arXiv:2611.29902v1 [physics.chem-ph] 23 Nov 2026

借助人工智能代理实现自主多参考量子化学

Younghwan Kim, †,¶ Ming-Hsiu Hsieh, †,¶ Ankit Mathanker, † Arpan Choudhury, † Joshua Zhou, ‡ Qiaohong Wang, ‡ MatthewR. Hermes, † 以及 Laura Gagliardi*,†,‡

†芝加哥大学化学系，美国伊利诺伊州芝加哥，邮编 *60637*

‡芝加哥大学普利兹克分子工程学院，美国伊利诺伊州芝加哥，邮编 *60637*，美国

¶对本研究贡献相同

电子邮件：lgagliardi@uchicago.edu

摘要

准确描述涉及强关联电子的化学反应，仍是计算化学中的一项重大挑战。多参考电子结构方法可以应对这一挑战，但其应用需要依赖专家判断，而这些判断难以自动化；当化学键形成和断裂导致电子结构发生变化时，尤其如此。本文介绍 CASPER，这是一种自主智能体工作流，可沿化学反应路径执行多参考计算。CASPER 将化学知识、电子结构计算和迭代评估相结合，以确定具有化学意义的活性空间，并在化学键形成和断裂过程中保持其一致性。我们在五种过渡金属反应中展示了 CASPER 的能力；对于这些反应，CASPER 几乎无需人工干预，便能生成一致的多参考势能剖面。值得注意的是，在其中两种反应中，CASPER 识别出了关键的电子结构问题，而独立运行的 Claude 会话若没有结构化工作流，则无法发现这些问题：

其中包括一个电子态交叉，以及遗漏了一个参与化学键断裂的轨道。我们的结果表明，将语言模型与化学知识以及对计算结果的独立评估相结合十分重要。此类智能体不仅能够自动执行计算，还有望帮助研究人员更好地理解其中的物理学和化学原理，并为研究复杂的化学反应性、规模化生成高质量量子化学数据开辟新的机会。

1 引言

要理解化学键的形成与断裂，需要准确描述电子结构。这在过渡金属配合物中尤具挑战性，因为多个电子组态都可能对波函数有所贡献，而且它们的相对重要性沿反应路径可能发生显著变化。[1,2]

多参考电子结构方法为描述这些强关联体系提供了强有力的框架，[3–6] 但其应用仍远未成为常规操作。主要挑战之一是选择合适的活性空间，即在多组态波函数中显式处理的电子和轨道子集。这一选择需要丰富的专业知识，而且往往需要反复手动干预。[7] 对于化学反应，这一问题尤其棘手，因为活性空间必须在化学键形成和断裂的过程中，对不断演变的电子结构作出平衡描述。

在自动化活性空间选择方面已取得显著进展。已有多种基于轨道纠缠的方法，例如autoCAS，[8,9] 近似关联度量，[10–12]机器学习，[13–15] 以及原子轨道投影，[16,17] 其中包括原子价层活性空间（AVAS）[18] 和 π-轨道体系（PiOS）[19]，。其他方法，包括加权活性空间协议（WASP），[20] 轨道映射策略，[21] 以及自适应轨道选择方法，[22] 则着眼于确保不同分子几何结构下活性空间的一致性。近来，大语言模型（LLM）智能体也开始用于

辅助电子结构计算和活性空间选择。[23–25]

然而，在单一几何构型下确定活性空间只是问题的一部分。在化学反应过程中，轨道的性质会发生变化，电子态可能变得近简并，而具有化学重要性的轨道可能会在完全活性空间自洽场（CASSCF）优化过程中被替换。因此，计算可能会收敛并生成平滑的能量曲线，却无法对反应提供适当的化学描述。识别这些问题并决定如何加以修正，需要的不只是计算自动化，还需要化学知识、电子结构分析能力，以及修订计算决策的能力。

本文介绍 CASPER（完全活性空间规划、执行与推理），这是一个用于沿化学反应路径开展自主多参考计算的人工智能智能体框架。CASPER 将化学知识与电子结构计算相结合，形成结构化工作流，用于选择活性空间、评估其化学性质与一致性、识别潜在问题，并在必要时修改计算。我们在五个涉及不同成键与断键过程的过渡金属反应中展示了 CASPER 的应用。在全部五个案例中，CASPER 都识别出一致的活性空间，并以极少的人工干预生成平滑的势能曲线。

本研究的一项重要成果，是比较 CASPER 工作流与独立的 Claude 会话。在所考察的两种反应中，受控的 Claude 会话能够执行计算，并得到看似一致的活性空间。然而，它未能识别其中一种反应的电子态交叉，并在另一种反应中排除了一个参与化学键断裂的轨道。CASPER 识别出了这些问题，并据此调整了计算。尽管这项比较仅涉及两种反应，但它表明，仅凭收敛和轨道一致性，并不足以确立多参考计算的化学有效性。这也说明，采用包含专门化学评估的结构化工作流，能够提升大语言模型所得结果的质量。

除了自动化电子结构计算之外，我们还设想，CASPER 这类智能体能够帮助科学家更深入地理解他们所研究系统的物理学和化学。通过分析电子结构、识别潜在问题并提出解决方法，这些智能体可以补充人类的专业知识，减少计算中技术环节所耗费的时间。这样，研究人员便能将更多精力用于解读结果和探究重要的科学问题。CASPER 是迈向让多参考计算更易于使用，并使其更有助于研究复杂化学反应性的初步一步。

2 用于自主多参考计算的智能体框架

我们开发了 CASPER，这是一种智能体框架，整合了化学推理、电子结构计算和迭代评估，能够沿化学反应路径自动执行多参考计算。核心挑战不仅在于为给定的分子几何构型确定合适的活性空间，还在于确保选定轨道始终具有化学相关性，并能在整个反应过程中一致地描述不断演变的电子结构。CASPER 将自动化计算程序与专门委员会成员相结合，由后者评估计算的不同方面，并在必要时提出纠正操作，以应对这一挑战。

该框架分为五个阶段：(1) 活性空间选择，(2) 过渡态（TS）计算（过渡态运行），(3) 过渡态评估，(4) 反应路径计算（路径运行），以及 (5) 路径审查（图1）。工作流从一个参考几何构型（通常为 TS）开始，在此构造基于化学原理的候选活性空间，并执行态特定完全活性空间自洽场（SS-CASSCF）计算并进行评估。确定合适的活性空间后，便沿反应坐标向两个方向推进计算。在整个过程中，CASPER 会监测

活性轨道的化学性质、优化活性空间的一致性以及电子态性质的变化。检测到不一致时，该框架可以尝试不同的初始猜测、调整轨道选择，或在继续计算前细化反应路径。只有当最终计算结果满足规定的一致性标准并通过最终的独立评估后，工作流才会结束。

科学决策由**协调器** 统筹，协调器会咨询七位专门委员会成员（图1）。**化学家** 负责解读反应化学，并识别与成键、键断裂和金属–配体相互作用相关的轨道。**构建者**负责构建计算输入、执行计算、监测收敛情况，并在计算失败时提出补救措施。**一致性审查员**负责评估优化后的活性轨道是否保留其预期的化学性质，以及在不同分子几何结构中是否保持一致。**态平均（SA）监测员** 负责监测电子态，并识别需要进行态平均计算的情况。**路径优化器** 负责判断不一致是否源于几何分辨率不足，并在必要时建议增加中间结构。**Librarian**负责查阅相关文献，以识别可供比较的电子结构和活性空间；而**审稿人** 则独立评估所提选择和计算结果的科学有效性。

每位委员会成员都依据一套明确的指令和评估标准开展工作，并向**协调器**提交结构化建议。成员之间不会直接交流，因此可以分别考量各自的评估。当建议相互冲突时，**协调器**会评估现有的计算证据，并在必要时启动额外计算以解决分歧。大多数成员都是**协调器**在自身上下文中依次扮演的角色，而 **Librarian** 和**审稿人**则会在最终评估阶段始终以全新且相互隔离的上下文作为子代理调用，以便独立审查结果。这种组织方式将计算决策的生成与批判性评估分开，减少了对任何单一成员评估的依赖。

，减少了对任何单一成员评估的依赖。

CASPER 的一个重要特点是，电子结构计算的收敛本身并不意味着成功。收敛的 CASSCF 计算中，轨道的化学性质可能已经发生变化，或者在轨道优化过程中被替换，从而可能损害反应剖面的一致性。因此，CASPER 会通过定量的轨道重叠度量、自然轨道占据数，以及结合化学知识对轨道特征进行的分析，评估优化后的活性空间。这些评估结果决定了是否可以接受某项计算，或是否需要进一步优化。该框架还会监测电子态性质的变化，并在适当时引入态平均。通过将这些诊断与迭代式决策相结合，CASPER 能够识别并纠正原本需要专家介入的问题。

该工作流受明确的计算约束和升级规则支配。只有在初始设定含糊不清、重复失败无法依据既定规则解决，或拟采取的操作超出允许的计算范围时，才会请求人工输入。所有决策、支持证据和纠正操作都会被记录，从而可追溯地说明最终活性空间是如何确定的。专业科学推理、独立评估与计算反馈相结合，为自主开展化学反应性的多参考计算奠定了基础。

2.1 输入与准备

CASPER 需要一系列描述反应路径的分子几何结构，以及基组、分子电荷、自旋多重度和应开始计算的几何构型。在本研究中，我们使用从参考密度泛函理论（DFT）计算中获得的过渡态（TS）几何结构作为起始点。尽管用户可以选择任意几何构型，我们将起始几何构型称为

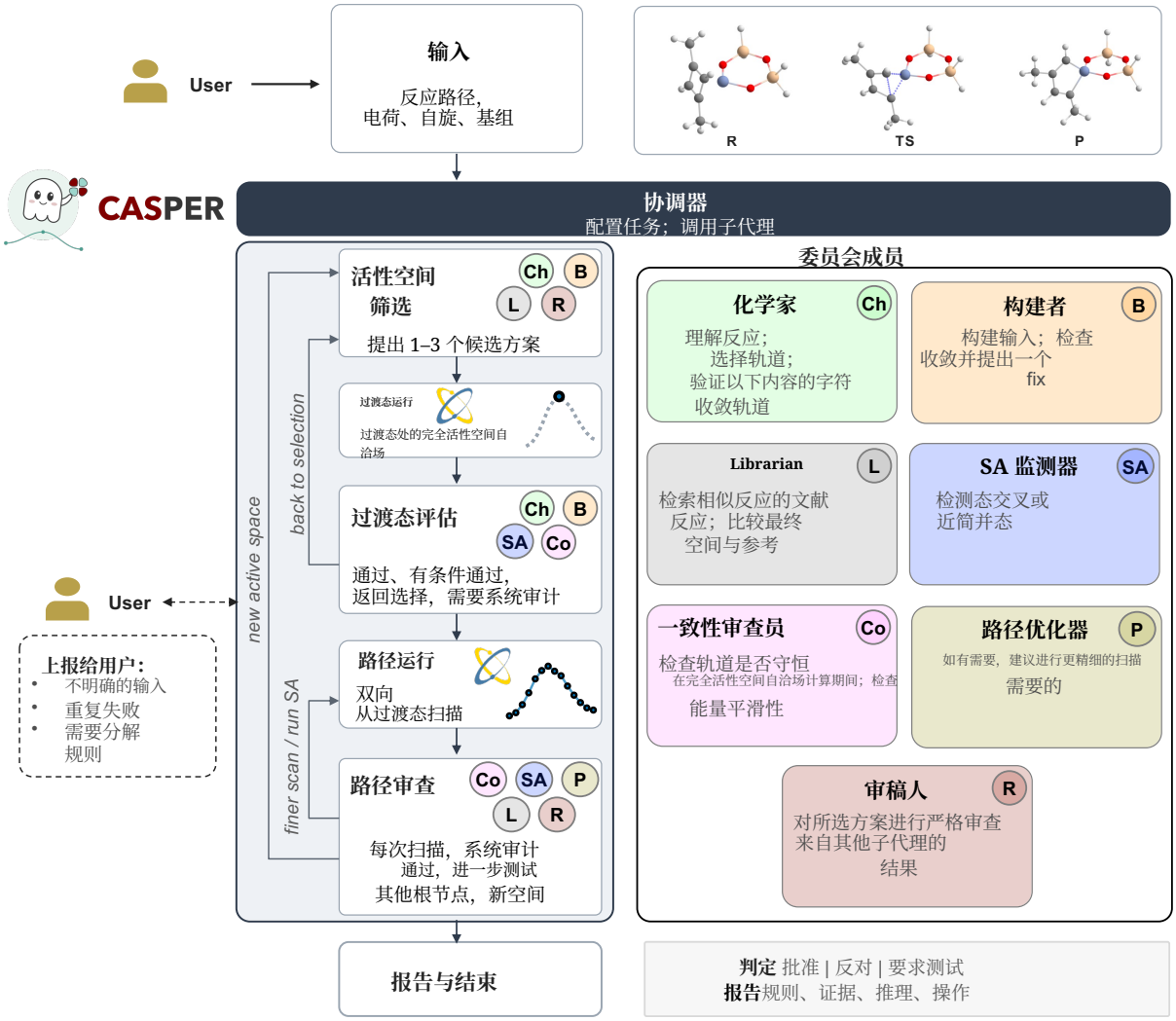


图1：用于反应路径活性空间选择的智能体工作流概览。该工作流需要将反应路径、系统自旋、电荷和基组作为输入。主智能体会话（**协调器**）将一条反应路径依次经过五个阶段：活性空间选择、过渡态（TS）计算、过渡态评估、路径运行、路径审查，随后进行报告生成。在每个阶段，工作流都可以一次调用一名委员会成员，即：**化学家**、**构建者**、**Librarian**、**审稿人**、**一致性审查员**、**路径优化器**和**SA 监测器**。**协调器**向委员会相关成员说明任务，并接收结构化评估意见。**Librarian** 和**审稿人**作为全新的子代理运行，无法看到其他代理的推理过程。只有当委员会无法继续推进时，才会咨询用户。

贯穿稿件的 TS。

首先，沿反应路径对分子几何结构进行对齐。随后在 TS 处进行限制性 Hartree–Fock 或限制性开壳层 Hartree–Fock 计算，

根据自旋态的不同，占据轨道使用内禀键轨道（IBO）进行局域化[26]，价层虚轨道则使用局域化内禀价层虚轨道（LIVVO）进行局域化。[27,28] 每个轨道均根据其内禀原子轨道（IAO）布居（按原子及原子壳层解析）、成键特征（σ、 π 或 δ）以及其成键或反键配对轨道进行表征。轨道还会按共轭 π 和金属 d 轨道集合进行分组。

使用局域化轨道键合分析（LOBA）确定 TS 处的金属氧化态和 d电子数，[29] 并计算基于IAO的键级。根据分子几何结构识别反应路径上的成键变化，并将其归类为成键、键断裂、键级变化、TS 附近的瞬态接触，或金属–配体距离变化（但未发生成键或断键）。这些信息为初始活性空间选择提供了化学依据。

2.2 活性空间选择

初始活性空间会在过渡态（TS）处自动生成（也可在用户指定的任一起始帧生成），采用 IBO/LIVVO 程序。最小活性空间包括单占轨道、五个金属 d 轨道，以及反应路径上参与成键和断键的轨道。其他轨道根据其化学相关性排序，优先考虑共价金属–配体轨道及其相关配对轨道，其次是与瞬态或次级金属–配体相互作用相关的轨道，以及涉及反应原子的共轭 π 轨道。这些其他轨道可用于扩展活性空间，但默认情况下不纳入其中。**化学家** 会检查分子结构和反应路径，并考虑金属氧化态、轨道性质及成键事件，依据预先制定的化学检查清单提出活性空间。**构建者** 会将这一方案与自动生成的选择进行比较，并构建一到三个候选活性空间，通常从最小选择到包含更多化学相关轨道的较大空间依次排列。当

IBO/LIVVO 程序无法得出明确的轨道选择时，**构建者** 也可以改用 AVAS 方法。为控制计算成本，**构建者** 所选的活性空间限制为最多包含 18 个电子和 18 个轨道（18e, 18o），即 2.4×10^9个行列式。对应单个行列式的活性空间也会被排除。与此同时，**Librarian** 会检索文献，并报告化学上相关的活性空间。**审稿人**会评估所提出的活性空间，并确定哪些候选方案应进入 CASSCF 计算。

2.3 过渡态（TS）运行与评估

对于用户指定的自旋多重度和基组，针对所选候选活性空间在过渡态（TS）处执行 SS-CASSCF 计算。CASSCF 收敛后，使用 tPBE 和 tPBE0 泛函进行多构型对密度泛函理论（MC-PDFT）[30] 计算，并执行三根态平均完全活性空间组态相互作用（SA-CASCI）计算。每项计算均使用自动化评估脚本进行分析，该脚本会检查收敛性、自旋污染、自然轨道占据数（NOONs）、主导电子组态，以及根据 meta-L¨öwdin 原子和壳层布居数确定的自然轨道（NOs）的化学性质。[31]该脚本还会监测初始活性空间中选定的轨道是否仍保留在优化活性空间内。一**致性审查员** 首先使用轨道相似性度量，将收敛的自然轨道与初始猜测进行比较；该度量定义为奇异值的平均值

$$s(\mathbf{C}_1, \mathbf{C}_2) = \frac{1}{n} \sum_{k=1}^{n} \sigma_k\big(\mathbf{C}_1^{\mathsf{T}}\, \mathbf{S}\, \mathbf{C}_2\big)\,, \tag{1}$$

其中，$\mathbf{C}_1$ 和$\mathbf{C}_2$ 是两组 n活性轨道的系数矩阵，$\mathbf{S}$ 是原子重叠矩阵。 σ_k 是重叠矩阵的奇异值。相似度 s低于 0.80 表明活性轨道子空间发生了显著变化，并会触发进一步检查，以确定是否需要进行轨道替换。一**致性审查员** 还会识别轨道

这些轨道在优化过程中可能带有芯层或里德伯特征，并进入了活性空间。

《style id='1'>化学家会检查这些结果，以确定优化后的活性空间是否保留了描述该反应所需的轨道。将特别关注参与成键和断键的轨道以及金属 d 轨道。对于 NOON 高于 1.995 或低于 0.005 的每个 NO，**构建者** 都会检查相应的相关轨道是否包含在活性空间中。如果缺少配对轨道，**构建者** 则会将其添加进来，或移除几乎完全双占据或空的轨道，并根据收敛的自然轨道构建修订后的活性空间。**SA 监测器**会检查每个态的自旋纯度和主导组态，并识别可能的态交叉或强态混合。当第一激发态与基态的能量差在 5.0 kcal/mol 以内，且满足 0.1 的规定权重判据时，**SA 监测器**会启动态平均 CASSCF 或线性化对密度泛函理论（L-PDFT）计算。[32] 否则，计算将保持为态特定计算。根据这些评估结果，过渡态计算将得到四种结果之一：接受；有条件接受，但需要沿反应路径进行额外检查；拒绝，并随后重新进行活性空间选择；或建议执行态平均计算。

2.4 PathRun 和审查

在 TS 处确定合适的活性空间后，CASSCF 计算会沿反应路径向两个方向传播，分别朝向反应物和产物。在每个几何构型处，CASSCF 之后会进行 MC-PDFT，或者在态平均计算中进行 L-PDFT，以及三根态平均 CASCI 计算。每个几何构型的初始轨道通过投影前一几何构型中已收敛的自然轨道获得，并在新的原子轨道基组中对其进行正交归一化。投影轨道的质量通过使用式 1 中定义的相似性度量 s 进行评估。比较不同几何构型时，使用跨几何构型重叠矩阵。在 CASSCF 优化期间，自动化脚本每隔

五次宏迭代（从第 10 次迭代开始）监测初始活性空间与当前活性空间之间的相似性。如果在连续两次检查中相似性低于 0.90，则停止计算。**构建者** 还会监测计算是否存在收敛问题，并终止停滞或失败的作业。

当路径运行到达反应路径的两端，或因收敛问题或轨道一致性问题而停止后，一致性审查员、SA 监测器和路径优化器会检查结果。一致性审查员使用轨道相似性度量和自然轨道的化学性质，对比相邻几何结构处的优化活性空间。当所有相邻活性空间之间的相似度均超过 0.97 时，反应路径通过一致性评估。如果不满足这一标准，则会找出导致变化的轨道，并将其列入评估报告。SA 监测器监测自然轨道占据数和主要组态相互作用系数的变化，以确定沿反应路径跟踪的是否为同一电子态。检测到态交叉或显著的态混合时，会执行额外的态平均完备活性空间自洽场测试，并在必要时引入态平均计算。路径优化器会判断不一致是否可能源于相邻结构之间几何变化过大。当键长变化超过 0.3 $^{\circ}$A，且投影初始轨道与前一几何构型处的轨道相似度较低时，系统会建议增加几何结构。这些中间结构通过依赖图像的成对势（IDPP）插值生成。[33] 因此，该框架不要求反应路径上的几何结构均匀精细，而仅在需要提高活性空间一致性时增加几何结构。

如果三位委员会成员中有任何一位拒绝接受结果，**协调器** 将根据其建议启动另一次反应路径计算。初步评估完成后，**一致性审查员**、**Librarian**和**审稿人** 使用全新的独立上下文执行最终审查，以减少早先决策的影响。**Librarian**将选定的活性空间与相关体系中报告的活性空间进行比较，**审稿人**评估其化学有效性，而**一致性审查员** 则再次检查反应路径中活性空间和计算设置的一致-

性。最终评估可能会促成接受、开展额外测试、改变所考虑的电子态，或重新进行活性空间选择。在后一种情况下，工作流将使用新的运行标识符重新启动。

2.5 报告与总结

在整个工作流中，**协调器** 会记录选定轨道的化学表征、每位委员会成员作出的决定、用于论证这些决定的标准，以及 successive assessments 的结果。每次活性空间试验都会单独建立记录，从而可以追溯选择过程及任何纠正操作。最终活性空间获得接受后，CASPER 会生成一份交互式报告，其中包含选定的活性空间、反应路径上各点的分子几何结构和自然轨道、通过不同电子结构方法获得的绝对能量和相对能量，以及完整的决策历史。分子结构和轨道可以通过三维表示进行检查。

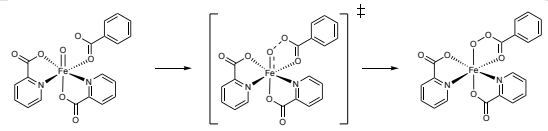
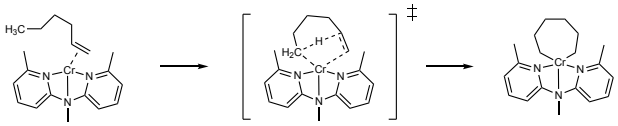
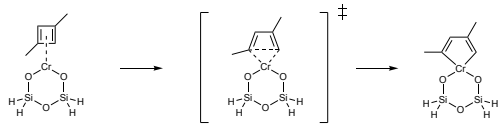
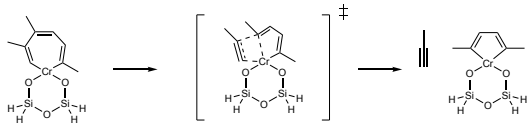
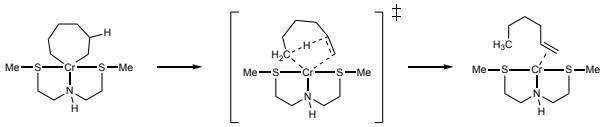
3 CASPER 的演示与结果

我们在 Gold-DIGR 数据集[34]中选取了五个有机金属反应，对 CASPER 进行了测试。表 1 展示了这些反应，它们涉及不同的断键和成键过程，可用于检验 CASPER 沿反应路径识别一致活性空间的能力。反应 A 涉及双重态下，铁氧代双(吡啶甲酸根)苯甲酸根配合物中配位的苯甲酸根与氧代之间形成双重态 O–O 键。[35]反应 B 涉及氢从 1-己烯转移，以形成带有双(吡啶基)胺配体的铬杂环庚烷。[36] 反应 C 描述配位的 η^4-二甲基环丁二烯扩环形成铬杂环戊二烯的过程。[37] 反应 D 涉及从铬杂环庚三烯配合物中脱去丙炔。[37] 反应 E 与反应 B 类似，但涉及

硫醚胺配体，以及氢从铬杂环庚烷转移至 1-己烯。[38]反应 B–E 处于三重态。

每个反应的反应物和产物结构均在 PBE0-D4/cc-pVDZ 理论水平下重新优化，采用 def2/J辅助基组，在库仑项的恒等分解下结合球链交换方法（RIJCOSX），并使用 `TightSCF` 收敛标准。在相同理论水平下使用 12 个图像进行爬升图像推挤弹性带（NEB）计算，随后以爬升图像为初始结构进行 TS 优化。当优化后的 TS 结构偏离爬升图像时，我们固定优化后的 TS，并对 NEB轨迹进行细化。目前的测试仅限于涉及单个过渡金属、具有单个能垒且不发生自旋交叉的反应。此外，还在 ωB97M-V/cc-pVDZ 理论水平下，采用相同的密度拟合和收敛标准进行额外的单点计算，以获得参考能量。DFT 和 NEB 计算使用 ORCA 6.0.1 完成。[39,40]

表 1：五种测试有机金属反应及其自旋多重度。

ID	公式	自旋多重度
A		2
B		3
C		3
D		3
E		3

使用 Claude Opus 5.5 并设置为中等努力程度运行 CASPER。使用 PySCF 2.14[41,42] 和 PySCF-forge 1.1.1.[43] 进行态特定和态平均 CASSCF、MC-PDFT、L-PDFT 以及强收缩二阶 n电子价态微扰理论（NEVPT2）计算。[44] GPU 加速代码来自 gpu4mrh。所有计算均采用 cc-pVDZ 基组，并使用 def2-universal-jkfit 辅助基组进行密度拟合。NEVPT2 计算未纳入 CASPER 工作流，但在结果中展示，以提供能量参考。

表 2 总结了 CASPER 选定的活性空间、所需的试验次数，以及采用的是 SS-CASSCF 还是态平均 CASSCF。活性空间一致性通过式 1 中定义的轨道相似性度量进行评估，该度量根据相邻几何结构的收敛活性轨道计算得出。CASPER 为全部五个反应确定了一致的活性空间，平均相似度介于 0.9946 和 0.9991 之间。试验次数从反应 B 的 4 次到反应 A 的 16 次不等。作者还检查了最终的活性空间，相应的自然轨道见补充图 S3–S8。在五个反应中，有四个采用 SS-CASSCF 即可满足要求。反应 C 沿反应路径发生了相同自旋多重度电子态之间的交叉，因此需要采用包含两个根的态平均 CASSCF。

我们会更详细地讨论反应 C 和 E，因为它们展示了活性空间选择中的两种不同挑战，以及 CASPER 如何应对这些挑战。

表 2：各反应的选择结果。

反应	试验次数	SS 或 SA	$(N_e,\ N_o)$	s (均值 ±标准差)
A	16	SS	(11, 11)	0.9991 ± 0.0028
B	4	SS	(10, 10)	0.9970 ± 0.0028
C	6	SA-2	(10, 11)	0.9957 ± 0.0020
D	10	SS	(14, 16)	0.9946 ± 0.0085
E	5	SS	(10, 10)	0.9977 ± 0.0017

3.1 反应 C：检测并处理态交叉

反应 C 展示了这样一种情形：在过渡态处，有不止一种活性空间被接受，但发生了态交叉，因此需要采用态平均处理，才能捕捉电子态的一致演变（图 2）。最初选定的活性空间为 (10e,9o)，包含五个 Cr 3d 轨道、参与断键的 Cr–C σ and σ^* 轨道，以及与环丁二烯环相关的一对 π/π^* 轨道。**化学家** 提议纳入另外两个 π^* 轨道，从而形成 (10e,11o) 活性空间。在过渡态（帧 06，F06）处对这两种选择都进行了测试。(10e,9o) 活性空间未能通过测试，因为在 CASSCF 优化过程中，一个 Cr 3d 轨道和一个 π^* 轨道被替换了。较大的 (10e,11o) 活性空间在过渡态处成功收敛，但沿反应路径进行的计算显示，在 F03 处，一个 Cr 3d 轨道被 Cr 4s 轨道替换，在 F09 处则被 Cr 4d 轨道替换。相应的轨道相似度分别降至 0.915 和 0.923（图 2c）。这些替换由**一致性审查员**识别出来，说明仅凭单一几何构型下收敛的 CASSCF 计算，不足以确定活性空间沿反应路径的质量。随后，CASPER 测试了更大的 (14e,13o) 活性空间，其中纳入了另外两个氧 σ 给体轨道。然而，在过渡态运行过程中，这些轨道被 Cr 3p 轨道替换。随后，**构建者** 提议移除第五个 Cr 3d 轨道，形成通过轨道相似性测试的 (10e,10o) 活性空间。尽管如此，**SA 监测器** 发现了轨道特征的变化，以及相邻几何结构之间自然轨道占据数的突变。三根态的 SA-CASSCF 计算显示，在 F06–F08 区间内电子态发生了交叉。因此，CASPER 对最低的两个三重态采用了态平均。所得的 (10e,10o) SA-CASSCF 计算通过了初步路径审查。然而，**Librarian** 和**审稿人** 建议纳入第五个 Cr 3d 轨道，以更完整地描述金属 d 轨道组。最终的 (10e,11o) SA-CASSCF 活性空间满足了整个反应路径上的一致性标准（图 2c）。相应的自然轨道见补充图 S5 和 S6，能量曲线

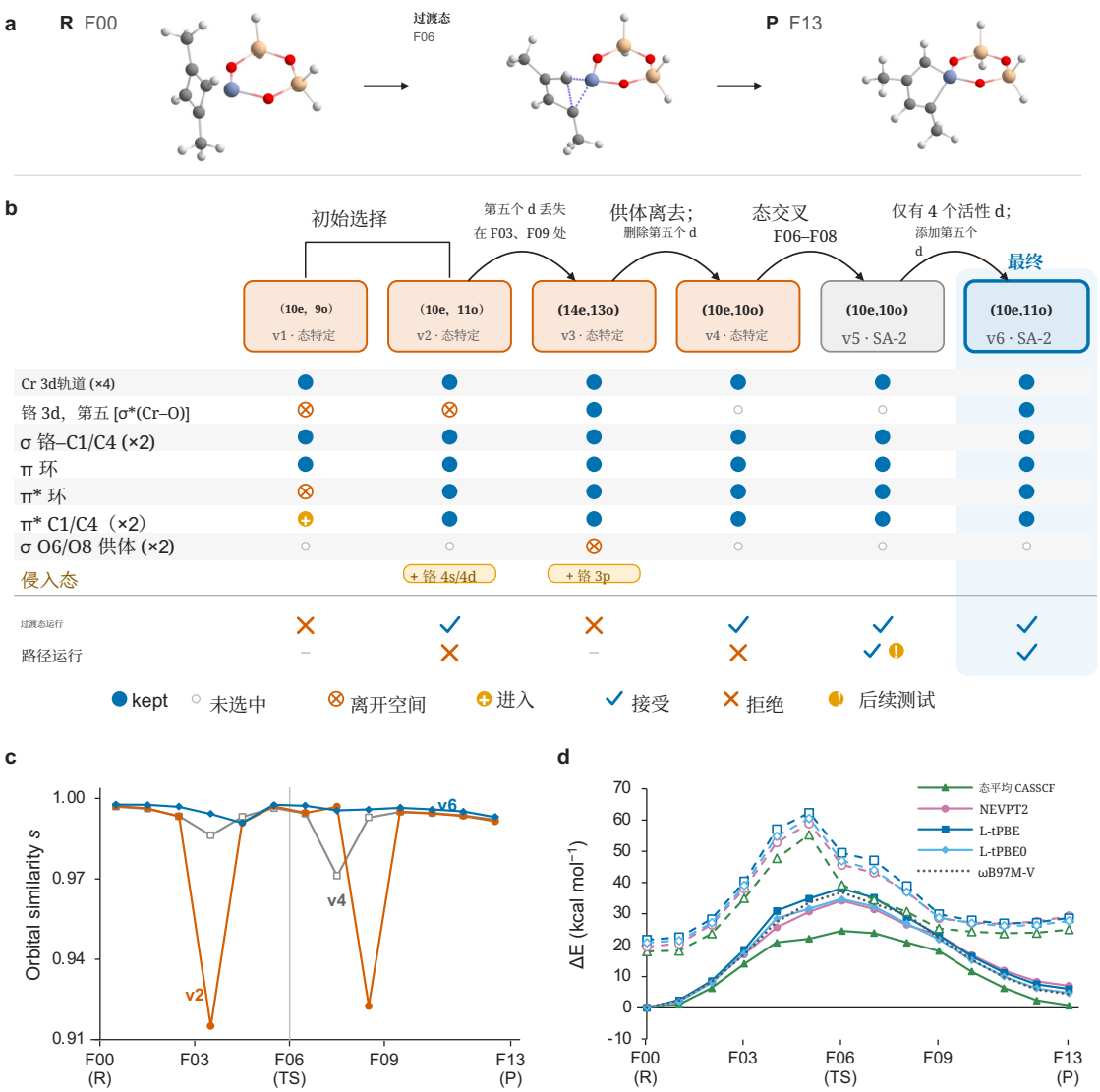


图 2：使用 CASPER 沿反应 C 路径逐步进行活性空间选择的过程。(a) 反应物、过渡态和产物的几何结构；(b) 工作流按选择顺序列出的六个试验活性空间。每个轨道行显示在每个试验中，该轨道在过渡态运行或路径运行期间是被保留、移出还是进入活性空间。底下两行显示过渡态运行和路径运行的判定结果。(c) 试验 v2、v4 和 v6 中相邻帧之间的轨道相似度 s。(d) 采用的 (10e, 11o) 相对于反应物 (F00) 的能量剖面，分别采用态平均 CASSCF、NEVPT2、L-PDFT（tPBE、tPBE0）方法计算，并附上相应的 ωB97M-V 结果。SS，态特定；SA-2，以相同权重对最低的 2 个三重态根进行态平均（分别以实线和虚线表示）。

使用态平均 CASSCF、NEVPT2 和 L-PDFT 得到的结果与图中的 ωB97M-V 进行比较

2d.

本例　　表明 CASPER 能够识别可能出现的两种不同问题　　e

在多参考计算中。第一种是在 CASSCF 优化期间替换化学相关轨道，即使计算已收敛，这种情况也可能发生。第二种是沿反应路径发生电子态性质变化，这需要采用不同的方法处理电子态。在反应 C 中，这两种问题都得到了识别和解决，且无需人工干预。

3.2 反应 E：维持具有化学意义的活性空间

反应 E 展示了活性空间选择中的另一项难题：在完全活性空间自洽场优化期间，与反应中的 C–H 键及 Cr 3d 轨道组相关的轨道很容易被替换（图3）。初始的 (8e,8o) 活性空间包含五个 Cr 3d 轨道，以及三个与氢转移所涉及的键相关的轨道。**化学家** 提出加入一个 Cr–C σ 轨道，从而形成 (10e,9o) 活性空间。在过渡态计算期间，一个与 Cr–配体 σ^* 相互作用相关的 Cr 3d 轨道被一个 Cr 4d 轨道替换。因此，**构建者** 构建了一个 (10e,10o) 活性空间，以纳入新增轨道，同时尝试保留原来的 3d 轨道。然而，在优化期间，同一个 3d 轨道再次被排除在外，而一个兼具反应氢原子和 Cr 4d 特征的轨道进入了活性空间。尽管发生了这种替换，所得活性空间沿反应路径仍保持一致。**化学家** 接受了这一选择，因为被排除的 3d 轨道在所有几何结构下均处于活性空间之外。然而，**审稿人** 质疑缺失的轨道是否会影响对金属电子结构的描述。因此，CASPER 又进行了两项测试。首先，使用先前计算得到的收敛轨道重新构建 (10e,9o) 活性空间，但同一个 Cr 3d 轨道再次被 4d 轨道替换。其次，将被排除的 3d 轨道加入收敛的 (10e,10o) 选择中，形成 (10e,11o) 活性空间。这项计算同样未能保留该 3d 轨道，也不满足轨道一致性判据（图 3c）。

因此，**化学家**保留了第三次试验中的（10e，10o）活性空间。

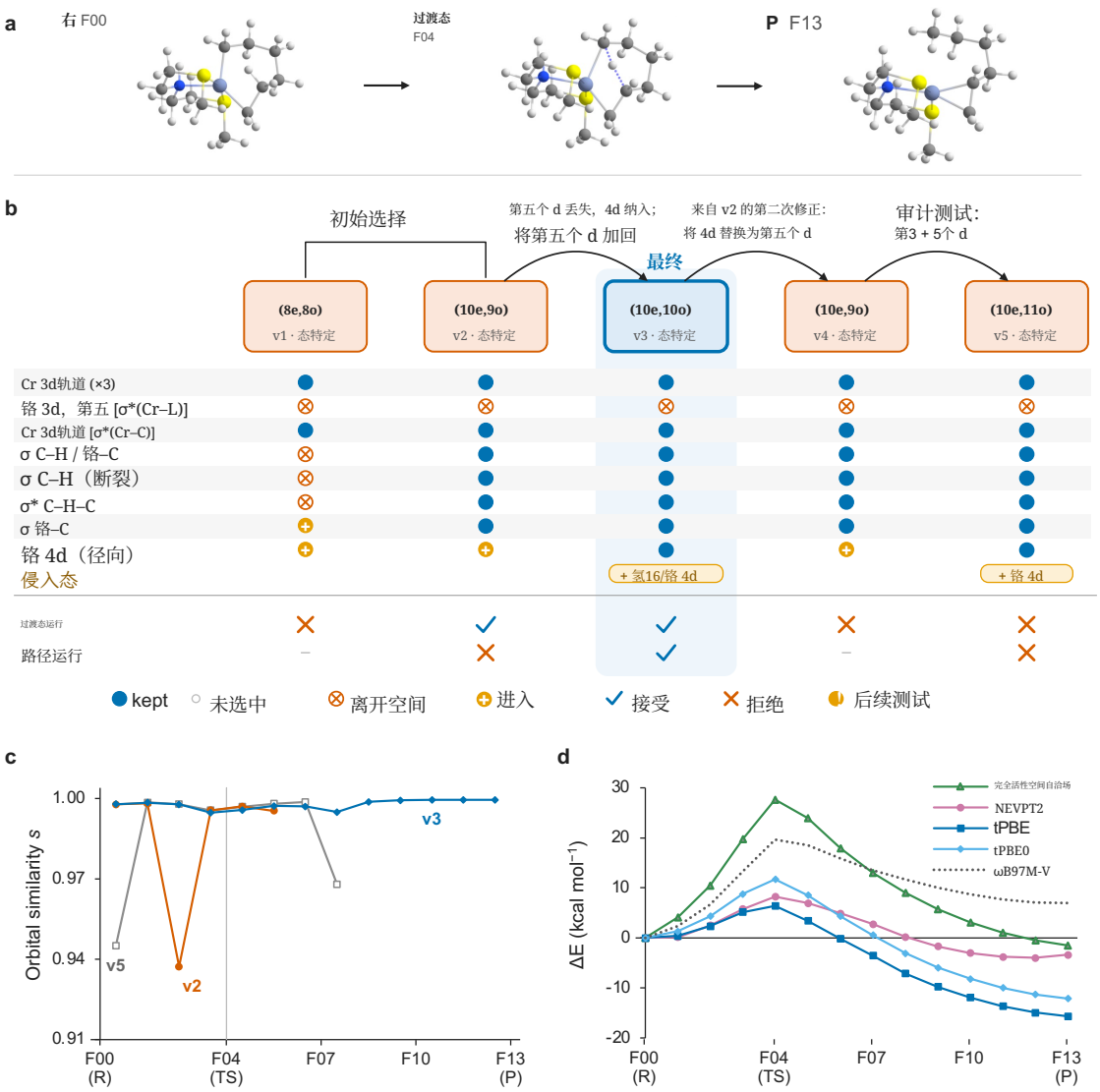


图3：使用CASPER沿反应 E路径逐步进行的活性空间选择过程。(a) 反应物、过渡态和产物的几何结构；(b) 工作流按其选择顺序排列的五个试验活性空间。每个轨道行显示，在每次试验的过渡态运行或路径运行期间，这些轨道是被保留，还是离开或进入了活性空间。底部两行显示过渡态运行和路径扫描的判定结果。(c) 相邻帧 s 之间针对试验v2、v3和v5的轨道相似度。(d) 所采用的(10e, 10o)基态相对于反应物(F00)在态特定完全活性空间自洽场、NEVPT2、MC-PDFT（tPBE、tPBE0）下的能量剖面，以及相应的 ωB97M-V结果。

反复排除同一个 Cr 3d轨道，表明它对该活性空间所描述的静态相关贡献不显著。最终的自然轨道如图S8所示。相应的完全活性空间自洽场、NEVPT2 和 MC-PDFT 能量曲线沿

反应路径（图 3d）。然而，与 ωB97M-V 相比，NEVPT2 和 MC-PDFT 对反应物与产物的相对能量预测不同，且更倾向于产物。这些差异表明，即使已确定一致的活性空间，电子结构方法的选择仍然很重要。

反应 E 也表明，活性空间更大并不一定更好。CASPER 执行的额外测试并未提高计算结果的一致性，而较小的（10e,10o）空间沿反应路径提供了更稳定的描述。重要的是，最终选择不仅依据计算的收敛情况，还考虑了轨道特征分析以及根据**审稿人**要求进行的额外测试结果。

3.3 不使用 CASPER 工作流与 Claude 的比较

为评估 CASPER 工作流的贡献，我们在不使用专门委员会成员及其指令的情况下，使用 Claude Opus 5.5 重复计算了反应 C 和 E。每项对照实验均在独立的 Claude 会话中进行，并可使用高性能计算资源。模型收到一条简短提示，其中指定了反应路径、基组，以及使用一致的活性空间执行 CASSCF 计算并获得平滑势能曲线的目标。各对照环境彼此隔离，且均与 CASPER 隔离。相应的轨道选择、决策历史、轨道相似度和能量曲线见补充信息。

仅 Claude 能够为两种反应确定轨道一致性较高的活性空间：反应 C 选用 (16e,14o)，反应 E 选用 (18e,14o)（图 S1 和 S2）。然而，这两个活性空间都包含几乎不活跃的 Cr 3s 和 3p 轨道，而描述相关电子结构并不需要这些轨道。更重要的是，对于反应 C，Claude 未能识别电子态之间的交叉，因此没有进行态平均 CASSCF 计算。对于反应 E，最终的活性空间未包含与反应中的 C–H 键相关的轨道；在 CASSCF

优化过程中，该轨道曾被反复替换。这一遗漏可能会影响对氢转移相关电子结构的描述。

这些独立的 Claude 会话能够获得收敛的计算结果和看似一致的活性空间。然而，轨道分析表明，重要的化学特征并不总能得到恰当描述。相比之下，CASPER 识别出了这些问题，并执行了额外的计算加以纠正。因此，这一比较表明，为语言模型提供结构化工作流、具体的化学判据以及对结果的独立评估十分重要。

尽管这些对照测试仅限于两个反应，但它们表明，仅具备运行电子结构计算的能力，并不足以实现可靠的自主多参考计算。CASPER 所实施的额外化学分析和迭代评估，对于识别那些原本可能一直未被察觉的问题至关重要。

4 结论与展望

我们开发了 CASPER，这是一种用于沿化学反应路径自动执行多参考计算的智能体框架。CASPER 将化学知识、电子结构计算和迭代评估相结合，用于选择活性空间，并确保随着电子结构沿反应坐标变化，活性空间保持一致。我们在五个过渡金属反应上测试了 CASPER；在所有情况下，它都能够识别一致的活性空间，并以极少的人工干预生成平滑的势能曲线。

这项工作的一项有趣发现是，独立会话能够执行计算并选择看似合理的活性空间。然而，更仔细的分析发现，电子结构的一些重要方面并未得到恰当描述。在反应 C 中，Claude 没有识别出电子态之间的交叉，也没有引入态平均。在反应 E 中，它选择的活性空间未能

包含参与反应的 C–H 键所涉及的轨道。CASPER 能够识别这些问题，并采取必要步骤加以纠正。尽管这一比较仅涉及两种反应，但结果表明，结构化工作流以及提供给智能体的专门化指令，对于获得具有化学意义的结果发挥着重要作用。

这仅仅是个开始。我们计划在数量更多的反应上测试 CASPER，包括涉及多个过渡金属、挑战性更大的体系。此外，我们还将研究不同语言模型的表现，以更好地理解底层模型与工作流各自的重要性。这些研究将帮助我们提高 CASPER 的效率和可靠性，并将其应用范围扩展到更复杂的化学体系。

CASPER 是迈向让先进的多参考方法更易于使用、更能有效解决化学和材料科学领域的难题的一个可喜开端。通过分析结果、识别可能存在的问题并提出改进计算的建议，CASPER 和其他人工智能代理能够帮助研究人员更深入地理解所研究体系背后的物理学和化学。我们并不认为这些代理会取代人类的专业知识，而是希望它们能起到补充作用，让科学家能够专注于最重要的问题。关于如何让这些工作流更可靠、更高效，我们仍有许多需要学习之处。

致谢

本材料基于美国能源部基础能源科学办公室化学科学、地球科学与生物科学部资助的工作，资助编号为 USDOE/DE-SC0023382；同时也基于美国能源部基础能源科学办公室重元素化学计划资助的工作，资助编号为 DE-SC0022572。M.R.H. 部分获得美国国家科学基金会 CHE-2513475号资助。A.C. 对多组态方法的开发工作获得美国国家科学基金会

CHE-2435218号资助。L.G. 感谢上述所有资助。作者感谢 NVIDIA学术资助计划提供本研究所用的图形处理器资源，感谢芝加哥大学研究计算中心（RCC）和国家能源研究科学计算中心（NERSC）。我们感谢在开发和实施本研究所述智能体工作流时使用 Claude Opus 5.5（Anthropic）。Claude 也协助了稿件的撰写和编辑。所有科学结果和结论均由作者审查并验证，作者对本研究内容承担全部责任。我们感谢 Gagliardi实验室成员提供的宝贵讨论。

数据可用性

CASPER代码、智能体指令和运行日志将在论文被接收后公开。

补充信息

不使用 CASPER 工作流的独立 Claude 会话结果。五种测试反应中 CASPER选定的活性空间的轨道和占据数。

参考文献

(1) Wardzala, J. J.; Kim, Y.; Khurana, R.; Mandal, M.; Delferro, M.; Liu, C.; Gagliardi, L. 从氧代到氧自由基再到双自由基：对金属有机框架节点上的甲烷活化进行系统的多参考计算。《美国化学会志》**2026年**,*148*, 27411–27420。

(2) Ghosh, A. 过渡金属自旋态能量学与非惰性体系：密度泛函理论在生物无机领域面临的挑战。《生物无机化学杂志》**2006**，*11*，712–724。

(3) Wardzala, J. J.; Hennefarth, M. R.; Agarawal, V.; Jangid, B.; Seal, A.; Hermes, M. R.; King, D. S.; Gagliardi, L. 化学与材料科学中的多参考方法：自动化活性空间、高效动态相关及扩展体系。《化学评论》**2026**，*126*，4592–4618。

(4) Vitillo, J. G.; Cramer, C. J.; Gagliardi, L. 多参考方法是用于催化建模的现实且实用的工具。《以色列化学杂志》2022年，62，e202100136。

(5) Choudhury, A.; Santra, S.; Ghosh, D. 理解生物系统中的光过程：需要准确的多参考处理。《化学理论与计算杂志》2024年，20，4951–4964。

(6) Mandal, M.; Hermes, M. R.; Berger, F.; Sauer, J.; Gagliardi, L. 使用多参考方法对负载型钒氧化物催化剂上丙烷的氧化脱氢进行建模。《物理化学杂志C》2025年，129，14418–14429。

(7) Veryazov, V.; Malmqvist, P. °A.; Roos, B. O. 如何为多组态量子化学选择活性空间?《国际量子化学杂志》2011年，111，3329–3338。

(8) Stein, C. J.; Reiher, M. 自动选择活性轨道空间。《化学理论与计算杂志》2016年，12，1760–1771。

(9) Stein, C. J.; Reiher, M. autoCAS：用于全自动多组态计算的程序。《计算化学杂志》2019年，40，2216–2226。

(10) King, D. S.; Gagliardi, L. 用于选择高通量多参考计算活性空间的轨道排序方法。《化学理论与计算杂志》2021年，17，2817–2831。

(11) King, D. S.; Hermes, M. R.; Truhlar, D. G.; Gagliardi, L. 通过自动化活性空间选择对多参考垂直激发计算进行大规模基准测试。《化学理论与计算杂志》**2022年**, *18*, 6065–6076.

(12) Wardzala, J. J.; King, D. S.; Gagliardi,L. 用于过渡金属化合物的自动多参考垂直激发。《物理化学杂志*A*辑》**2025年**, *129*, 2683–2691.

(13) Golub, P.; Antalik, A.; Veis, L.; Brabec, J. 面向强关联过渡金属体系的活性空间机器学习辅助选择。《化学理论与计算杂志》**2021年**,*17*, 6053–6072。

(14) Jeong, W.; Stoneburner, S. J.; King, D.; Li, R.; Walker, A.; Lindh, R.; Gagliardi, L. 通过对化学键解离进行机器学习，实现多参考方法的活性空间选择自动化。《化学理论与计算杂志》**2020**,*16*, 2389–2399。

(15) Osaro, E.; Mitra, A.; Jenkins, A. J.; Parker, K. A.; Lavroff, R. H.; Neufeld, V. A.; Kundu, A.; Kakekhani, A.; Rocca, D. RLEASE：强化学习高效活性空间引擎。*arXiv* **2026年**, arXiv:2606.07879。

(16) Kolodzeiski, E.; Stein, C. J. 用于量子嵌入的自动化、一致且不偏不倚的活性轨道空间选择。《化学理论与计算杂志》**2023年**，*19*，6643–6655。

(17) Tarocco,F.; Haase, P. A.; Pavo˘sevi´c, F.; Krishna, V.; Guidoni, L.; Knecht, S.; Stella, M. AEGISS——用于空间选择的原子轨道与基于熵的引导推断——一种面向量子化学和量子计算应用的新型半自动活性空间选择工作流。*arXiv* **2025年**, arXiv:2508.10671。

(18) Sayfutyarova, E. R.; Sun, Q.; Chan, G. K.-L.; Knizia, G. 从原子价轨道自动构建分子活性空间。《化学理论与计算杂志》 **2017年**，*13*，4063–4078。

(19) Sayfutyarova,E. R.; Hammes-Schiffer,S. 为共轭体系的多参考计算构建分子 π-氢轨道活性空间。《化学理论与计算杂志》**2019年**，*15*，1679。

(20) Seal, A.; Perego, S.; Hennefarth, M. R.; Raucci, U.; Bonati, L.; Ferguson, A. L.; Par- rinello, M.; Gagliardi, L. 用于多参考机器学习势的加权活性空间方案。美国国家科学院院刊 **2025年**，*122*，e2513693122。

(21) Bensberg, M.；Reiher, M. 化学反应路径上的对应活性轨道空间。《物理化学快报》**2023年**，*14*，2112–2118。

(22) Li, C.；Pan, K.；Sun, X.；Song, Y.；He, L.；Lu, T.；Ma, H.；Lan, Z.；Ma, Y. 迈向自动驾驶多参考计算：信念感知轨道选择。*arXiv* **2026年**，arXiv:2611.15682。

(23) Lee, V. C.；Rondinelli, J. M. 自主大语言模型智能体能否执行多参考量子化学计算？*arXiv* **2026年**，arXiv:2609.13357。

(24) P´erez-S´anchez, J. B.；Zou, Y.；Campos-Gonzalez-Angulo, J. A.；M¨uller, M.；Gustin, I.；Wang, A.；Hao, H.；Ko, T. W.；Choi, C.；Isbrandt, E. S. 等。El Agente Quntur：用于量子化学的科研协作智能体。*arXiv* **2026年**，arXiv:2602.04850。

(25) Li, T.; Sun, C. 使用关联感知智能体框架对强关联体系进行自动化多体模拟。*arXiv* **2026年**，arXiv:2610.00943.

(26) Knizia, G. 内禀原子轨道：量子理论与化学概念之间一座无偏的桥梁。《化学理论与计算杂志》**2013年**，*9*，4834–4843.

(27) Schmidt, M. W.; Hull, E. A.; Windus, T. L. 价层虚轨道：对最低未占分子轨道概念进行明确的从头算定量表征。《物理化学杂志*A*辑》**2015年**，*119*，10408– 10427.

(28) Derricotte, W. D.; Evangelista, F. A. 将局域化内禀价层虚轨道用作自动分类芯激发态的工具。《化学理论与计算杂志》**2017年**,*13*, 5984–5999。

(29) Thom, A. J. W.; Sundstrom, E. J.; Head-Gordon,M.LOBA：一种局域化轨道成键分析方法，用于计算氧化态，并应用于模型水氧化催化剂。物理化学化学物理**2009**,*11*, 11297–11304。

(30) Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. 多构型对密度泛函理论。《化学理论与计算杂志》**2014年**，*10*，3669–3680。

(31) L¨owdin, P. 关于分子和晶体理论中使用原子波函数所涉及的非正交性问题。《化学物理学杂志》**1950**，*18*，365–375。

(32) Hennefarth, M. R.; Hermes, M. R.; Truhlar, D. G.; Gagliardi, L. 线性化对密度泛函理论。《化学理论与计算杂志》**2023年**，*19*，3172–3183。

(33) Smidstrup, S.; Pedersen, A.; Stokbro, K.; J´onsson, H. 用于最小能量路径计算的改进初始猜测。《化学物理学杂志》**2014**，*140*，214106。

(34) Li, Z.; Yang, Y.; Savoie, B. 大规模从补充信息中挖掘有机金属反应路径。*ChemRxiv* **2026年**，DOI：10.26434/chemrxiv-2025-ccgfs/v2。

(35) Li, J.; Yang, L.; Cui, Y.; Yang, Z.; Zeng, G.; Pan, B. 过氧化物引导铁循环，在水相芬顿反应中定制生成活性氧化剂。*Environ.Sci. Technol.* **2025年**，*59*，3316–3324。

(36) Liu, L.; Liu, Z.; Tang, S.; Cheng, R.; He, X.; Liu, B. 是什么触发了Cr/(2,2'-二吡啶胺)催化剂从乙烯选择性三聚转向四聚？*ACS Catal.***2019年**，*9*，10519–10527。

(37) Liu, Z.; Cheng, R.; He, X.; Liu, B. Phillips Cr/二氧化硅催化剂催化甲基乙炔环三聚的反应性和区域选择性：一项密度泛函理论研究。《*ACS Catal.*》**2013年**,*3*, 1172–1183.

(38) Yang, Y.; Liu, Z.; Zhong, L.; Qiu, P.; Dong, Q.; Cheng, R.; Vanderbilt, J.; Liu, B. 在 SNS-Cr 介导的乙烯三聚中，无需去质子化即可发生铬 (I)/ 六重态与铬 (III)/ 四重态之间的自旋势能面交叉。《有机金属化学》**2011年**, *30*, 5297–5302.

(39) Neese, F. ORCA程序系统。《计算与分子科学评论》**2012年**，*2*，73–78。

(40) Neese, F. 软件更新：ORCA程序系统——6.0版。《计算与分子科学评论》**2025年**，*15*，e70019。

(41) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H. 等。PySCF程序包的最新进展。《化学物理学杂志》**2020年**，*153*，024109。

(42) Sun, Q.; Hermes, M. R.; Wu, X.; Zhai, H.; Zhang, X.; Ahmed, A. M.; Aucar, J. J.; Backhouse, O. J.; Banerjee, S.; Bao, P. 等。Python化学模拟框架：开源量子化学项目的十年历程。《化学物理学杂志》**2026年**，*165*，102502。

(43) pyscf-forge. `https://github.com/pyscf/pyscf-forge`，访问日期：2026-11-18。

(44) mrh. `https://github.com/MatthewRHermes/mrh`，访问日期：2026-11-18。