向量空间 $H^0(L^k)$。

凯勒量子化的实质，是将由凯勒势（一个无限维函数空间）描述的几何，与有限维复向量空间 $H^0(L^k)$ 联系起来。这一主题贯穿数值分析和形状描述，例如傅里叶分析、球谐函数和泰勒级数等理论，它们都从某个无限维函数集合中构造出一个有限维向量空间。

2.2 量子化与田刚定理

数据（L, h）$\to X$ 在截面向量空间 $H^0(L^k)$ 上自然诱导出一个 $\mathcal{L}^2$-内积。给定截面s_1, $s_2 \in H^0(L^k)$，我们计算

$$\langle s_1, s_2 \rangle := \int_X h_k(s_1, s_2) \frac{\omega^n}{n!},$$

其中，h_k 是由 h 在 L^k 上诱导的厄米度量，而 $\omega^n/n!$ 是由凯勒形式产生的体积元。正是这个内积构成了数据（L, h）$\to X$的量子化。复 N 维向量空间上所有（厄米）内积构成的空间可以看作 $GL(N;\mathbb{C})/U(N)$。这是一个负曲率对称空间，并具有自然的距离概念；我们将用此距离来衡量形状相似性（见第 2.5 节）。

为了从量子化中恢复由（L, h）$\to$ X 定义的几何，我们选取向量空间 $H^0(L^k)$ 的一组基 $\{s_j\}$，由此得到内积的矩阵表示

$$\mathbb{M}_{ij} := \langle s_i, s_j \rangle.$$

若令 v 为截面向量

$$v = (s_1, s_2, \ldots s_N),$$

则可以定义 一个凯勒势（回想一下，截面是局部定义的全纯函数） $\tilde{\varphi}$ by

$$\tilde{\varphi} := -\frac{1}{k} \log \left(v^* \mathbb{M}^{-1} v \right).$$

定理 2.1（田刚， [9]）。设（X, L, h）是一个带有全纯线丛 L 和正曲率厄米度量 hwi 曲率 ω。如果我们再构造一个凯勒形式

$$\widetilde{\omega} = \omega_0 + \sqrt{-1}\partial\bar{\partial}\tilde{\varphi},$$

then

$$\|\omega - \widetilde{\omega}\|_{C^0} = O(k^{-2}).$$

换句话说，根据这个定理，任何由凯勒势 φ 所得的凯勒形式，都可以用由“代数”函数 $\tilde{\varphi}$ 所得的凯勒形式很好地近似。如果我们选取局部复坐标 z_1, z_2, $\ldots$, z_n，那么项 $v^*\mathbb{M}^{-1}v$ 就只是坐标的一个幂级数。对于分子表面来说，我们会得到类似多项式的东西。这就是函数 $\tilde{\varphi}$ 与原函数的截断泰勒级数 φ 相似的意义。该定理接着说明，这个级数确实收敛。

田刚定理针对光滑度量（即可以对凯勒势 φ 求任意阶导数的度量）表述；但在实际应用中（见第2.3节），我们将使用势属于 $C^2(X)$ 的度量，也就是说，势是二次连续可微的。关于以代数方式近似这类度量的理论尚未成文，但我们将展示，这种方法确实能够进行有意义的形状比较。我们预计，只要针对这一情形作适当调整，类似田刚定理的结果仍然成立；例如， [10] 讨论了较低正则性势的情形。

2.3 实践中的实现

如前所述，在实际应用中，我们取 $X = \mathbb{CP}^1$ 黎曼球面，并将线丛取为反典范丛$K^*_{\mathbb{CP}^1} = \mathcal{O}(2)$。凯勒形式 ω可以由黎曼度量 g显式构造；在分片立体投影映射 Φ_{ps}所提供的坐标中，我们可以利用度量（2）的形式得到

$$\omega = F(z)\sqrt{-1}dz \wedge d\bar{z},$$

其中 $F\colon\ \mathbb{C}\to\mathbb{R}_+$ 是由下式给出的“度量函数”

$$F(z)=\begin{cases}\frac{2r_B^2}{(1+|z|^2)^2} & \text{if } z\in\mathcal{C},\\ \frac{C_1}{(|z-A_1|^2+B_1)^2} & \text{if } z\in\mathbb{D}(a_1,R_1),\\ \frac{C_2}{(|z-A_2|^2+B_2)^2} & \text{if } z\in\mathbb{D}(a_2,R_2),\\ \vdots & \vdots\\ \frac{C_{N-1}}{(|z-A_{N-1}|^2+B_{N-1})^2} & \text{if } z\in\mathbb{D}(a_{N-1},R_{N-1}).\end{cases}\tag{3}$$

请注意，我们已在度量 g 中，用反对称形式（$\sqrt{-1}/2$）$dz\wedge d\bar{z}$ 替换了实对称二阶张量 dx^2+dy^2，其中 $dz=dx+\sqrt{-1}dy$ 和 $d\bar{z}=dx-\sqrt{-1}dy$。

为求凯勒势 $\varphi\colon\ \mathbb{C}\to\mathbb{R}$，我们求解“$\partial\bar{\partial}$-方程”

$$\omega=\sqrt{-1}\partial\bar{\partial}\varphi.$$

若考虑复微分算子

$$\frac{\partial}{\partial z}=\frac{1}{2}\left(\frac{\partial}{\partial x}-\sqrt{-1}\frac{\partial}{\partial y}\right)\qquad\text{and}\qquad\frac{\partial}{\partial\bar{z}}=\frac{1}{2}\left(\frac{\partial}{\partial x}+\sqrt{-1}\frac{\partial}{\partial y}\right),$$

那么，$\partial\bar{\partial}$-方程等价于求解泊松方程

$$\frac{\partial^2\varphi}{\partial z\partial\bar{z}}=\frac{1}{4}\Delta_{Euc}\varphi=F,$$

其中 Δ_{Euc} 是通常的二维拉普拉斯算子。我们可以显式求解泊松问题以求得 φ。该解可视为由两部分组成：一部分是“局部”部分，只需观察即可得出

$$\frac{\partial^2}{\partial z\partial\bar{z}}\left(\frac{C\log(|z-A|^2+B)}{B}\right)=\frac{C}{(|z-A|^2+B)^2},$$

以及一个“修正项”，之所以这样命名，是因为需要这一项来确保函数属于 $C^2(\mathbb{C})$。修正项是以下函数的线性组合 形式的函数

$$\log(|\alpha z+\beta|^2),$$

其中 w 每个球面得到一项。由于每个修正项都是调和函数，因此 s

$$\Delta\log(|\alpha z+\beta|^2)=0,$$

修正项之和仍是泊松方程的一个解。看起来修正项在点 $z=-\beta/\alpha$ 处是奇异的；然而，这些点总是位于函数取此特定形式的圆盘之外。我们将势的形式记录为一个定理，并请读者参阅附录（第 5 节），了解该解的推导。

定理 2.2（凯勒势的形式）。设 g 的形式为公式（2）。在与 i^{th} 球面相关的区域中，凯勒势可写为

$$\varphi(z)=\frac{C_i}{B_i}\log(|z-A_i|^2+B_i)+\sum_{j=1}^{N}\mathcal{K}_{ij}\log(|\alpha_{ij}z+\beta_{ij}|^2),$$

其中 $\mathcal{K}\in M^{N\times N}$（$\mathbb{R}$），且 $\alpha,\beta\in M^{N\times N}$（$\mathbb{C}$）。

前一定理中的矩阵 $\mathcal{K}$，α，和 β 可根据与分子相关的几何数据轻易计算，因此凯勒势的显式描述十分直接。

整体截面空间 $H^0(\mathcal{O}(2k))\cong\mathbb{C}^{2k+1}$ 可视为由以下函数张成的空间

$$\langle 1,z,z^2,\dots,z^{2k}\rangle.$$

因此，与该曲面相关联的形状描述符是厄米矩阵 $(2k+1)\times(2k+1)$ $\mathbb{M}$ 其中（考虑从 0 到 $2k$的指标）

$$\mathbb{M}_{ij}=\iint_{\mathbb{C}}z^i\bar{z}^je^{-k\varphi}F(z)\sqrt{-1}dz\wedge d\bar{z}.\tag{4}$$

2.4 计算相关积分

对公式（4）所述积分进行朴素的数值计算会带来两个明显的问题：首先，积分域无界（即整个复平面 $\mathbb{C}$）；其次，用于描述度量和凯勒势 φ 的区域及数值可能变得非常小，以至于数值不稳定性开始主导相关原子的贡献。第二个问题已被讨论为拉普拉斯算子谱近似中的一项局限 [1]。本文利用了这样一个事实： $\mathbb{CP}^1$ 的自同构群是莫比乌斯变换群， $PSL(2,\ \mathbb{C})$；我们可以利用该群的元素，确保计算所用的坐标始终处于数值受控的区域内（这里我们使用单位圆盘）。

更具体地说，令 $m \in \{1, 2, \ldots, N\}$ 为构成分子表面的 m^{th} 球面 的索引，则存在一个元素 $\mathcal{T}_m \in PSL(2,\ \mathbb{C})$，它将单位圆盘映射

$$\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\},$$

到公式（2）中的区域 $\mathbb{D}(a_m,\ R_m)$。注意，如果 m^{th} 球面 的层级为 l，那么与相交于 m^{th} 球面 的层级为（$l+1$）的球面所对应区域的原像，将描述某些完全包含于 $\mathbb{D}$ 中的圆盘。因此， m^{th} 球面 对公式（4）所述矩阵的贡献为

$$\iint_{\mathbb{D}-\hat{D}} (\mathcal{T}_m(w))^i (\overline{\mathcal{T}_m(w)})^j e^{-k\varphi(\mathcal{T}_m(w))} F(\mathcal{T}_m(w))\, d\mathcal{T}_m(w) \wedge d\overline{\mathcal{T}_m(w)}, \tag{5}$$

其中，$\hat{}D$ 表示与相交于 m^{th} 球面 的下一级球面相对应的圆盘的并集。实际计算中，我们通过在体积形式 $F(\mathcal{T}_m(w))\ d\mathcal{T}_m(w) \wedge d\overline{\mathcal{T}_m(w)}$ $\hat{}$中令 $w \in D$ 时赋值 0，来计入这些更高层级的球面（请注意，这会导致体积形式出现跳跃间断）。公式（5）这类积分的数值计算通过将积分拆分为角向和径向两部分，再依次应用梯形法则完成；我们选择的径向步长对应于 $n_r = 15$ 个积分点，角向步长则对应于取 $n_\theta = 10$个点。这似乎能达到合理的精度；例如，可以针对给定的积分方案检验面积积分。我们还发现，缩小步长似乎不会显著改变形状描述符之间的距离（第 3.1 节）。

2.5 求形状描述符之间的距离

给定两个正定厄米矩阵 $\mathbb{M}_1$， $\mathbb{M}_2$，例如由公式（4）生成的矩阵，定义此类矩阵之间距离的方式不胜枚举。就凯勒量子化理论而言，自然可以将 $\mathbb{M}_1$， $\mathbb{M}_2$ 视为固定复向量空间 $H^0(\mathcal{O}(2k))$ 上的两个厄米内积。该空间自然可视为流形 $GL(2k+1;\ \mathbb{C})/U(2k+1)$。内积由指定某个基为正交归一基来确定；在 $U(2k+1)$ 作用下共轭的任意基都定义相同的内积。该空间上存在一种自然的距离；这种距离的一种刻画是：最短路径（测地线）由 $GL(2k+1;\mathbb{C})$ 的单参数子群给出，也就是说，这些路径由形式为 $\exp(tA)$ 的矩阵构成，其中 A 是某个 $(2k+1) \times (2k+1)$ 复矩阵。

更明确地说，如果 $\{v_1, v_2, \ldots, v_{2k+1}\}$ 是 $H^0(\mathcal{O}(2k))$ 的一个基，使得两个内积都由对角矩阵表示

$$\mathbb{M}_1 = \mathrm{Diag}\left(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_{2k+1}}\right), \qquad \mathbb{M}_2 = \mathrm{Diag}\left(e^{\mu_1}, e^{\mu_2}, \ldots, e^{\mu_{2k+1}}\right),$$

then

$$d(\mathbb{M}_1, \mathbb{M}_2) = k^{-\frac{3}{2}} \sqrt{\sum_{i=1}^{2k+1} (\lambda_i - \mu_i)^2}. \tag{6}$$

$k^{-3/2}$ 的因子确保当 $k \to \infty$ 时距离趋于稳定（参见 [11] 中的定理 1.1）。考虑以下更紧凑的距离表达式将很有用

$$d(\mathbb{M}_1, \mathbb{M}_2) = k^{-\frac{3}{2}} \sqrt{\sum_{i=1}^{2k+1} (\log(\eta_i))^2}, \tag{7}$$

其中 $\{\eta_i\}$ 是矩阵 $\mathbb{M}_1^{-1}\mathbb{M}_2$的特征值。

众所周知，黎曼球面 $\mathbb{CP}^1$ 的自同构群是莫比乌斯变换群 $PSL(2,\ \mathbb{C})$。粗略地说，子群 $PSU(2) \subset PSL(2,\ \mathbb{C})$ 对应于原曲面的旋转，其余映射则对应于保持复结构不变的重参数化。如果$\varpi \in PSL(2,\ \mathbb{C})$ 是如下形式的自同构

$$\varpi(z) = \frac{\alpha z + \beta}{\gamma z + \delta},$$

那么 ϖ 也作用于向量空间 $H^0(\mathcal{O}(2k))$。用表示论的术语来说，这个作用是 $SL(2,\ \mathbb{C})$ 的标准表示在 $\mathrm{Sym}^{2k}(\mathbb{C}^2)$ 上诱导的表示。若将 $SL(2k+1,\ \mathbb{C})$ 中的元素记为 $\vartheta(\varpi)$（见 [12], 引理 8），并将以 z-坐标计算得到的原始形状描述符记为 $\mathbb{M}$，那么以 $\varpi(z)$-坐标计算得到的形状描述符将为

$$\left(\vartheta(\varpi)\right)^* \mathbb{M}\left(\vartheta(\varpi)\right).$$

如第 2 节所述，Kähler 势只定义到一个加法常数，这意味着我们也可以将厄米矩阵 $\mathbb{M}$ 乘以一个正数。因此，计算两个形状描述符 $\mathbb{M}_1$和 $\mathbb{M}_2$之间距离的问题，具体来说就是在 $(p,\ \vartheta) \in \mathbb{R} \times SL(2,\ \mathbb{C})$ 上求最小值，

$$\zeta(p, \vartheta) = \sum_{i=1}^{2k+1} (\log(\eta_i))^2,$$

其中，$\{\eta_i\}$ 是矩阵 $\mathbb{M}_1^{-1} e^p \left(\vartheta(\varpi)\right)^* \mathbb{M}_2(\vartheta(\varpi))$ 的特征值。

容易看出，在 ζ 的临界点处，p 的值与元素 ϑ无关。根据初等微积分，p 的值为

$$p = -\frac{1}{2k+1} \sum_{i=1}^{2k+1} \log(\tilde{\eta}_i),$$

其中 $\{\tilde{\eta}_i\}$ 是矩阵 $\mathbb{M}_1^{-1}\mathbb{M}_2$的特征值。由于矩阵 $(\vartheta(\varpi))$的行列式为 1，p的值不受厄米矩阵 $\mathbb{M}_2$上的 $SL(2,\mathbb{C})$ 作用影响。因此，我们将距离计算简化为在六维李群 $SL(2,\mathbb{C})$上的最小化问题。

请注意，方程 (6) 给出的形状描述符之间的距离，是分子形状在重新缩放至面积为 4π之后 的距离。因此，两个分子表面 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 之间的距离应包含一个分量，以反映 $\mathcal{S}_1$ 与 $\mathcal{S}_2$之间的面积差异。由于我们旨在得出相似度分数，而非计算两个输入之间的距离，因此本文不再进一步讨论这一点。我们在构建相似度分数方面的初步尝试将在下一节详细介绍。

对 $SL(2,\mathbb{C})$ 的剩余最小化通过将一般矩阵参数化为 6 个实变量 $x_1, ...x_6$来完成，并取

$$\varpi(x_1, x_2, \ldots, x_6) = \begin{pmatrix} x_1 + \sqrt{-1}x_2 & x_3 + \sqrt{-1}x_4 \\ x_5 + \sqrt{-1}x_6 & * \end{pmatrix},$$

其中，选择 $*$ 以确保 $\det(\varpi) = 1$。为了执行最小化，我们采用无需输入梯度向量的算法，例如 Nelder–Mead 或 Powell方法 [13]。这些算法通过 SciPy [14] 中现成的软件包实现。我们发现，对于 $k = 1$，两种方法得到的结果几乎没有差别；最小化算法能够收敛并给出稳健的距离值。对于 $k = 2$，最小化方法似乎不太稳定，偶尔无法收敛。一种解决办法是将 $k = 1$ 最小化得到的 $SL(2,\mathbb{C})$ 元素作为 $k = 2$ 步骤的初始猜测（否则使用单位矩阵）。我们预计这一过程还可以改进；例如，可以显式计算待最小化函数的梯度，然后在共轭梯度下降等算法中使用该梯度。

在实现两个矩阵之间的距离度量时，还需考虑形状描述符中 $k > 2$（某些情况下也包括 $k = 2$）存在数值不稳定性的问题。偶尔会产生非正定矩阵，无法使用上述方法进行比较。由于仅相差一个缩放比例的厄米矩阵可视为等价，因此在必要时，我们会将其中一个矩阵乘以 10、100 或 1000，以使特征值落入 Python 所需的考察范围内。

3 初始案例研究：磷酸二酯酶5（PDE5）抑制剂

3.1 调整参数 n_r 和 n_θ

为确定改变参数 n_r 和 n_θ （第 2.4 节）对所生成形状描述符质量的影响，我们考察了三组参数：$n_r = 200$ 和 $n_\theta = 100$；$n_r = 50$ 和 $n_\theta = 25$；$n_r = 15$ 和 $n_\theta = 10$。此处报告了 Sildenafil（表 1）、Vardenafil（表2）和 Tadalafil（表 3）各自描述符与计算相关积分时得到的面积之间的距离（对于准确的描述符，该距离应为 ~ 12.57；这是将曲面面积缩放至 4π 所施加的约束）。

表 1： 计算得到的距离 b 使用不同的 n_r 和 n_θ 值生成的西地那非描述符之间的距离，适用于 $k = 1$。报告的面积是由 th e 积分步骤。

	(200, 100)，面积 = 12.59	**(50, 25)，面积 = 12.62**	**(15, 10)，面积 = 12.62**
(200, 100)	-	0.032	0.032
(50, 25)	0.038	-	0.040
(15, 10)	0.038	0.040	-

表2： 使用不同的 n_r 和 n_θ 值为 $k = 1$生成的伐地那非描述符之间的距离。报告的面积为积分步骤返回的面积。

	(200, 100) 面积 = 12.57	**(50, 25)，面积 = 12.58**	**(15, 10)，面积 = 12.58**
(200, 100)	-	0.005	0.005
(50, 25)	0.005	-	0.004
(15, 10)	0.005	0.004	-

表 3： 计算得到的距离 b etween 使用不同的 n_r 和 n_θ 值为 $k = 1$ 生成的他达拉非描述符。报告的面积是由 e 积分步骤返回的。

	(200, 100)，面积 = 14.32	**(50, 25)，面积 = 14.37**	**(15, 10)，面积 = 14.37**
(200, 100)	-	0.003	0.003
(50, 25)	0.003	-	0.001
(15, 10)	0.003	0.001	-

在每种情况下，这些距离都很小，因此减少所考虑的点数不会显著降低质量。西地那非和伐地那非的面积也都接近 12.57，表明其描述符质量很高。不过，他达拉非的面积略微被高估，这是由于将五元环与另外两个环之间的结构单元替换为其他结构单元时出现了问题，而非 n_r 和 n_θ的选择所致。考察 $k = 2$时也观察到了类似结果。由于质量未受影响，最终描述符采用了 $n_r = 15$ 和 $n_\theta = 10$ 的最小参数，以提高计算速度。

3.2 构建相似度分数

为了便于以熟悉的方式比较分子，我们希望构建一个相似度分数，而不是简单地取两个矩阵之间的距离。在化学信息学中，该分数通常介于 0（不相似）和 1（完全相同）之间 [4]。为此，我们取距离的倒数，并通过两个表面积之比来考虑大小因素。方程 8 给出了两个分子表面 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 之间的相似度分数，

$$score(\mathcal{S}_1, \mathcal{S}_2) = x(A_{min}/A_{max}) + y\frac{1}{1 + d(\mathbb{M}_1, \mathbb{M}_2)}, \tag{8}$$

其中，A_{min} 是两个表面积中较小的一个，A_{max} 是较大的一个，以确保分数介于 0 和 1 之间。因此，我们需要选择一组合适的权重 x 和 y，使得 $x + y = 1$ 且 $x < 0.5$，以确保形状是相似度分数的主要贡献因素。

表 4 给出了 PDE5 抑制剂成对比较所得的相似度分数。正如预期，在这三种情况下，随着表面积项所占比重增加，相似度也随之提高。西地那非与伐地那非之间的相似度仅略有提高，而对于他达拉非，加入面积项的影响则更大。最终选定 $x = 0.3$ 和 $y = 0.7$ 的权重，以平衡表面积项的贡献，避免其影响超过形状项。选择 PDE5 抑制剂进行调参，是因为它们已知具有相似性；不过，要进行全面的虚拟筛选，可能还需要使用更大规模的样本集进一步优化这些参数。

表4： 表面积权重从 0 到 0.5 时，PDE5抑制剂的相似度分数。

x	y	西地那非-伐地那非	西地那非-他达拉非	伐地那非-他达拉非
0	1	0.884	0.286	0.275
0.1	0.9	0.892	0.340	0.328
0.2	0.8	0.900	0.394	0.380
0.3	0.7	0.908	0.449	0.432
0.4	0.6	0.916	0.503	0.485
0.5	0.5	0.924	0.557	0.537

对于全规模虚拟筛选，可能需要使用更多样例来调整这些参数。

3.3 探究三维构象异构体的变化

如前文所述 [1], ，使用三维形状描述符时，考虑分子可能采取的不同取向（称为构象异构体）十分重要。理论上，同一分子的不同构象异构体得分应在 $0.7 < score < 1$ 范围内，因为预期它们具有较高的自相似性（在化学信息学中，得分高于 0.7），同时仍能彼此区分。

与 RGMolSA 一样，我们使用 PDE5抑制剂的两组小型数据集（各含 10 个构象异构体），以探究 KQ-MolSA 如何看待不同的构象异构体。一组包含 10 个随机构象，预计其变化略大；另一组包含 10 个低能构象，预期其相似性较高。两组数据均使用 ETKDG算法 [15] 生成，并采用 MMFF94力场 [16], 进行能量优化，所有操作均通过 RDKit [17]实现。图 3 给出了每组的最小、最大和平均形状相似性，以及平均均方根偏差（根据原子位置比较构象异构体）。完整的均方根偏差和形状相似性比较见**补充数据**。

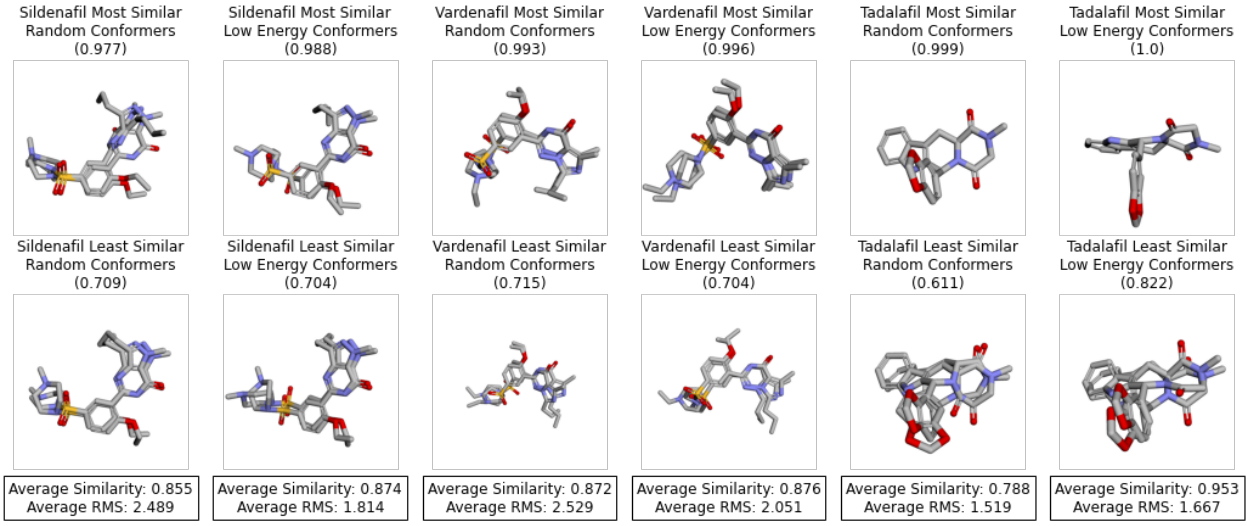
图4中的蜂群图比较了各组的均方根偏差和形状相似性。对于 $k = 1$，总体上观察到较高的相似度分数，但他达拉非的随机构象中有些分数略低于0.7。对于 $k = 2$，观察到的变化更大，其中一些构象对的分数低于0.6。对于 $k = 2$ ，相似性降低是意料之中的，因为这些描述符比 $k = 1$ 中的描述符更详细地近似了原始曲面，因此对几何差异更加敏感。不过，整体而言，得到的相似度分数低于RGMolSA；在RGMolSA中，大多数构象对之间的相似度都高于0.8 [1]。对于随机组，构象之间的相似度变化比RGMolSA更大；在RGMolSA中则观察到了相似构象的聚类。虽然KQMolSA能够较好地处理构象，但由于拉普拉斯–贝尔特拉米算子的谱对曲面形变不敏感，RGMolSA似乎处理得更好。在虚拟筛选中，将这些构象视为相似构象，就无需在计算形状相似性前进行预对齐；这或许还能帮助识别出能够发生形变以适应结合口袋的分子，并将其作为潜在命中化合物，而依赖原子坐标的方法则会将这些分子判定为形状不匹配。

3.4 与现有方法的比较

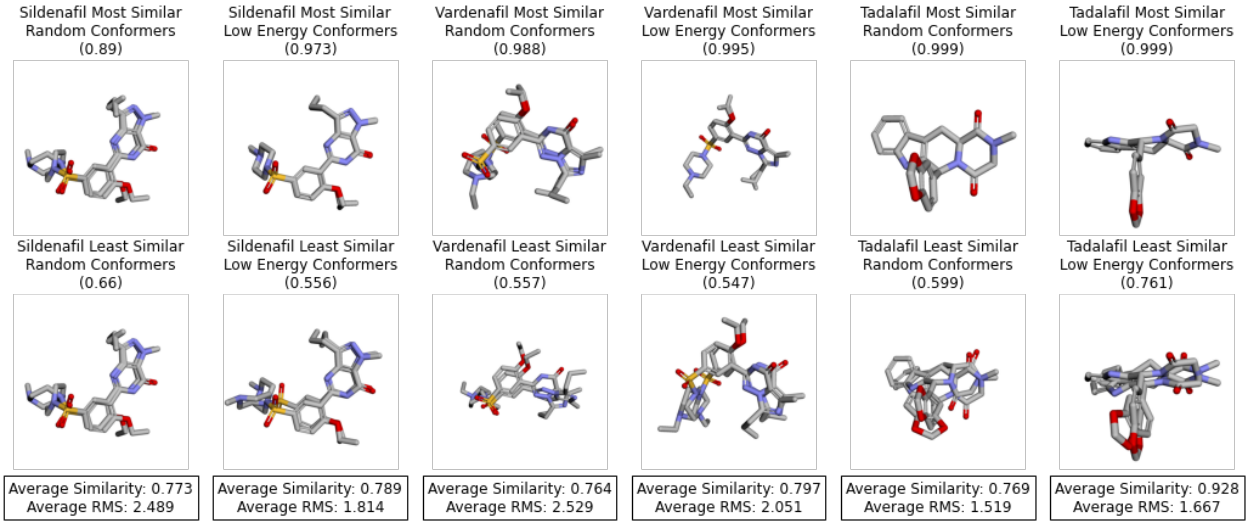
我们还使用 PDE5 抑制剂系列研究了 KQMolSA 与先前工作的比较效果，以及它与其他开源形状相似性方法的比较效果。表 5 列出了 PDE5抑制剂之间使用 KQMolSA（针对 $k = 1$ 和 $k = 2$）、RGMolSA [1], USRCAT [18, 17], Shape-It [19] 和MolSG [20]得到的形状相似性分数。此外，还列出了以 1024 位 Morgan 指纹（半径为 3）表示的二维表示。每种描述符的相似度分数范围均为 0（不同）至 1（相同）。

表 5： 将本文提出的方法（KQMolSA）与先前工作（RGMolSA）以及现有的基于原子间距离、以原子为中心和基于分子表面的描述符进行比较。 [1] [18], [19] [20] 所有情况下，相似度分数的范围均为 0（不相似）至 1（完全相同）。

	KQMolSA (k=1)	KQMolSA (k=2)	RGMolSA	USRCAT	Shape-It	MolSG	Morgan 指纹
西地那非-伐地那非	0.907	0.652	0.903	0.384	0.388	0.704	0.667
西地那非-他达拉非	0.449	0.482	0.809	0.269	0.278	0.746	0.201
伐地那非-他达拉非	0.432	0.470	0.725	0.291	0.353	0.887	0.209



(a) $k = 1$

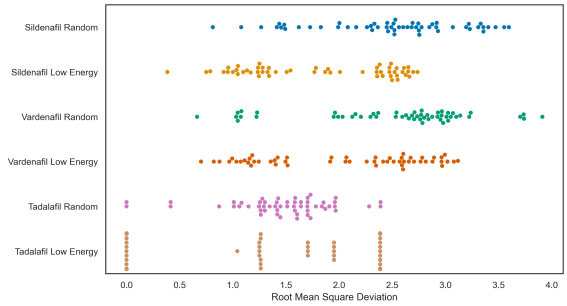


(b) $k = 2$

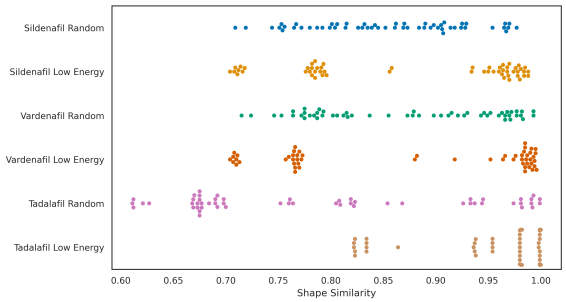
图 3： 西地那非、伐地那非和他达拉非中形状相似性最高和最低的构象异构体叠合图，以及每组的平均形状相似性和均方根偏差：(a) $k = 1$ 和(b) $k = 2$。尽管原子位置相似性存在差异，这些构象异构体平均而言仍表现出高度的自相似性。

正如本文前篇所述，西地那非和伐地那非是结构相近的类似物，因此应同时具有较高的形状相似性和指纹相似性。已知与其他抑制剂相比，他达拉非在 PDE5 中占据的体积相似，因此我们也预期其形状相似度分数较高，但二维相似性较低。为简便起见，每种分子仅考虑一种构象。

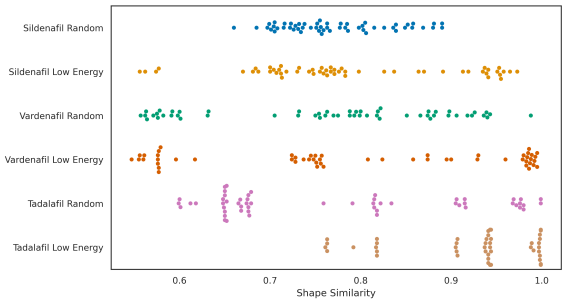
与 RGMolSA 一样，西地那非和伐地那非被评为高度相似，相似度分数为 0.907（$k = 1$）。然而，他达拉非的评分没有那么高；若采用 0.7 这一常见阈值，KQMolSA 会将其归类为不相似。对于 $k = 2$，观察到的相似度较低，这与前文所述的预期一致。 $k = 2$ 的相似度分数对比较顺序有轻微依赖（A 与 B 比较所得的分数，在小数点后第二位可能与 B 与 A 比较所得的分数不同，见表6）。这是因为距离计算涉及数值最小化过程，而非精确表达式；不过，这对化学信息学应用不会产生实际影响。在这项简单研究中，两种提出的方法（RGMolSA 和 KQMolSA）表现良好：与所有其他三维方法相比，它们对西地那非和伐地那非预测出的相似度更高；与 MolSG 相比，它们对相对相似度的排序也更符合直觉。不过，仍需开展全面的基准测试研究，以验证其性能。



(a)RMS相似性



(b)形状相似性 ($k = 1$)



(c)形状相似性 ($k = 2$)

图4： 我们这组构象异构体的均方根偏差（单位：埃）与形状相似性蜂群图显示了一个普遍趋势：尽管原子位置存在显著差异，不同构象异构体仍被归类为形状相似。均方根偏差小于 1 埃的构象异构体被认为相似，而大于 3 埃的则存在显著差异。

表6： k=2 时 PDE5抑制剂的相似度分数，突显比较顺序的影响。

	西地那非	**伐地那非**	**他达拉非**
西地那非	-	0.652	0.462
伐地那非	0.648	-	0.470
他达拉非	0.482	0.470	-

3.5 与势诱饵分子的相似性

与 RGMolSA 一样，我们也希望检验该方法如何处理那些应归类为与 PDE5抑制剂分子确实不同的分子。因此，我们将其与另外四种分子进行比较（图5）：精氨酸（补充剂），其分子量较低，但整体形状相似（由球面构成的长链）；赖甲环素（抗生素），其分子量较高，且含有四环基序，可能使分子的部分形状与西地那非相似；双氟拉松（外用皮质类固醇），其分子量相近，也含有四个环，但治疗靶点/适应症不同；以及 S-辛基谷胱甘肽（寡肽），其分子量同样相近，但不含环，且由于分子中心存在支链，可能具有一定相似性。

这项比较的结果如图 6 所示。对于 $k = 1$ 和 $k = 2$，大多数得分都明显低于相似度的典型阈值 0.7，因此这些分子会被归类为真正不同，且很可能对 PDE5 无活性。例外是他达拉非与双氟拉松之间的比较，其得分较高，达到 0.74（$k = 1$）。由于两者的结构相似（都含有由 4 个稠合环构成的基序），我们预期它们会有一定相似性。通过观察这两个分子的空间填充模型和曲面，也可以看出它们的形状确实相似（图 7）。RGMolSA 也将它们归类为可能相似（相似度为 0.872）。

4 结论

我们概述了一种全新形状描述符所依据的理论，

$$\mathbb{M}_{ij} = \iint_{\mathbb{C}} z^i \bar{z}^j e^{-k\varphi} F(z) \sqrt{-1} dz \wedge d\bar{z}, \tag{9}$$

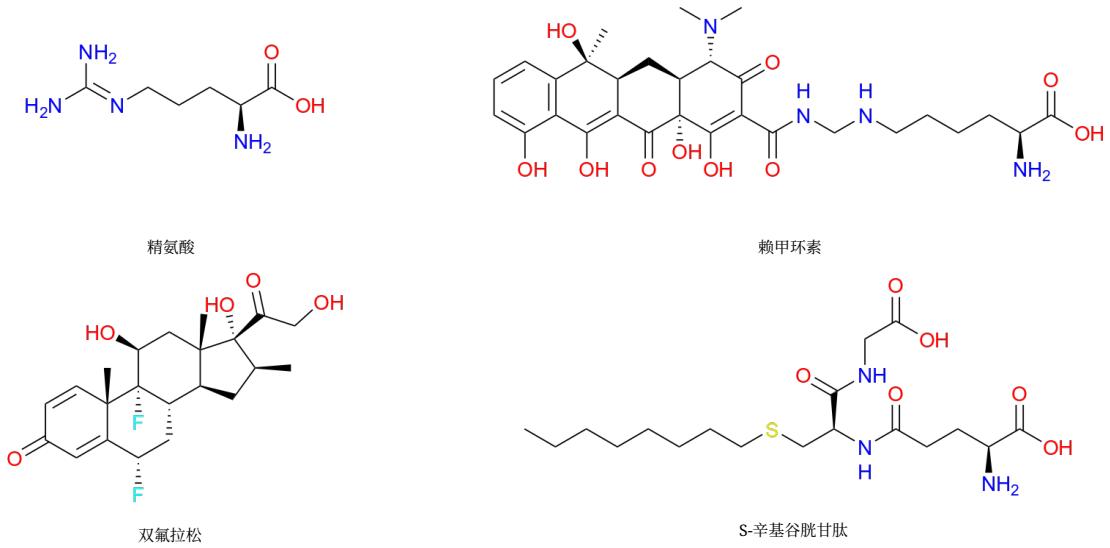


图 5： 潜在诱饵分子的化学结构。

即捕捉分子表面几何特征的 $(2k+1)\times(2k+1)$ 厄米矩阵。两个此类矩阵表示之间的距离为

$$d(\mathbb{M}_1,\mathbb{M}_2)=k^{-\frac{3}{2}}\sqrt{\sum_{i=1}^{2k+1}(\lambda_i-\mu_i)^2}. \tag{10}$$

随后得到一个总体相似度分数：相同分子的分数为 1，而完全不相似的分子分数为 0，具体如下

$$score(\mathcal{S}_1,\mathcal{S}_2)=0.3(A_{min}/A_{max})+0.7\frac{1}{1+d(\mathbb{M}_1,\mathbb{M}_2)}. \tag{11}$$

与此前报道的研究一样，通过一系列已知形状相似的 PDE5抑制剂对 KQMolSA 的能力进行了考察。该方法通常能够较好地处理构象异构体，相似度分数通常高于 0。7。由于细节更加丰富，对几何变化的敏感性更高，因此 $k=1$ 的分数高于 $k=2$，这符合预期。曲面对形变不敏感，使得 RGMolSA 在这一方面的表现优于 KQMolSA。与现有方法相比，KQMolSA 的表现相对较好，能够识别出西地那非和伐地那非具有高度相似性，但对他达拉非给出的相似度分数较低。这项小规模研究表明，RGMolSA 的表现可能仍然更好，但需要开展完整的回顾性基准测试研究加以确认。与 RGMolSA 相比，KQMolSA 的优势在于对基球选择的依赖性较低。因此，尽管 KQMolSA 的表现看似较差，在某些情况下使用它可能更为合适；例如，在考察环较少的长链分子时，RGMolSA 往往会出现数值误差。与一组潜在诱饵分子进行比较后，除他达拉非与双氟拉松的比较外，所有比较均得到较低分数；RGMolSA 也将这两者归类为相似分子。通过目测分子的空间填充模型和曲面模型可以看出，这一判定是合理的，因为它们的形状看起来相似。识别出这种相似性，证明了这些方法进行骨架跃迁的潜力。

尽管上述测试表明，矩阵 $\mathbb{M}$ 确实能够很好地描述分子形状，但该方法也存在一些缺点，主要在于计算两个描述符之间的距离。虽然两个厄米内积（分别由矩阵 $\mathbb{M}_1$ 和 $\mathbb{M}_2$表示）之间的距离概念已得到充分理解，但要计算两个分子表面之间的距离，就需要对 $SL(2,\ \mathbb{C})$-轨道上一点的距离进行最小化。尽管采用了现有的优化最小化算法，这一过程仍然相当缓慢，具体取决于所需最小化的程度，而且无法保证找到全局最小值。与 RGMolSA 几乎瞬时完成计算相比，这一步通常每对需要耗费几秒钟。若要将该方法用于药物发现流程中的超大型化学库筛选，还需进一步改进这一步骤。

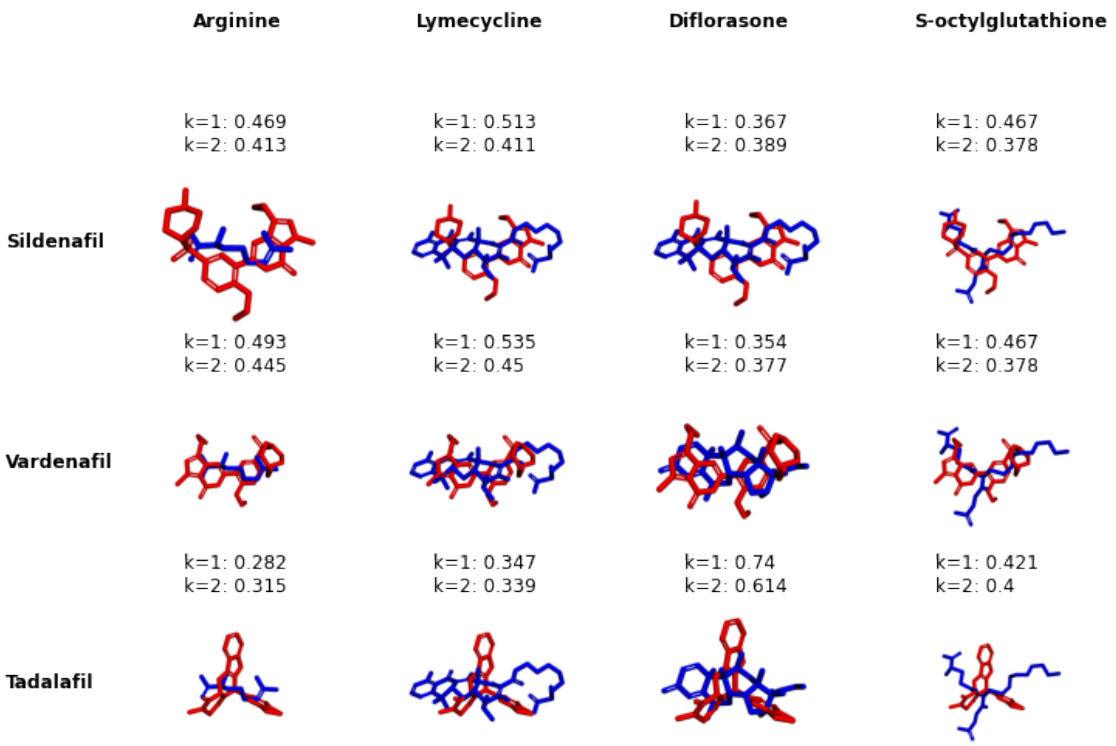


图 6： KQMolSA 对四种“不同”分子（蓝色）与 PDE5 抑制剂测试系列（红色）的相似度（针对 $k = 1$ 和 $k = 2$）。结构叠合是使用 Open3DAlign [21] 计算的。

当然，测量两个厄米矩阵之间距离的方法还有很多。人们或许可以期待，利用适当的数据集训练某种机器学习模型，能够发现描述符空间上的其他有用几何结构。

该方法在 $k = 2$ 以上也存在数值不稳定性（少数情况下在 $k = 2$ 也存在），会产生非正定的厄米矩阵。由于仅相差一个比例因子的厄米矩阵可视为等价，我们通过将一个矩阵乘以 10–1000 的因子进行缩放，使特征值落入 Python 数值容差范围内，以处理此类情况。

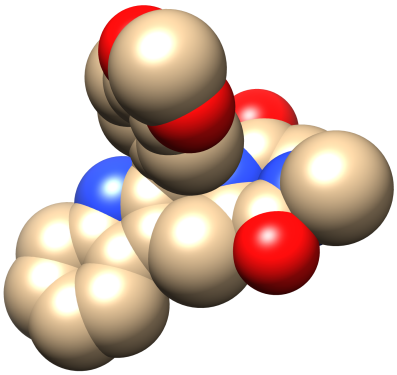
除了解决这些问题外，还可以通过考虑药效团特征进一步改进这两种方法，例如在考虑形状的同时纳入芳香环、氢键供体和受体。由于这些特征对结合至关重要，这样做可能比单独考虑形状带来更准确的预测。对于 RGMolSA，还可以进一步研究将 KQMolSA 生成的厄米矩阵描述符用作机器学习中的特征描述符。

5 附录：求凯勒势

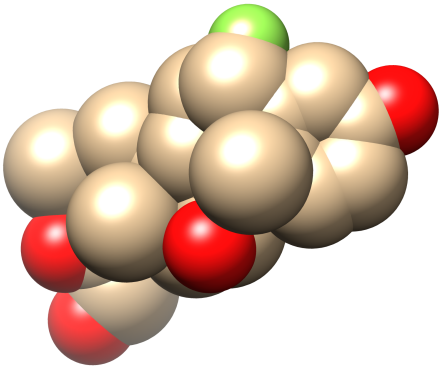
在给出 凯勒势的形式证明之前，我们先处理一个小小的技术问题。从 f
描述……的视角。 凯勒形式 ω 的角度来看

$$\sqrt{-1}\partial\bar{\partial}\varphi = \omega,$$

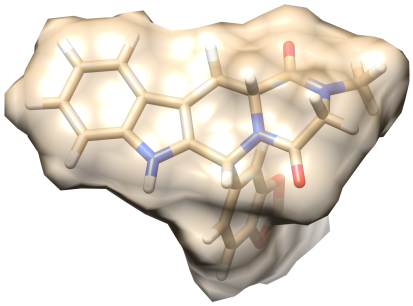
凯勒势 φ 仅在局部有定义，而添加任何满足 $\sqrt{-1}\partial\bar{\partial}H = 0$ 的函数 H
也将为 ω定义一个凯勒势。在底层复流形为 $\mathbb{CP}^1$ 且我们使用标准坐标 z的设定下，可以加上任意调和函数 H：$\mathbb{C} \to \mathbb{R}$，从而得到一个有效的凯勒势。



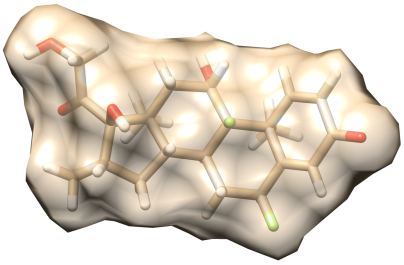
(a)他达拉非 - 空间填充模型



(b)双氟拉松 - 空间填充模型



(c) 他达拉非 - 曲面



(d)双氟拉松 - 曲面

图 7: 通过目视比较他达拉非和双氟拉松的空间填充模型与曲面，突显出二者的相似性 .

然而，在凯勒量子化中，势 φ 实际上描述的是一个全局对象，即线丛 L上的厄米度量 h。这意味着这些函数

$$h(z^j, z^j) = e^{-k\varphi(z)}|z|^{2j},$$

定义在整个球面 $\mathbb{CP}^1$ 上。具体而言，它们可以延拓为定义在无穷远点处的函数。例如，圆度量的凯勒势为 $\varphi = -2\log(|z|^2+1)$，因此，若加上一个调和函数 H，则要求

$$\frac{|z|^{4k}}{(1+|z|^2)^{2k}}e^{-kH}$$

有界。于是刘维尔定理推出 H 必为常数。

定理 5.1 (凯勒势的形式)。 设 ω 是由方程 *(3)* 给出的凯勒度量。若将与 i^{th} 球面对应的区域记为 $R_i \subset \mathbb{C}$, 则满足 $\sqrt{-1}\partial\partial\varphi = \omega$ 的凯勒势 φ 具有如下形式

$$\varphi(z) = \frac{C_i}{B_i}\log(|z-A_i|^2+B_i) + \sum_{j=1}^{N}\mathcal{K}_{ij}\log(|\alpha_{ij}z+\beta_{ij}|^2),$$

其中 $\mathcal{K} \in M^{N\times N}(\mathbb{R})$，且 α、$\beta \in M^{N\times N}(\mathbb{C})$。

证明。 证明采用归纳法，针对球面的数量 N进行。对于 $N=1$，度量 ω 是圆度量，因此我们可以取 $\mathcal{K}=0$。在曲面上添加一个新的球面，会使度量增加一个新的区域 R_k，该区域是一个圆盘，在其中度量的形式为

$$\omega(z)|_{R_k} = \frac{C_k}{(|z-A_k|^2+B_k)^2}\sqrt{-1}dz\wedge d\bar{z}.$$

我们可以通过莫比乌斯变换 $\mathcal{M}$ 将 R_k 映射到以原点为中心的单位圆盘，使得在单位圆盘的坐标下，度量表示为

$$\widetilde{\omega}(w)=\begin{cases} F(w)\sqrt{-1}dw\wedge d\overline{w} & \text{if} \quad |w|>1,\\ \dfrac{\kappa}{(|w|^2+\varepsilon)^2}\sqrt{-1}dw\wedge d\overline{w} & \text{if} \quad |w|\leq 1,\end{cases}$$

对于某个函数 $F\colon\ \mathbb{C}\to\mathbb{R}$ 以及常数 $\kappa,\ \varepsilon\in\mathbb{R}$。

我们使用 Dolbeault 方法求解 $\bar{\partial}$ 方程；对于一个紧支撑的[1] 连续函数 $H\colon\ \mathbb{C}\to\mathbb{C}$,

$$\psi(w)=\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{C}}\frac{H(p)}{p-w}dp\wedge d\overline{p},$$

满足 $\overline{\partial}\psi=H(w)d\overline{w}$。我们根据度量的形式将积分拆分，并考虑

$$\psi(w)=\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{D}}\frac{\kappa}{(|p|^2+\varepsilon)^2(p-w)}dp\wedge d\overline{p}+\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{C}\setminus\mathbb{D}}\frac{F(p)}{p-w}dp\wedge d\overline{p}.$$

为了计算第一个积分，我们使用柯西–庞培积分公式以及以下事实:

$$\frac{\kappa}{(|p|^2+\varepsilon)^2}=\frac{\partial}{\partial\overline{p}}\left(\frac{(\kappa/\varepsilon)\overline{p}}{(|p|^2+\varepsilon)}\right),$$

从而得到

$$\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{D}}\frac{\kappa}{(|p|^2+\varepsilon)^2(p-w)}dp\wedge d\overline{p}=$$
$$\begin{cases}\left(\dfrac{(\kappa/\varepsilon)\overline{w}}{(|w|^2+\varepsilon)}\right)-\dfrac{1}{2\pi\sqrt{-1}}\displaystyle\int_{\partial\mathbb{D}}\frac{(\kappa/\varepsilon)\overline{p}}{(|p|^2+\varepsilon)(p-w)}dp & \text{if} \quad |w|<1,\\ -\dfrac{1}{2\pi\sqrt{-1}}\displaystyle\int_{\partial\mathbb{D}}\frac{(\kappa/\varepsilon)\overline{p}}{(|p|^2+\varepsilon)(p-w)}dp & \text{if} \quad |w|>1.\end{cases}$$

围道积分

$$\frac{1}{2\pi\sqrt{-1}}\int_{\partial\mathbb{D}}\frac{(\kappa/\varepsilon)\overline{p}}{(|p|^2+B)(p-w)}dp,$$

可利用柯西留数公式轻松计算，得到

$$\frac{1}{2\pi\sqrt{-1}}\int_{\partial\mathbb{D}}\frac{(\kappa/\varepsilon)\overline{p}}{(|p|^2+\varepsilon)(p-w)}dp=\begin{cases}0 & \text{if } |w|<1,\\ -\frac{(\kappa/\varepsilon)}{(1+\varepsilon)w} & \text{if } |w|>1.\end{cases}$$

最后，我们得到

$$\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{D}}\frac{\kappa}{(|p|^2+\varepsilon)^2(p-w)}dp\wedge d\overline{p}=\begin{cases}\left(\frac{(\kappa/\varepsilon)\overline{w}}{|w|^2+\varepsilon}\right) & \text{if } |w|<1,\\ \frac{(\kappa/\varepsilon)}{(1+\varepsilon)w} & \text{if } |w|>1.\end{cases}$$

为了计算第二个积分，我们再次将定义域拆分，并考虑

$$\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{C}\setminus\mathbb{D}}\frac{F(p)}{p-w}dp\wedge d\overline{p}=\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{C}}\frac{F(p)}{p-w}dp\wedge d\overline{p}-\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{D}}\frac{F(p)}{p-w}dp\wedge d\overline{p}.$$

积分

$$S(w)=\frac{1}{2\pi\sqrt{-1}}\iint_{\mathbb{C}}\frac{F(p)}{p-w}dp\wedge d\overline{p},$$

是.....的解

$$\frac{\partial S}{\partial\overline{w}}=F(w).$$

[1] Our 函数不是紧支撑的，但我们可以在任意半径处截断，从而得到这样的函数 n.

在单位圆盘 $\mathbb{D}$ 中， F 的形式为

$$F(w) = \frac{\tilde{\kappa}}{(|w|^2 + \tilde{\varepsilon})^2},$$

其中 $\tilde{\kappa}$ 和 $\tilde{\varepsilon}$ 是正常数。因此

$$\psi(w) = \begin{cases} S(w) + \left(\dfrac{(\kappa/\varepsilon)}{|w|^2 + \varepsilon} - \dfrac{(\tilde{\kappa}/\tilde{\varepsilon})}{|w|^2 + \tilde{\varepsilon}} \right) \overline{w} & \text{if} \quad |w| < 1, \\ S(w) + \left(\dfrac{(\kappa/\varepsilon)}{1 + \varepsilon} - \dfrac{(\tilde{\kappa}/\tilde{\varepsilon})}{1 + \tilde{\varepsilon}} \right) w^{-1} & \text{if} \quad |w| > 1, \end{cases}$$

是 $dw \wedge \partial\psi = \widetilde{\omega}(w)$ 的解。

如果 $Q(w)$ 是 $F(w)\sqrt{-1}dw \wedge dw$ 的凯勒势，那么

$$\widetilde{\varphi}(w) = \begin{cases} Q(w) + (\kappa/\varepsilon)\log(|w|^2 + \varepsilon) - (\tilde{\kappa}/\tilde{\varepsilon})\log(|w|^2 + \tilde{\varepsilon}) - K & \text{if} \quad |w| < 1, \\ Q(w) + \left(\dfrac{(\kappa/\varepsilon)}{1 + \varepsilon} - \dfrac{(\tilde{\kappa}/\tilde{\varepsilon})}{1 + \tilde{\varepsilon}} \right) \log(|w|^2) & \text{if} \quad |w| > 1, \end{cases}$$

其中

$$K = (\kappa/\varepsilon)\log(1 + \varepsilon) - (\tilde{\kappa}/\tilde{\varepsilon})\log(1 + \tilde{\varepsilon}),$$

是 $\widetilde{\omega}$ 的凯勒势。通过莫比乌斯变换将函数 $\widetilde{\varphi}$ 拉回

$$\mathcal{M}(z) = \frac{\alpha z + \beta}{\gamma z + \delta}$$

我们得到

$$\varphi_k(z) = \begin{cases} Q\left(\dfrac{\alpha z + \beta}{\gamma z + \delta} \right) + (\kappa/\varepsilon)\log\left(\left| \dfrac{\alpha z + \beta}{\gamma z + \delta} \right|^2 + \varepsilon \right) - K & \text{if} \quad z \in R_k \\ Q\left(\dfrac{\alpha z + \beta}{\gamma z + \delta} \right) + \left(\dfrac{(\kappa/\varepsilon)}{1 + \varepsilon} - \dfrac{(\tilde{\kappa}/\tilde{\varepsilon})}{1 + \tilde{\varepsilon}} \right) \log\left(\left| \dfrac{\alpha z + \beta}{\gamma z + \delta} \right|^2 \right) & \text{if} \quad z \notin R_k \end{cases}$$

是度量的凯勒势，该度量在点 $z = -\delta/\gamma$ 处是奇异的。我们可以将 Q 项替换为前一个 φ 所对应的适当函数，然后加上适当倍数的 $\log(|\gamma z + \delta|^2)$，从而得到具有适当形式的凯勒势。

□

6 致谢

作者感谢 EPSRC 博士培训合作项目提供的博士生奖学金（资助编号 EP/R51309X/1）、艾伦·图灵研究所的 Enrichment Scheme（R.P.），以及 UKRI 未来领袖奖学金（资助编号 MR/T019654/1）（D.J.C.）。S.J.H. 感谢 R. L. Hall 博士对本项目的关注以及围绕项目进行的有益交流。T.M. 感谢加利福尼亚大学欧文分校在本文部分工作完成期间给予的热情接待。

参考文献

[1] Daniel J. Cole、Stuart J. Hall 和 Rachael Pirie。黎曼几何与分子表面 I：拉普拉斯算子的谱，（预印本），2022年。[2] Mark A. Johnson 和 Gerald M. Maggiora。分子相似性的概念与应用。1990年。[3] Sumudu P. Leelananda 和 Steffen Lindert。药物发现中的计算方法。*Beilstein J. Org. Chem.*，12：2694–2718，2016年。[4] Ashutosh Kumar 和 Kam Y. J. Zhang。形状相似性方法的发展及其在药物发现中的应用进展。*Front. Chem.*，6：1–21，2018年。[5] S. K. Donaldson。复微分几何中的一些数值结果。*Pure Appl.Math. Q.*，5(2)：571–618，2009年。

[6] Ann E. Cleves 和 Ajay N. Jain。归纳偏置对基于配体的建模计算评估及药物发现的影响。*J. Comput. Aided Mol.Des.*，22(3)：147–159，2008年。[7] Stuart J. Hall 和 Thomas Murphy。分子表面的凯勒几何，准备中。[8] Phillip Griffiths 和 Joseph Harris。《代数几何原理》。纯粹与应用数学。Wiley-Interscience 出版物。纽约等地：John Wiley & Sons。XII，813页。£ 29.60；$ 58.00（1978）。，1978 年。[9] Gang Tian。代数流形上的一组极化凯勒度量。*J. Differ. Geom.*，32(1)：99–130，1990 年。[10] Robert Berman 和 Julien Keller。伯格曼测地线。载于凯勒度量空间中的复蒙日-安培方程与测地线。讲义，第283–302页。柏林：施普林格，2012年。[11] Xiuxiong Chen 和 Song Sun。凯勒度量空间（V）——凯勒量子化。载于度量与微分几何，第297卷数学进展，第19–41页。Birkhäuser/施普林格，巴塞尔，2012年。[12] Yoshinori Hashimoto。极值凯勒度量的量子化。*J. Geom.Anal.*，31(3)：2970–3028，2021年。[13] William H. Press、Saul A. Teukolsky、William T. Vetterling 和 Brian P. Flannery。数值方法。剑桥大学出版社，剑桥，第三版，2007年。科学计算的艺术。[14] Pauli Virtanen 等。SciPy 1.0：Python科学计算的基础算法。《自然·方法》，17：261–272，2020年。[15] Sereina Riniker 和 Gregory A. Landrum。更可靠的距离几何：利用已知信息改进构象生成。*J. Chem. Inf. Model*，55(12)：2562–2574，2015年。[16] Paolo Tosco、Nikolaus Stiefl 和 Gregory Landrum。将MMFF力场引入RDKit：实现与验证。*J.Cheminformatics*，6(1)，2014年。[17] Greg Landrum。RDKit：开源化学信息学软件。版本2021.09.1。[18] Adrian M Schreyer 和 Tom Blundell。USRCAT：利用药效团约束实现实时超快速形状识别。*J.Cheminform.*，4：1489–1495，2012年。[19] Jonatan Taminau、Gert Thijs 和 Hans De Winter。Pharao：药效团对齐与优化。*J. Mol.Graph*，27(2)：161–169，2008年。[20] Matthew P. Seddon、David A. Cosgrove、Martin J. Packer 和 Valerie J. Gillet。基于谱几何的无对齐分子形状比较：框架。*J. Chem. Inf. Model*，59：98–116，2019年。[21] Paolo Tosco、Thomas Balle 和 Fereshteh Shiri。Open3DAlign：旨在实现无监督配体对齐的开源软件。*Journal of Computer-Aided Molecular Design*，25(8)：777–783，2011年。