高等学校化学学报
CHEMICAL JOURNAL OF CHINESE UNIVERSITIES
研究论文

基于自由能微扰与机器学习协同优化的黑色素瘤抗原肽设计

王泰中[1,2]，钟青庐[2]，张　珊[2]，吴毅剑[2]，李金辉[2]，
庄　齐[2]，吴　松[1,2,3,4]，付　蕾[2,4]
（1. 安徽医科大学第二临床医学院，合肥 230601；
2. 深圳大学附属华南医院医学人工智能研究所，深圳 518111；
3. 深圳大学附属华南医院泌尿外科，深圳 518111；
4. 深圳大学电子与信息工程学院，深圳 518060）

摘要　针对黑色素瘤抗原肽与主要组织相容性复合体结合亲和力不足、疫苗疗效受限的问题，须建立基于物理与人工智能交互的协同优化框架，以高效筛选高亲和力候选肽段. 本文以 $PRAME_{190—198}$ 肽段及 HLA-A*11:01 复合物为研究对象，利用自由能微扰(FEP)精确量化抗原的氨基酸突变对结合自由能的影响，并驱动 FEPaML 模型进行多轮优化学习；应用该抗原优化框架对海量抗原突变组合进行高通量预测与排序，并运用分子动力学模拟的自由能微扰计算对机器学习模型预测的抗原进行验证. 结果表明，自由能微扰的精确物理数据校正后的机器学习模型 FEPaML 可以高效、准确地进行抗原的优化设计，并能够捕捉多位点突变的非线性组合效应；筛选出的最佳三突变体 I7F-E8S-K9R 的相对结合自由能低至−24.43 kJ/mol，其结合亲和力较野生型显著提升. 该方法实现了抗原肽亲和力的高效、精准优化，能够显著降低筛选成本与周期，可为黑色素瘤个体化疫苗设计提供理论依据与技术支持.

关键词　分子动力学；自由能微扰；机器学习；抗原优化
中图分类号　O643；O642　　**文献标志码**　A　　doi：10.7503/cjcu20260155

Free Energy Perturbation-Machine Learning Collaborative Optimization for Melanoma Antigenic Peptide Design

WANG Taizhong[1,2], ZHONG Qinglu[2], ZHANG Shan[2], WU Yijian[2],
LI Jinhui[2], ZHUANG Qi[2], WU Song[1,2,3,4*], FU Lei[2,4*]
(*1. The Second Clinical College of Anhui Medical University, Hefei* 230601, *China*;
2. Institute of Medical Artificial Intelligence, South China Hospital, Medical School, Shenzhen University, Shenzhen 518111, *China*;
3. Department of Urology, South China Hospital, Medical School, Shenzhen University, Shenzhen 518111, *China*;
4. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, *China*)

Abstract　To address the limited efficacy of peptide vaccines caused by insufficient binding affinity between melanoma antigens and the major histocompatibility complex (MHC) proteins, this study proposes a free energy

收稿日期：2026-04-16. 网络首发日期：2026-07-06.
联系人简介：吴　松，男，博士，教授，主要从事医学人工智能与智慧医疗转化方面的研究. E-mail：wusong@szu.edu.cn
付　蕾，女，博士，特聘研究员，主要从事计算化学与计算生物学方面的研究. E-mail：leifu@szu.edu.cn
基金项目：深圳市医学研究专项资金(批准号：E250200620，E250200622，E250200623)和重庆市自然科学基金创新发展联合基金(批准号：CSTB2025NSCQ-LZX0079)资助.
Supported by the Shenzhen Medical Research Fund, China(Nos.E250200620, E250200622, E250200623) and the Chongqing Natural Science Foundation Innovation and Development Joint Fund, China(No.CSTB2025NSCQ-LZX0079).

perturbation (FEP)-AI synergistic design framework for efficiently screening high-affinity candidate peptides. We focused on the $PRAME_{190—198}$ peptide-HLA-A*11:01 complex, and employed FEP to accurately quantify the impact of antigenic amino acid mutations on binding free energy. The resulting precise physical data were used to calibrate the FEPaML model through multiple rounds of iterative optimization. The optimized framework was then applied to perform high-throughput prediction and ranking of large-scale combinatorial mutations, followed by validation using the FEP calculation. The FEPaML model, calibrated with high-precision physical data from FEP, enabled efficient and accurate antigen optimization design and successfully captured the synergistic effects of multi-site mutations. The identified optimal triple mutant, I7F-E8S-K9R, achieved a relative binding free energy of −24.43 kJ/mol, showing significantly improved binding affinity compared to the wild type. This framework enables efficient and precise optimization of antigen peptide affinity, substantially reducing screening cost and duration while providing robust theoretical and technical support for personalized melanoma vaccine design.

Keywords Molecular dynamics; Free energy perturbation; Machine learning; Antigen optimization

黑色素瘤是一种来源于黑色素细胞的高度恶性肿瘤，因其易转移至脑、肺、肝等重要器官，致死率远高于其它类型皮肤癌[1~3]. 近年来，全球黑色素瘤的发病率持续攀升. 美国皮肤癌基金会报告显示，2025年美国约有105000例新发黑色素瘤病例，其中，超过一半的患者在诊断时已经处于转移性阶段[4]. 目前，临床治疗主要依赖手术及免疫检查点抑制剂，但多数患者仍面临免疫耐受、免疫逃逸及严重的免疫相关不良反应等问题，患者的5年生存率在疾病晚期急剧下降[5~7]. 因此，临床迫切需要更加精准、个体化的免疫治疗策略.

在多种新兴的肿瘤免疫疗法中，基于肿瘤抗原的个体化疫苗已成为当前研究的核心方向. 这类抗原由肿瘤细胞突变产生，通常不在正常细胞上表达，因此具有极高的特异性[8]. 肿瘤疫苗通过呈递特异性的抗原肽激活患者自身的免疫系统，识别并清除肿瘤细胞；相较于传统的抗肿瘤方法，肿瘤疫苗具有较低的免疫毒性和持久的免疫记忆等优势，有望实现精准免疫干预，并形成体内持续的监视体系[9~12]. 然而，许多潜在的肿瘤抗原肽与主要组织相容性复合体(Major histocompatibility complex, MHC)结合力较弱，导致抗原呈递效率低下，无法有效启动免疫级联反应[13]. 因此，如何提高抗原肽与主要组织相容性复合体的结合亲和力，从而增强抗原肽的呈递效率与免疫原性，成为亟待解决的难题[12,14].

传统的抗原筛选与优化方法主要依赖体外实验，如ELISPOT和流式细胞术等，需要消耗大量时间和资源，且往往集中于已知抗原或特定突变，难以满足大规模肿瘤抗原筛选的临床需求[15]. 随着计算生物学的飞速发展，基于结构的理性设计已成为肿瘤疫苗开发的重要辅助手段：其中，分子动力学(Molecular dynamics, MD)模拟能够为抗原设计提供原子级机理，而自由能微扰(Free energy perturbation, FEP)技术能够在热力学循环中精确计算氨基酸突变前后的相对结合自由能[16,17]. 上述方法能够为抗原优化提供可靠的物理依据，然而其计算成本随突变位点个数的增多而增长：在当前的算力条件下，FEP计算可以实现抗原全表位的单突变穷举扫描，对于多位点的组合突变空间，通常需要结合结构信息、实验数据等已有信息，在候选突变空间中进行有限采样，尚难以实现系统性的穷举，因此难以应用于肿瘤抗原序列空间的高通量筛选[18].

为克服以上挑战，基于大数据的人工智能方法应运而生. 传统的卷积神经网络和循环神经网络方法(如NetMHCpan[19,20]和MHCflurry[21]等)，利用海量的实验数据捕捉肽段序列中的特征基序进行亲和力预测，成为目前抗原表位筛选的主流方法[15,22]. 近年来，随着人工智能技术的发展，基于结构的预测模型(如BioEmu[23]等)开始引入三维结构层面的原子互作信息，以提高预测的效能. 然而，数据驱动的模型存在着显著局限：一方面，训练集的范围决定了模型性能，面对罕见人类白细胞抗原(Human leukocyte antigen, HLA)亚型或复杂突变时，模型易发生过拟合；另一方面，模型缺乏物理层面的可解释性，导致预测结果的假阳性率较高，限制了其在精准医疗中的应用效能[24]. 因此，将高精度物理计算引入机器学习模型，并利用物理数据对模型进行约束与校准，已成为提升抗原肽优化可靠性的重要

发展方向[25,26]. 本文基于自主研发的自由能微扰辅助机器学习(Free energy perturbation-assisted machine learning, FEPaML)模型[24], 面向黑色素瘤抗原肽优化任务, 构建了一个由FEP驱动的多轮协同的优化流程. 图1总结了该流程的整体设计思路: 与以往多侧重于提升模型预测精度的研究不同, 该流程不仅用于提高预测准确性, 还用于指导关键突变位点识别、候选突变采样、多突变组合空间排序及验证, 从而形成面向抗原肽-HLA体系优化的完整设计框架.

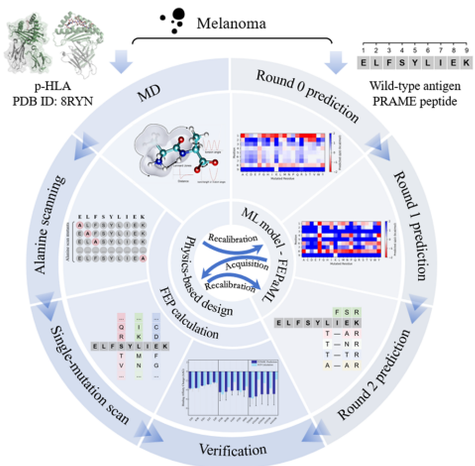


Fig. 1 Workflow of physics-driven and knowledge-guided antigenic peptide optimization

HLA-A*11:01亚型在东亚人群中具有高频率分布, 具有重要的人群覆盖意义[27]; PRAME(黑色素瘤优先表达抗原)是一种典型的癌-睾丸抗原, 在黑色素瘤组织中呈高水平过表达, 而在正常体细胞中几乎不表达, 是理想的免疫治疗靶点[28~30]. 本文以黑色素瘤相关抗原$PRAME_{190—198}$肽段及其与HLA-A*11:01形成的复合物为研究对象, 以野生型肽段ELFSYLIEK为初始序列, 首先通过全原子分子动力学模拟获得稳定的抗原肽-HLA复合物构象, 并利用自由能微扰计算精确量化突变前后的相对结合自由能变化. 在此基础上, 将FEP计算结果逐轮反馈至FEPaML模型, 使模型逐步适应特定体系的局部物理规律, 并进一步用于单突变、双突变和三突变组合空间的高通量预测与排序. 通过这种由FEP物理数据驱动的多轮协同优化策略, 实现了FEPaML模型在具体体系上的精准优化, 并成功筛选出具有较高HLA结合潜力的多位点突变候选肽段, 为黑色素瘤个体化疫苗的精准设计提供了新的理论依据与技术支持.

1 实验部分

1.1 分子动力学模拟

分子动力学模拟是基于牛顿经典力学原理的计算机模拟方法, 在原子水平上, 通过对多体系统随时间的运动轨迹进行积分, 来探究生物大分子的动态构象特征、热力学性质及相互作用机制[31]. 对于抗原肽-HLA复合物, 静态的晶体结构无法体现结合口袋内的动态适配过程与溶剂效应, 分子动力学模拟则能够全面展示该生物大分子在生理环境下的构象系综, 通过分析其动态轨迹可以评估复合物稳定性, 并可以通过计算自由能微扰得到不同序列的抗原肽和HLA的结合亲和力变化[32].

在本研究中, 所有分子动力学模拟均使用GROMACS 2023.5软件包[31]进行. 选用CHARMM 36全原子力场[33]来描述蛋白质与多肽的相互作用参数, 水模型选用TIP3P三点水模型[34]以精确模拟溶剂环境. 首先, 利用pdb2gmx工具生成复合物的拓扑文件, 将抗原肽-HLA复合物置于一个十二面体的模拟盒子中, 盒子边缘距离溶质表面至少1.0 nm, 以满足周期性边界条件, 并消除镜像相互作用的影响. 随后, 在盒子中填充显式水分子, 并添加钠离子和氯离子以中和体系电荷, 并将离子强度维持在0.15 mol/L的生理浓度.

为了消除初始结构中可能存在的不合理几何接触, 采用最速下降法对体系进行能量最小化, 以最大受力小于1000 $kJ\cdot mol^{-1}\cdot nm^{-1}$为收敛标准. 随后, 依次进行两个阶段的平衡模拟: 首先是100 ps的等温等容(NVT)系综平衡, 使用V-rescale恒温器[35]将系统温度耦合至310 K, 耦合时间常数为0.1 ps, 以此给予原子初始速度并稳定温度. 310 K接近人体正常生理温度37 ℃, 本研究关注黑色素瘤抗原肽与HLA-A*11:01复合物在免疫相关生理环境中的结合稳定性和界面相互作用, 因此, 采用接近人体生理温度的模拟条件, 有助于更合理地表征抗原肽-HLA复合物在体内环境下可能呈现的动态构象特征[36~39]. 接着进行100 ps的等温等压(NPT)系综平衡, 使用Parrinello-Rahman恒压器[40]将系统压力维持

在1×10^5 Pa，耦合时间常数为2.0 ps，以此调整体系密度．在平衡阶段，对蛋白质主链原子施加1000 kJ·mol^{-1}·nm^{-2}的位置限制，防止结构发生过度形变．

完成平衡步骤后，取消位置限制，并进行长时间的产品模拟．为保证充分采样，对该体系进行了3次独立的平行模拟，单次模拟时长为500 ns，总计1.5 μs．模拟采用LINCS算法[41]约束所有涉及氢原子的键长，积分步长设定为2 fs；长程静电相互作用采用粒子网格埃瓦尔德方法[42]计算，傅里叶网格间距设为0.16 nm，库仑相互作用与范德华相互作用的截断半径均设定为1.0 nm．模拟轨迹坐标每400 ps保存一次．模拟结束后，对模拟得到的轨迹进行去除周期性边界条件处理．随后，利用GROMACS软件包内置的标准分析工具gmx rms计算复合物主链相对于初始结构的均方根偏差(Root mean square deviation，RMSD)，在计算过程中进行了基于最小二乘法的平移与旋转拟合校正；通过分析RMSD随时间的演化曲线来评估体系的收敛性与稳定性，并基于轨迹聚类分析提取最具代表性的构象用于后续自由能微扰计算．

1.2　自由能微扰计算

采用自由能微扰方法定量评估抗原肽上氨基酸的点突变对其与HLA结合亲和力的影响[22]．自由能微扰是一种基于统计热力学的计算方法，其原理是采用“炼金术”方法建立热力学循环，并在计算中引入耦合参数λ，使系统逐步从一个状态(如野生型肽段)平滑转变为另一个状态(如突变型肽段)，从而精确计算该转变过程的自由能变化(ΔG)[43,44]．如图2所示，根据热力学循环原理，突变引起的相对结合自由能变化($\Delta\Delta G$)等于抗原肽在与HLA复合物结合态(Bound state)下的突变自由能与在溶液游离态(Free state)下的突变自由能之差[45]，即：

$$\Delta\Delta G = \Delta G_{\mathrm{bound}} - \Delta G_{\mathrm{free}} \tag{1}$$

式中：$\Delta G_{\mathrm{bound}}$和$\Delta G_{\mathrm{free}}$(kJ/mol)分别为结合态和游离态下由野生型肽段转变为突变型肽段的自由能变化．在该定义下，$\Delta\Delta G < 0$表明突变后抗原肽-HLA复合物的结合在热力学上更有利；而$\Delta\Delta G > 0$则表明突变后结合自由能升高，提示结合稳定性或亲和力可能降低．

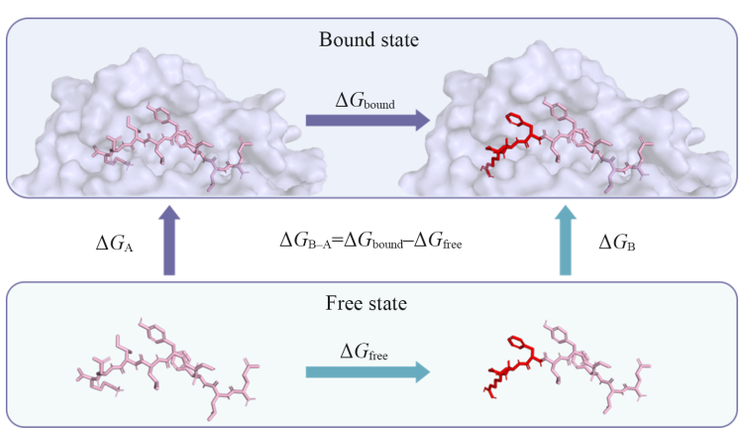


Fig. 2　Thermodynamic cycle of free energy perturbation calculations

所有的FEP模拟均使用NAMD软件(版本2.14)[16]进行．为了精确描述突变过程中的能量变化，采用双拓扑策略描述野生型与突变型残基：在体系中同时构建野生型与突变型残基的拓扑结构，并通过耦合参数λ共同控制其相互作用的强度，实现从初始状态(λ=0)至终态(λ=1)的演化[46]．

所有FEP计算均采用NAMD内置alchemical FEP模块完成，突变体系基于CHARMM36 hybrid topology构建．对炼金术转化路径进行离散化处理，从而克服原子变化过程中可能出现的能量势垒，并确保相邻状态之间的相空间充分重叠．窗口采用非线性分布，在λ接近0和1的端点区域设置更密集的采样点，以精确捕捉范德华势变化引起的能量波动，而在中间区域的排布则相对稀疏．具体而言，计算共设置35个λ值(0，0.00001，0.0001，0.001，0.01，0.02，0.06，0.10，0.14，0.18，0.22，0.26，0.30，0.34，0.38，0.42，0.46，0.50，0.54，0.58，0.62，0.66，0.70，0.74，0.78，0.82，0.86，0.90，0.94，0.98，0.99，0.999，0.9999，0.99999和1.00)，对应34个相邻λ窗口．为避免λ端点附近的数

值奇异性，采用soft-core vdW[46]修正Lennard-Jones势中的距离项：$r_{ij}^2 \rightarrow r_{ij}^2 + \delta(1-\lambda_{LJ})$[47,48][其中，$r_{ij}$(nm)表示原子$i$与$j$之间的距离；$\lambda_{LJ}$为vdW有效耦合参数；$\delta$(nm^2)为soft-core vdW距离平方修正系数，对应NAMD中的alchVdWShiftCoeff]. 本文设置alchVdWShiftCoeff=1.0，按NAMD输入约定换算为0.01 nm^2；同时，设置alchElecLambdaStart=0.1，alchVdwLambdaEnd=1.0，alchDecouple=off. 多项针对相似肽-蛋白系统的FEP计算研究支持了该设置的合理性，且足以实现准确的自由能估算[38,39,49].

所有FEP窗口的模拟均在NPT系综(300 K，1×10^5 Pa)下进行. 对于每一个λ窗口，总运行时长为300 ps，指定前100 ps为体系弛豫阶段，用以适应哈密顿量的微扰，随后的200 ps用于生产采样. 为确保充分覆盖构象空间，并验证结果的统计鲁棒性，对每一个突变体均执行5次不同初始速度分布的独立平行重复实验；对于每一个单点突变，有效采样时间累积达到70 ns：200 ps模拟×35窗口×5次重复×2个状态(结合态，游离态)=70 ns. 这种多重独立采样的策略能够有效降低统计误差，确保自由能计算的收敛性[16].

在数据处理阶段，采用加权直方图分析法，对34个窗口的采样数据进行全局积分与重加权[50]；最终报告的相对结合自由能变化值由NAMD FEP基于Zwanzig指数平均公式计算，最终结果以5次重复的mean±SEM(均值标准误差)表示，附带95%置信区间，以量化预测结果的可靠性[51].

1.3 自由能微扰辅助机器学习模型

自由能微扰计算能够提供高精度的结合能数据，然而其计算成本高昂，难以实现庞大突变空间的穷举搜索[18]；传统的机器学习方法虽然支持高通量筛选，但对于模拟肽结合亲和力的排序难以达到物理级精度[25,26]. 为此，使用了自主开发的FEPaML模型[24]，该模型设计了一种基于贝叶斯优化的深度学习框架(图S1，见本文支持信息)：首先，基于免疫表位数据库(Immune epitope database，IEDB)中筛选的大规模抗原肽-HLA结合数据集进行训练，构建基于序列表征与免疫知识的先验模型；然后，利用高精度的FEP计算数据作为物理依据，不断修正预训练的先验模型，从而以较小的计算成本实现高精度、鲁棒性预测，兼顾物理规律的严谨性与数据驱动的高效性.

FEPaML的核心机制包括深度学习预测与贝叶斯优化两个层面：模型采用基于Transformer的深度神经网络架构，在输入层，肽段序列通过One-hot编码及正弦位置编码转化为矩阵表示，以显式保留序列信息；在编码器层，利用多头自注意力机制并行捕捉氨基酸残基间的长程依赖关系[52]. 在优化层面，该模型通过自适应双重损失函数，将大量粗糙的二分类数据与少量精确的FEP数据有效融合. 模型的优化遵循贝叶斯推断逻辑：将深度学习的初始预测视为先验分布，通过采集最具潜力的样本进行FEP物理验证，并将物理计算结果反馈回模型进行校正，从而获得更逼近真实物理化学特征的后验模型[53,54].

采用FEPaML模型预测候选突变肽与HLA分子的结合自由能变化. 模型分别预测突变肽和野生型肽段的结合自由能，并计算相对结合自由能变化：

$$\Delta\Delta G_{\mathrm{pred}} = \Delta G_{\mathrm{mutant}} - \Delta G_{\mathrm{original}} \tag{2}$$

式中：$\Delta\Delta G_{\mathrm{pred}}$(kJ/mol)为FEPaML预测的突变体相对于野生型肽段的相对结合自由能变化；$\Delta G_{\mathrm{mutant}}$和$\Delta G_{\mathrm{original}}$(kJ/mol)分别为突变肽和野生型肽段与同一HLA分子的预测结合自由能. $\Delta\Delta G_{\mathrm{pred}}$与FEP计算得到的突变相对结合自由能变化$\Delta\Delta G$在物理含义上相对应，因此作为候选突变体亲和力变化的主要评价指标.

为了在特定抗原肽-HLA体系中实现多点突变的高效寻优，提出了如图3所示的协同优化流程：首先，利用FEPaML预训练权重，以野生型肽段序列为输入进行初步预测，并将预测结果与丙氨酸扫描的计算结果进行比对. 随后，将丙氨酸扫描得到的9组高精度物理数据整合至训练集，对模型进行第一轮优化，使模型迅速适应该体系的物理特征. 依据模型该轮更新后的预测结果，不仅针对核心位点(P4，P6，P8)进行了单突变全扫描计算，还设计了“补充突变”环节：即对于非核心位点，从模型第1轮输出的单突变预测热图中，筛选预测$\Delta\Delta G$值表现相对最优的前两名突变方案纳入FEP计算. 这一过程是在物理层面对全肽段的二次扫描，通过引入更广泛的采样数据来校准模型在非核心区域的预测偏

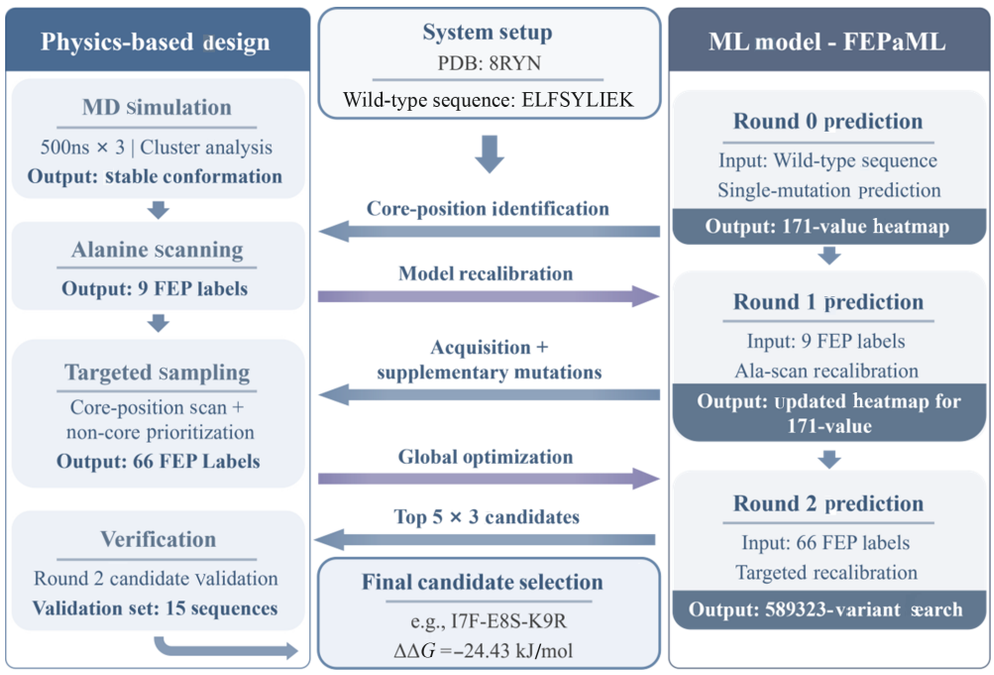


Fig. 3　FEP-FEPaML collaborative optimization strategy for antigenic peptides

差，有效抑制模型对局部核心位点的过拟合，从而提升全局预测的鲁棒性．紧接着，将包含单突变全扫描计算与补充突变的共66组高精度物理数据再次输入模型进行第二次优化，输出第二轮包含单突变、双突变及三突变全空间的$\Delta\Delta G$预测值．最后，依据该轮优化结果，分别遴选出单、双、三突变预测亲和力最优的前五名候选序列，进行FEP计算验证．

通过优化策略与采样机制的任务化整合，本研究将机器学习模型从通用的初筛工具拓展为面向特定抗原肽-HLA体系的筛选流程．该流程将FEP计算、分层采样、模型的多轮校准和多突变组合空间搜索任务整合起来，以丙氨酸扫描初步建立体系特异性的物理约束，再结合模型与结构分析共同提示的关键位点进行全扫描，同时通过补充突变扩展非核心区域的采样范围，降低模型对局部位点的依赖；最后应用多轮校准后的模型遍历单突变、双突变及三突变组合空间，并对高潜力候选突变体进行FEP验证．通过上述设计，模型能够逐步修正特定体系下的预测偏差，并在有限FEP计算成本下完成候选突变体的优先级排序与验证．

2　结果与讨论

2.1　抗原肽-HLA体系的分子动力学模拟

HLA-A*11:01在东亚人群中有着高频率分布，该等位基因在中国也具有较高的群体覆盖率，与肿瘤免疫应答的密切关联[55]．本研究基于HLA-A*11:01等位基因进行黑色素瘤相关抗原PRAME免疫优势表位的优化[27,56]．体系的构建源于RCSB PDB数据库(https://www.rcsb.org/)中PDB ID为8RYN的晶体结构，野生型肽段为PRAME蛋白第190~198位序列(ELFSYLIEK)[57]．如图4(A)所示，从8RYN中截取了抗原肽与HLA的二元复合物部分，作为后续分子动力学模拟的初始构象．抗原肽与HLA的结合状态如图4(B)所示，肽段呈延伸构象横跨于HLA结合槽内，其中第二位点P2(亮氨酸)与第九位点P9(赖氨酸)的残基分别深埋于槽底的锚定口袋中，起到核心固定作用．随后进行了能量最小化处理，以消除截取操作可能造成的局部原子冲突以及不合理的空间位阻．

对上述的抗原肽-HLA结构分别进行了3组独立的分子动力学模拟(平行组1~3)，并通过计算骨架原子的均方根偏差来量化模拟结构的收敛性．如图4(C)所示，3组模拟轨迹在经历短时间的构象调整后均趋于平稳．为提取最具代表性的稳定构象，选取了3组轨迹的稳定阶段(模拟后段300~500 ns数据)分别进行聚类分析．其中，平行组1在模拟后段表现出最优的稳定性，RMSD曲线收敛于0.26 nm

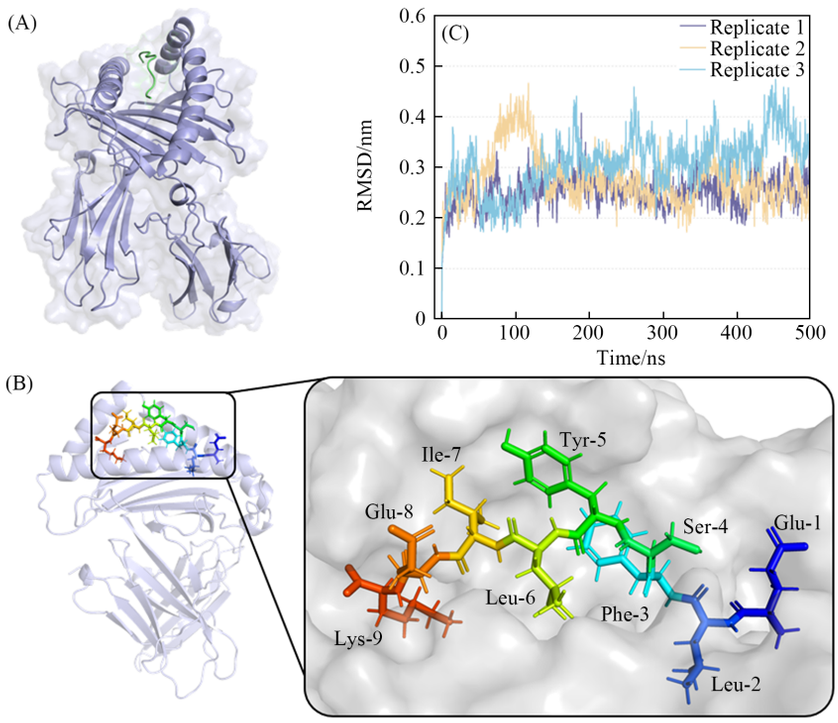


Fig. 4 Structural construction and dynamic stability analysis of the PRAME$_{190—198}$-HLA-A*11∶01 complex

(A) Initial structure of the antigenic peptide-HLA complex, where HLA is shown in purple cartoon representation and the peptide is shown in light green cartoon representation; (B) representative complex structure at 400 ns of molecular dynamics simulation; in the magnified view, HLA is shown as a gray surface model and the peptide is shown as a colored stick model; (C) backbone RMSD profiles from three independent MD simulations.

左右；平行组1在400 ns处的结构位于最大核心构象族群的中心位置，其构象最具有典型性，因此，作为后续自由能微扰计算的统一初始结构.

为进一步阐明肽段与HLA结合口袋的相互作用情况，计算了模拟过程中各残基的埋藏表面积(Buried surface area，BSA)，以评估残基侧链被HLA口袋包裹的程度. 其中，如图5(A)和(B)所示，抗原肽第二位点P2(亮氨酸)和第九位点P9(赖氨酸)氨基酸的平均埋藏百分比分别为96%和98%，显著高于其它位点，其绝对埋藏面积分别为1.351和1.809 nm^2，定量地证明该位点氨基酸提供了巨大的疏水接触面，被HLA结合口袋深度包裹；同时，这一结果与特异性HLA-A*11∶01锚定基序相一致，确认了P2和P9位点对固定肽段构象的决定性作用[58]. 值得注意的是，P3(苯丙氨酸)也表现出了极高的埋藏特征，其绝对面积达1.415 nm^2，占比92%，这提示P3在该体系中也具有关键的辅助锚定作用. 相比之下，P4(丝氨酸)、P6(亮氨酸)和P8(谷氨酸)的BSA值显著较低；其中，P4位点绝对埋藏面积仅为0.117 nm^2，平均埋藏百分比为16%，说明其侧链几乎完全伸出结合槽并朝向溶剂一侧. 这类残基具有高暴露、低接触的性质，一方面说明其具备典型的T细胞受体(T cell receptor，TCR)识别位点的特征；另一方面则提示这些位点与HLA之间存在未被填充的空腔，或存在不饱和的接触界面[37,59,60].

此外，BSA时间序列分析显示，野生型肽段的总BSA值围绕8.738 nm^{2}水平波动[图5(C)]，证明该基准复合物在动力学模拟过程中整体保持稳定[37].

上述结构分析能够为亲和力优化提供直接的物理依据：在P4，P6，P8等低埋藏位点引入体积更大或互补性更强的氨基酸，有望通过增加接触面积和填补界面空腔来提升结合自由能. 然而，低BSA并不必然意味着该位点一定具有可优化性，低BSA位点也可能出现突变无效甚至不利的情况；BSA主要反映残基侧链在结合界面中的空间埋藏程度和接触面积，而突变是否真正有利仍取决于侧链大小、电荷性质、疏水性、构象适配以及局部相互作用网络等多种因素[13,61,62]. 因此，上述位点在本研究中被作为优先优化位点，而非全部潜在可优化位点；其它位点的潜在贡献仍需结合FEP计算和FEPaML模型预测进一步判断.

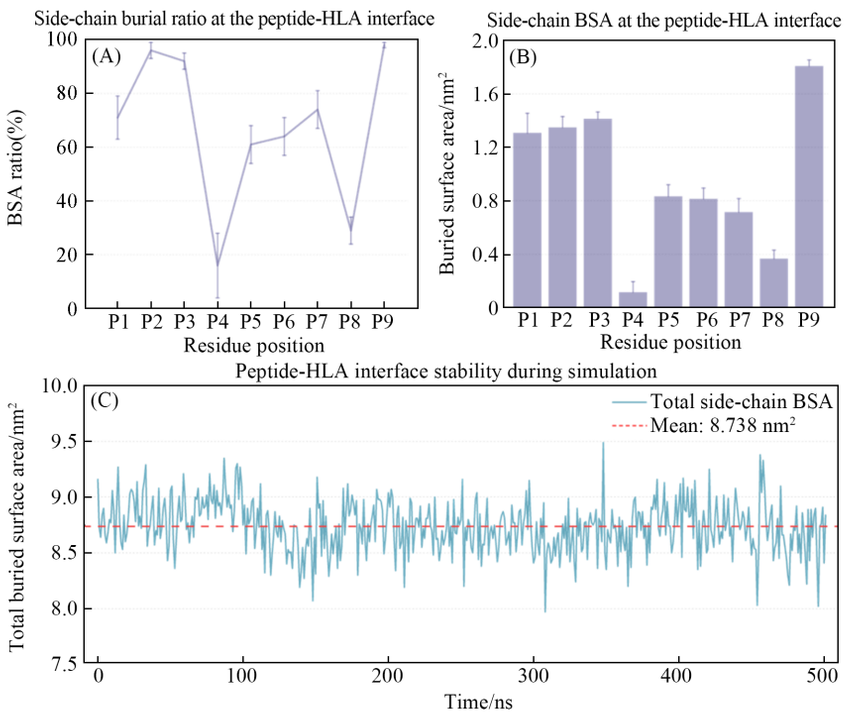


Fig. 5　Buried surface area analysis of the antigenic peptide-HLA interface

(A) Average burial percentage of each peptide residue; (B) absolute buried surface area of each peptide residue; (C) time evolution of total buried surface area of the wild-type peptide.

2.2　FEPaML 用于初步筛选及丙氨酸扫描

在构建复合物结构的基准模型后，应用FEPaML进行了第0轮粗筛，并同时开展丙氨酸扫描用于交叉验证. 将野生型肽段的序列信息输入FEPaML基于IEDB数据库的预训练模型，输出得到最初版的预测热图. 在热力学层面，相对结合自由能变化为负值表示突变增强了肽段与HLA的结合亲和力，在热力学上有利；正值则表示突变削弱了结合稳定性. 如图6(A)所示，在单突变预测热图中，红色区域代表预测结合亲和力提升，蓝色代表降低：模型在P1，P4及P8位点呈现出较多红色色块，表明这些位点在进行氨基酸突变时，预测自由能为负值，提示具有较高的亲和力提升空间.

丙氨酸扫描的计算结果如图6(B)所示，P4，P8位点的氨基酸突变为丙氨酸时，FEP计算的$\Delta\Delta G$值分别为−5.52和−9.04 kJ/mol，表明突变结果呈现有利于结合的趋势；P4位点突变的数值为

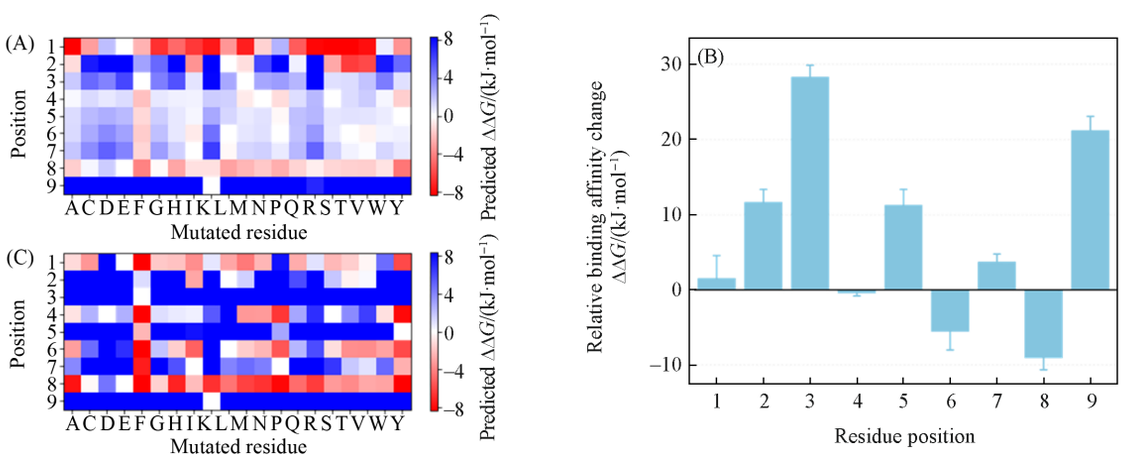


Fig. 6　Preliminary FEPaML screening and alanine scanning

(A) Round 0 FEPaML prediction heatmap for single mutations. Red regions represent predicted affinity enhancement, while blue represents reduction; (B) FEP-based $PRAME_{190—198}$ alanine scanning of the wild-type peptide; (C) round 1 FEPaML prediction heatmap after alanine scanning feedback.

−0.42 kJ/mol，表明突变对结合无显著负面影响．FEPaML的机器学习预测与FEP的物理计算在判别优化位点时结果基本一致，证实了模型在初始阶段应用的可靠性．

FEPaML的预测与FEP的物理计算结果说明了P4，P6，P8位点在热力学上的可塑性，而这些位点较低的埋藏表面积则证明，原有的侧链可能尚未与HLA口袋形成相对紧密的相互作用[59]．FEP计算结果为负值或在零点附近，说明这些位置的氨基酸突变具有增强抗原肽和HLA结合亲和力的潜力[63]．因此，通过数据驱动的机器学习模型FEPaML与物理驱动的FEP计算相结合的方法，进行突变位点的大规模筛选与交叉验证，并进一步引入更加匹配的侧链来增加肽段与HLA的结合亲和力．

2.3 FEP计算用于机器学习模型的协同优化

为了解决FEP计算通量较低、无法完全覆盖多突变的组合空间，而数据驱动的机器学习方法缺乏体系特异性和精准度普遍较低的双重难题，构建了基于物理与人工智能深度交互的协同优化流程．首先，将丙氨酸扫描获得的高精度$\Delta\Delta G$数据作为第一批物理数据输入给FEPaML模型，进行第1轮的针对性优化训练．第1轮优化的FEPaML预测结果如图6(C)所示，模型进一步确定了P4(丝氨酸)、P6(亮氨酸)、P8(谷氨酸)为潜在关键位点，并给出了具体的氨基酸突变方向建议．基于此，对P4，P6及P8位点进行了单突变的全扫描FEP计算，并在后续的优化中将计算结果再次反馈给模型．

单突变的全扫描FEP计算结果揭示了各关键位点独特的构效关系特征．如图7(A)所示，P4位点通过引入小体积疏水(甘氨酸、丙氨酸)、大体积芳香族(色氨酸)或碱性残基(赖氨酸)均能提升亲和力，其中，S4W突变体(将P4位点的丝氨酸突变为色氨酸)效果最优，其$\Delta\Delta G$为−2.80 kJ/mol．如图7(B)所示，P6位点能够兼容多种理化性质侧链：将该位点的亮氨酸突变为丙氨酸、半胱氨酸、组氨酸、异亮氨酸、赖氨酸、脯氨酸、精氨酸、苏氨酸或缬氨酸均能获得有利的结合自由能变化，证明该位点具备多样的构效优化路径．其中，最优突变体为L6P(将P6位点的亮氨酸突变为脯氨酸)，其$\Delta\Delta G$达到−8.03 kJ/mol．如图7(C)所示，P8位点表现出更宽的结构容忍度，共有14种不同种类的残基突变产生了有利的能量贡献；该位点优化的关键是消除原有的静电排斥作用，通过引入小体积极性或中性残基替代原有的谷氨酸，大幅改善了抗原肽-HLA复合物的结合自由能[64]；其中，最优突变体为E8S(将

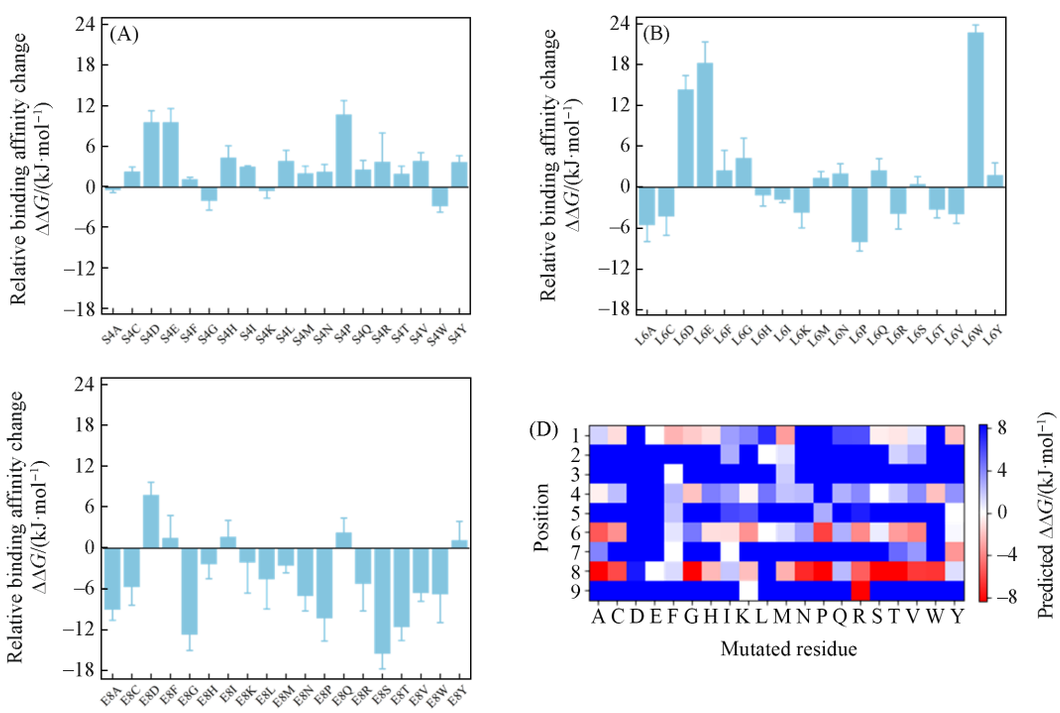


Fig. 7　Full-scan FEP results at key sites and round 2 FEPaML prediction heatmap

(A—C) Full-scan FEP results for single mutations at P4(A), P6(B), and P8(C) sites; (D) round 2 FEPaML prediction heatmap after FEP-feedback correction.

P8位点的谷氨酸突变为丝氨酸），其ΔΔ*G*达到−15. 52 kJ/mol.

为防止模型在后续优化过程中过度拟合于上述位点，依据第1轮模型和FEP交互之后输出的预测热图，设计了针对抗原肽其余位点的“补充突变”验证，相关数据列于表S1（见本文支持信息）.

汇集上述“单突变的全扫描计算”与“补充突变”的所有FEP数据后，将共计66组高精度物理数据再次输入FEPaML模型，对模型进行第2轮优化训练. 经过此轮校准，FEPaML模型能够准确捕获该抗原肽-HLA体系的物理规律，并在此轮预测中完成了对全位点单突变、全组合双突变及全组合三突变的遍历预测与排序.

单突变预测热图显示，经过第2轮物理数据校正，FEPaML模型将高潜力突变位点进一步收敛于P6和P8位点[图7(D)]，该预测趋势与前述单突变全扫描FEP计算的结果一致；此外，模型通过对“补充突变”数据的学习，识别出K9R同样为具有亲和力提升潜力的单点突变. 在多位点组合的预测中，FEPaML模型能够有效捕捉残基间的非线性组合效应：如表S2~表S4（见本文支持信息）所示，随着突变位点数目的增加，预测的ΔΔ*G*均值呈现下降趋势. 具体而言，模型通过学习单突变的特征，在双突变空间中预测出了能量更低的组合，其中，最佳双突变体E8T-K9R的预测ΔΔ*G*值为−20. 50 kJ/mol；在三突变空间，模型进一步探索到了理论上的能量极小值区域，最佳突变组合I7F-E8S-K9R的预测ΔΔ*G*值为−24. 43 kJ/mol. 第2轮的预测结果说明，优化后的FEPaML模型具备了从局部优化扩展至多位点协同设计的能力.

2.4 FEPaML模型预测值与FEP计算值的对比验证

为验证第2轮筛选策略的准确性，并确定最终的候选序列，用FEP的计算结果来验证机器学习的输出结果. 依据第2轮输出的ΔΔ*G*预测排序，分别从单突变、双突变及三突变类别中选取模型预测亲和力最优的前5名方案[如图8(A)所示的共15种候选肽]，进行FEP验证计算.

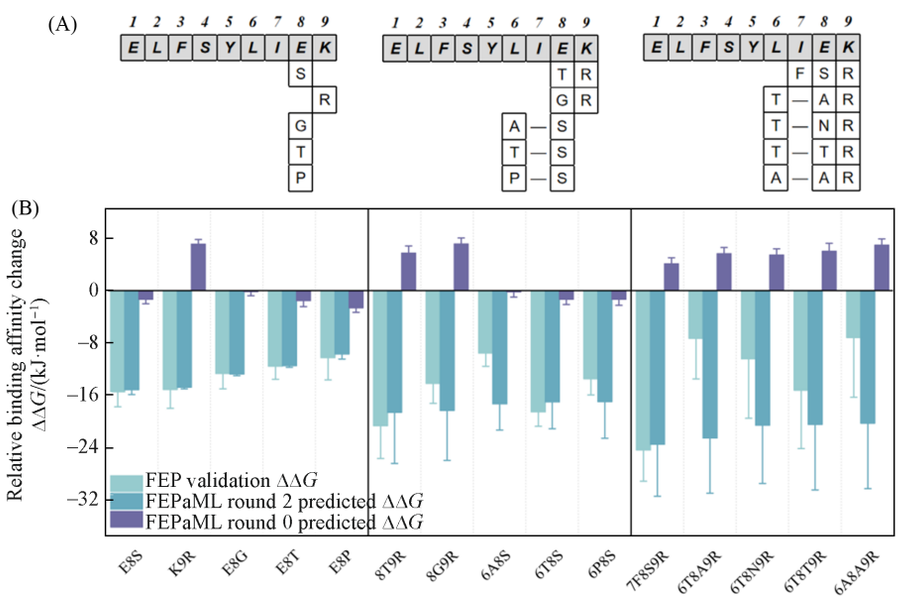


Fig. 8 FEP validation of selected candidate peptides and FEPaML predictions before and after FEP-feedback correction

(A) Fifteen candidate peptides selected from round 2 FEPaML rankings for FEP validation; (B) comparison of FEPaML-predicted and FEP-validated ΔΔ*G* values. The FEPaML prediction values represent the mean outputs of the Bagging ensemble models, and the error bars indicate the standard deviation among multiple parallel base models.

验证结果如图8(B)所示，物理与数据双驱动的FEPaML模型预测值与FEP计算的结合亲和力ΔΔ*G*在正负的方向性与数值的量级上均高度一致. 对于单点突变，模型展现了极高的预测精度. 其中，排名靠前的E8S与K9R突变，FEPaML预测值分别为−15. 19和−14. 81 kJ/mol，FEP计算值分别为

−15.52和−15.15 kJ/mol，二者预测值与计算值的平均绝对偏差均为0.33 kJ/mol左右，证明模型能够精准捕捉侧链优化的热力学特征．在多突变层面，模型能够区分位点间的非线性组合效应，以最佳双突变组合 E8T-K9R（8T9R）为例，FEPaML预测其结合能为−18.66 kJ/mol，FEP验证值为−20.71 kJ/mol，两者平均绝对偏差为2.05 kJ/mol. 在更加复杂的三突变空间中，模型的筛选能力依然较稳定．其中，模型推荐的优势方案I7F-E8S-K9R（7F8S9R）的$\Delta\Delta G$预测值为−23.56 kJ/mol，与FEP计算得到的−24.43 kJ/mol极为接近，进一步说明模型能够在多位点突变空间中识别具有潜在优势的候选组合．

为进一步评估校正模型的必要性，比较了第0轮未更新的FEPaML模型与第2轮交互校正后的FEPaML模型对于上述15条候选肽的预测表现．如图8（B）所示，第0轮FEPaML模型的预测值多数分布在接近0或正值区间，与FEP验证结果在方向性和数值尺度上均存在明显偏差；经过两轮FEP反馈交互校正后，FEPaML预测值整体转移至负$\Delta\Delta G$区间，与FEP验证值在亲和力变化方向上更加一致．该结果说明，FEP反馈能够有效修正模型的预测尺度和方向偏差，提高其在$PRAME_{190—198}$肽段与HLA-A*11∶01复合物体系中的适用性．

需要注意的是，随着突变数量由单突变增加至双突变和三突变，FEP计算值与FEPaML预测值的波动均有所增加，且图8（B）中部分多突变体的FEPaML预测值较FEP验证值更负，提示模型在该体系下的复杂多突变组合空间中可能存在一定的偏负预测趋势．这一现象可能与模型预测的外推不确定性有关：由于模型校准数据来自丙氨酸扫描、关键位点单突变全扫描及少量补充突变，当预测双突变或三突变体时，模型可能将多个单突变带来的有利自由能信号进行近似叠加或放大，从而预测出更强的结合自由能降低．同时，在真实抗原肽-HLA复合物中，多位点突变会同时改变多个残基的侧链体积、电荷性质和局部空间分布，可能诱导抗原肽-HLA结合界面发生局部构象重排、空间位阻变化、相互作用网络重构及残基间非线性耦合效应．上述因素会增加FEP计算中的构象采样难度，也会提高机器学习模型预测外推的不确定性．因此，采用多重复FEP计算和Bagging集成预测分别评估物理计算和模型预测的稳定性．总体来看，尽管多突变体的不确定性有所增加，FEPaML预测值与FEP验证值在方向性和数值量级上仍保持较好一致．

综上所述，第2轮的机器学习预测与物理验证结果高度吻合，有力证明了以精准的FEP计算值为物理依据、利用机器学习进行大规模突变组合高效筛选的协同优化体系，能够在保持物理可信度的同时，大幅提升黑色素瘤肿瘤抗原肽HLA-A*11∶01体系的优化效率与设计精准度．

2.5 FEPaML协同优化方法的适用性与展望

2.5.1 *本策略与现有方法的比较* 当前传统基于序列的亲和力预测模型，如NetMHCpan[19,20,65]，依赖大规模粗粒度公共数据库进行训练，适合对候选抗原肽进行快速初筛；然而，在同一抗原肽背景下进一步区分高度相似突变体的相对结合自由能差异时，仅依赖通用序列特征往往难以精准捕捉模拟表位与野生型肽段间的微小亲和力差异．FEP计算能够提供高精度的结合能数据，然而受限于计算成本，难以进行大规模突变组合的全扫描[18]．本研究提出的协同优化框架将基于物理的方法与机器学习模型进行多轮交互、反馈，以少量的FEP数据作为物理依据，校正机器学习模型的预测偏差，显著提升了模型在局部突变空间的预测精度．具体而言，首先，利用分子动力学模拟方法确定稳定结构，通过丙氨酸扫描识别核心突变位点，并进行对应点位的单突变全扫描计算以及补充突变．接着，将上述少量的、高质量的精确FEP结果逐轮输入到模型中，作为机器学习模型的高质量约束；其中，在FEPaML和FEP交互方面，在第0轮和第1轮交互中逐步缩小搜索空间并校正预测偏差，第2轮则实现对单突变、双突变与三突变结合亲和力的可靠遍历预测与排序．这种逐轮校准的策略有效平衡了计算精度与效率，说明物理知识对机器学习的约束与再训练可能是提升抗原肽-HLA结合亲和力预测准确度的关键[25]．

从方法定位上看，本文提出的策略主要指FEP计算的循环反馈与功能互补，而非在神经网络结构内部嵌入物理模块．从样本效率角度，若采用纯FEP方法对$PRAME_{190\sim198}$肽段的单突变、双突变和三突变组合空间进行穷举验证，理论上需完成589323组FEP计算．相比之下，本研究实际使用75组FEP数据进行模型体系特异性校准，并额外使用15组FEP数据进行候选突变体验证，大幅降低了计算成本．

另一方面，相较于纯机器学习方法，本策略虽然增加了少量FEP计算任务，但这些体系特异性高精度物理数据可显著提高模型对相似突变体和多突变组合的局部分辨能力，增强了模型在$PRAME_{190—198}$-HLA-A*11∶01局部突变空间中的物理约束和候选排序可靠性.

因此，相较于单纯的机器学习方法，本策略通过多轮校准与物理反馈机制，降低模型在特定体系中的预测偏差；而相较于单纯FEP计算方法，我们借助机器学习模型对组合空间进行先验分层与优先级分配，在双突变、三突变组合空间中实现了快速、低成本的广域搜索. 本研究通过部分体系特异性数据与较小的FEP计算成本获得有效的物理约束，提升模型对特定抗原肽-HLA体系内候选突变体筛选和排序的可靠性，为解决免疫组学中数据多样性不足与高精度预测需求之间的矛盾提供了可行路径.

2.5.2 *亲和力优化的免疫学意义与局限性* 以$PRAME_{190—198}$-HLA-A*11∶01体系为优化实例，验证了物理计算在个体化肿瘤疫苗设计中的初步指导作用. 基于抗原肽侧链和HLA的结构与丙氨酸扫描分析，在热力学层面识别出具有显著可塑性的位点；接着，结合机器学习的大规模筛选与物理计算的验证，找到了能够有效改善相对结合自由能的突变组合. 本研究基于FEP计算，主要探讨了不同突变带来的能量收益，其中，突变体结合自由能呈现降低趋势，说明抗原肽与HLA结合亲和力在理论上得到了显著提升，为后续探索其微观适配机理的相关实验提供了物理依据[66]. 从免疫学机制角度，抗原肽-HLA复合物的热力学稳定性直接影响抗原在细胞表面的呈递，进而决定$CD8^+$ T细胞的激活效率[67]. 本研究选择的PRAME抗原在黑色素瘤中表达广泛且具有组织特异性，是理想的疫苗靶点[28,29]；结合HLA-A*11∶01在东亚人群中的高频率分布，本策略具备较强的临床应用潜力[58]. 本文主要从计算层面验证候选突变肽对HLA结合亲和力的改善，尚未开展抗原肽-HLA结合实验或稳定性实验. 因此，筛选得到的I7F-E8S-K9R等候选突变体应被理解为具有较高HLA结合潜力的计算候选肽，而不能直接等同于已被实验证实具有增强免疫原性的疫苗候选物. 尽管如此，已有研究表明，基于分子动力学模拟及FEP的相对结合自由能计算有助于为抗原肽或模拟表位的突变优化提供有效的物理依据，并可作为实验验证前候选突变肽优先级排序的重要计算工具[38,39,49].

抗原肽-HLA复合物的热力学稳定性是抗原递呈和T细胞识别的重要前提，但真实免疫原性仍受到TCR特异性识别、抗原加工与递呈效率、共刺激信号、细胞因子释放及免疫微环境等多重因素共同调控[12,68~70]. 此外，部分突变虽然可能增强抗原肽与HLA结合槽之间的相互作用，但也可能改变肽段暴露区域的局部构象，尤其是TCR接触面残基的空间取向，从而影响TCR对抗原肽-HLA复合物的识别[69]. 基于此，后续研究仍需结合TCR-抗原肽-HLA复合物结构建模、抗原肽-HLA稳定性检测、T细胞激活实验和细胞毒性实验等，系统评估候选肽段是否能够真正诱导有效的抗肿瘤免疫应答.

2.5.3 *未来研究方向与拓展应用* 整合物理驱动与数据驱动的策略具有广阔的拓展潜力. 关于模型的跨体系适用性，FEPaML的模型研究已在p53R175H、NY-ESO-1和WT1等多个代表性抗原体系中进行了验证. 结果显示，经过约25组FEP采样数据微调后，不同体系中的预测精度均可提升至0.9以上[24]. 本研究进一步基于$PRAME_{190—198}$-HLA-A*11∶01的晶体结构，开展分子动力学模拟与FEP计算，使模型逐步适应该抗原肽-HLA体系的局部物理规律. 已有研究表明，对于结构准备充分、力场参数合理且采样收敛性较好的蛋白-配体结构，FEP对相对结合自由能的预测精度可接近实验亲和力测定的可重复性范围，并已在前瞻性候选分子筛选中显示出较好的优先级排序能力[71,72]. 基于此，本优化策略在不同抗原肽-HLA体系间具有一定的迁移潜力，即利用少量体系特异性FEP数据对模型进行校准，使其适应该体系的局部结合自由能特征；但其预测性能可能受到目标抗原和HLA亚型差异的影响. 未来仍需在更多肿瘤抗原、HLA等位基因及实验体系中进一步验证该策略的适用范围.

在肿瘤疫苗开发的后续研究中，将继续深化该策略.（1）构建涵盖TCR识别环节的三元复合物预测模型. 将构建TCR-抗原肽-HLA相互作用的专用数据集，用以支撑高精度的泛化预测；将针对特定TCR克隆应用该策略，通过少量高精度的三元复合物FEP计算校正机器学习模型，实现对特定患者TCR库中T细胞激活潜力的精准评估[73].（2）为解决TCR-抗原肽-HLA界面的高度动态性和构象多样

性的挑战，将发展多模态深度学习框架，整合序列、结构和动力学信息，以突破静态结构分析的局限. 将整合分子动力学模拟轨迹中的动态特征作为模型输入，提取残基间距离分布、氢键形成概率、溶剂可及性变化等动态特征，同时引入FEP自由能等物理量作为损失函数，确保模型预测具备物理合理性[74]. (3) 将本研究策略扩展至其它肿瘤类型与HLA等位基因. 本研究聚焦黑色素瘤相关的癌-睾丸抗原PRAME，未来可将该策略应用于体细胞突变产生的特异性抗原. 此外，针对不同HLA等位基因，如HLA-A*02∶01、HLA-A*24∶02等高频等位基因，建立相应的优化流程，提升肿瘤疫苗设计的人群覆盖率[58].

3 结　论

提出了一种基于自由能微扰计算与机器学习反馈校准的黑色素瘤抗原肽协同优化策略，以$PRAME_{190—198}$-HLA-A*11∶01体系为例验证了其可行性. 通过分子动力学模拟、埋藏表面积分析、FEP计算及FEPaML多轮校准，识别了P4，P6和P8等具有优化潜力的关键位点，并将75组高精度FEP数据逐轮反馈至模型，实现了对单突变、双突变和三突变组合空间的高通量预测与候选排序. 结果显示，第2轮FEPaML预测与FEP验证结果具有较好一致性，其中，单突变E8S和K9R的平均绝对偏差仅0.33 kJ/mol，最优三突变体I7F-E8S-K9R的$\Delta\Delta G$预测值为−23.56 kJ/mol，与FEP计算值−24.43 kJ/mol接近，提示其HLA结合亲和力较野生型明显提升. 与单纯依赖序列特征的预测方法相比，该策略通过引入体系特异性的FEP物理数据，增强了模型在局部突变空间中的分辨能力和候选排序可靠性. 本研究为黑色素瘤抗原肽的计算优化提供了一种兼具物理依据与筛选效率的设计框架，筛选得到的突变肽可为后续HLA结合实验、抗原递呈稳定性检测及免疫功能验证提供优先候选.

支持信息见http://www.cjcu.jlu.edu.cn/CN/10.7503/cjcu20260155.

完整FEP计算数据以及FEPaML第0~2轮预测结果已上传至GitHub：https://github.com/leifugroup/PRAME190-198-HLA-A1101-FEP-ML-Prediction-Data.git.

参　考　文　献

[1] Hamid O., Robert C., Daud A., Hodi F. S., Hwu W. J., Kefford R., Wolchok J. D., Hersey P., Joseph R., Weber J. S., Dronca R., Mitchell T. C., Patnaik A., Zarour H. M., Joshua A. M., Zhao Q., Jensen E., Ahsan S., Ibrahim N., Ribas A., *Ann. Oncol.*, **2019**, *30*(4), 582—588
[2] Larkin J., Chiarion-Sileni V., Gonzalez R., Grob J. J., Rutkowski P., Lao C. D., Cowey C. L., Schadendorf D., Wagstaff J., Dummer R., Ferrucci P. F., Smylie M., Hogg D., Hill A., Márquez-Rodas I., Haanen J., Guidoboni M., Maio M., Schöffski P., Carlino M. S., Lebbé C., McArthur G., Ascierto P. A., Daniels G. A., Long G. V., Bastholt L., Rizzo J. I., Balogh A., Moshyk A., Hodi F. S., Wolchok J. D., *New Engl. J. Med.*, **2019**, *381*(16), 1535—1546
[3] Tian H. C., Zhang J. R., Si L., *J. Pract. Oncol.*, **2023**, *38*(6), 509—512(田惠春, 张佳冉, 斯璐. 实用肿瘤杂志, **2023**, *38*(6), 509—512)
[4] Siegel R. L., Kratzer T. B., Giaquinto A. N., Sung H., Jemal A., *CA Cancer J. Clin.*, **2025**, *75*(1), 10—45
[5] Topalian Suzanne L., Drake Charles G., Pardoll Drew M., *Cancer Cell*, **2015**, *27*(4), 450—461
[6] Sanmamed M. F., Chen L., *Cell*, **2018**, *175*(2), 313—326
[7] Li W., Hu J. L., Wang K., He Y. J., *Chin. J. Clin. Pharmacol. Ther.*, **2021**, *26*(9), 1053-1064(李玮, 胡佳丽, 王凯, 贺毅憬. 中国临床药理学与治疗学, **2021**, *26*(9), 1053—1064)
[8] Marty R., Kaabinejadian S., Rossell D., Slifker M. J., van de Haar J., Engin H. B., de Prisco N., Ideker T., Hildebrand W. H., Font-Burgada J., Carter H., *Cell*, **2017**, *171*(6), 1272—1283
[9] Singhaviranon S., Dempsey J. P., Hagymasi A. T., Mandoiu I. I., Srivastava P. K., *Nat. Immunol.*, **2025**, *26*(2), 240—251
[10] Blass E., Ott P. A., *Nat. Rev. Clin. Oncol.*, **2021**, *18*(4), 215—229
[11] Hu T. T., Qiu X. D., Zhang Y., Li K. G., Du J., *J. Mod. Oncol.*, **2021**, *29*(19), 3495—3499(胡天天, 邱戌旦, 张嬿, 李凯刚, 杜娟. 现代肿瘤医学, **2021**, *29*(19), 3495—3499)
[12] Chowell D., Morris L. G. T., Grigg C. M., Weber J. K., Samstein R. M., Makarov V., Kuo F., Kendall S. M., Requena D., Riaz N., Greenbaum B., Carroll J., Garon E., Hyman D. M., Zehir A., Solit D., Berger M., Zhou R., Rizvi N. A., Chan T. A., *Science*, **2018**, *359*(6375), 582—587

[13] Matsumura M., Fremont D. H., Peterson P. A., Wilson I. A., *Science*, **1992**, *257*(5072), 927—934
[14] Givechian K. B., Rocha J. F., Liu C., Yang E., Tyagi S., Greene K., Ying R., Caron E., Iwasaki A., Krishnaswamy S., *Nat. Mach. Intell.*, **2025**, *8*, 70—83
[15] Paul S., Grifoni A., Peters B., Sette A., *Front. Immunol.*, **2020**, *10*, 3151
[16] Phillips J. C., Hardy D. J., Maia J. D. C., Stone J. E., Ribeiro J. V., Bernardi R. C., Buch R., Fiorin G., Hénin J., Jiang W., McGreevy R., Melo M. C. R., Radak B. K., Skeel R. D., Singharoy A., Wang Y., Roux B., Aksimentiev A., Luthey-Schulten Z., Kalé L. V., Schulten K., Chipot C., Tajkhorshid E., *J. Chem. Phys.*, **2020**, *153*(4), 044130
[17] Jorgensen W. L., *Acc. Chem. Res.*, **1989**, *22*(5), 184—189
[18] Meng Y., Dashti D. S., Roitberg A. E., *J. Chem. Theory Comput.*, **2011**, *7*(9), 2721—2727
[19] Reynisson B., Alvarez B., Paul S., Peters B., Nielsen M., *Nucleic Acids Res.*, **2020**, *48*(W1), W449—W454
[20] Nilsson J. B., Greenbaum J., Peters B., Nielsen M., *Front. Immunol.*, **2025**, *16*, 1616113
[21] O'Donnell T. J., Rubinsteyn A., Laserson U., *Cell Systems*, **2020**, *11*(1), 42—48
[22] Zhang C., Li M., *J. Sci. Teachers Coll. Univ.*, **2021**, *41*(7), 33—36, 47(张驰, 黎明. 高师理科学刊, **2021**, *41*(7), 33—36, 47)
[23] Lewis S., Hempel T., Jiménez-Luna J., Gastegger M., Xie Y., Foong A. Y. K., Satorras V. G., Abdin O., Veeling B. S., Zaporozhets I., Chen Y., Yang S., Foster A. E., Schneuing A., Nigam J., Barbero F., Stimper V., Campbell A., Yim J., Lienen M., Shi Y., Zheng S., Schulz H., Munir U., Sordillo R., Tomioka R., Clementi C., Noé F., *Science*, **2025**, *389*(6761), eadv9817
[24] Zhong Q., Chan K. C., Fu L., Zhou R., *Briefings Bioinf.*, **2025**, *26*(4), bbaf254
[25] Noé F., Tkatchenko A., Müller K. R., Clementi C., *Annu. Rev. Phys. Chem.*, **2020**, *71*, 361—390
[26] Zhang H., Gao Y. Q., *Chin. J. Chem. Phys.*, **2024**, *37*(2), 177—191
[27] Gonzalez-Galarza F. F., McCabe A., Santos E., Jones J., Takeshita L., OrtegaRivera N. D., Cid-Pavon G. M. D., Ramsbottom K., Ghattaoraya G., Alfirevic A., Middleton D., Jones A. R., *Nucleic Acids Res.*, **2020**, *48*(D1), D783—D788
[28] Epping M. T., Wang L., Edel M. J., Carlée L., Hernandez M., Bernards R., *Cell*, **2005**, *122*(6), 835—847
[29] Zhao C. L., Yang C., *Med. J. Chinese People's Armed Police Forces*, **2016**, *27*(1), 92—96(赵铖龙, 杨诚. 武警医学, **2016**, *27*(1), 92—96)
[30] Ikeda H., Lethé B., Lehmann F., van Baren N., Baurain J. F., de Smet C., Chambost H., Vitale M., Moretta A., Boon T., Coulie P. G., *Immunity*, **1997**, *6*(2), 199—208
[31] Abraham M. J., Murtola T., Schulz R., Páll S., Smith J. C., Hess B., Lindahl E., *SoftwareX*, **2015**, *1/2*, 19—25
[32] Feng S. L., Hu S., Liu B., Liu W., *Acta Chim. Sinica*, **2013**, *71*(9), 1313—1320(冯石磊, 胡墅, 刘兵, 刘伟. 化学学报, **2013**, *71*(9), 1313—1320)
[33] Best R. B., Zhu X., Shim J., Lopes P. E. M., Mittal J., Feig M., MacKerell A. D. Jr., *J. Chem. Theory Comput.*, **2012**, *8*(9), 3257—3273
[34] Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L., *J. Chem. Phys.*, **1983**, *79*(2), 926—935
[35] Bussi G., Donadio D., Parrinello M., *J. Chem. Phys.*, **2007**, *126*(1), 014101
[36] Fu L., Chan K. C., Zhong Q., Buratto D., Wu S., Zhou R., *J. Phys. Chem. B*, **2025**, *129*(50), 12799—12810
[37] Zhou H., Chan K. C., Buratto D., Zhou R., *Int. J. Biol. Macromol.*, **2023**, *253*, 127199
[38] Ahmed R., Omidian Z., Giwa A., Cornwell B., Majety N., Bell D. R., Lee S., Zhang H., Michels A., Desiderio S., Sadegh-Nasseri S., Rabb H., Gritsch S., Suva M. L., Cahan P., Zhou R., Jie C., Donner T., Hamad A. R. A., *Cell*, **2019**, *177*(6), 1583—1599
[39] Song Y., Bell D. R., Ahmed R., Chan K. C., Lee S., Hamad A. R. A., Zhou R., *Proc. Natl. Acad. Sci. USA*, **2023**, *120*(16), e2214430120
[40] Parrinello M., Rahman A., *J. Appl. Phys.*, **1981**, *52*(12), 7182—7190
[41] Hess B., Bekker H., Berendsen H. J. C., Fraaije J. G. E. M., *J. Comput. Chem.*, **1997**, *18*(12), 1463—1472
[42] Darden T., York D., Pedersen L., *J. Chem. Phys.*, **1993**, *98*(12), 10089—10092
[43] Zwanzig R. W., *J. Chem. Phys.*, **1954**, *22*(8), 1420—1426
[44] Boresch S., Tettinger F., Leitgeb M., Karplus M., *J. Phys. Chem. B*, **2003**, *107*(35), 9535—9551
[45] Deng Y., Roux B., *J. Chem. Theory Comput.*, **2006**, *2*(5), 1255—1273
[46] Mobley D. L., Chodera J. D., Dill K. A., *J. Chem. Phys.*, **2006**, *125*(8), 084902
[47] Beutler T. C., Mark A. E., van Schaik R. C., Gerber P. R., van Gunsteren W. F., *Chem. Phys. Lett.*, **1994**, *222*(6), 529—539
[48] Steinbrecher T., Joung I., Case D. A., *J. Comput. Chem.*, **2011**, *32*(15), 3253—3263
[49] Bessell C. A., Isser A., Havel J. J., Lee S., Bell D. R., Hickey J. W., Chaisawangwong W., Glick Bieler J., Srivastava R., Kuo F., Purohit T., Zhou R., Chan T. A., Schneck J. P., *JCI Insight*, **2020**, *5*(8), e135597
[50] Kumar S., Rosenberg J. M., Bouzida D., Swendsen R. H., Kollman P. A., *J. Comput. Chem.*, **1992**, *13*(8), 1011—1021
[51] Petrov D., Perthold J. W., Oostenbrink C., de Groot B. L., Gapsys V., *J. Chem. Theory Comput.*, **2024**, *20*(2), 914—925
[52] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I., *NeurIPS*, **2017**, *30*, 5998—6008
[53] Shahriari B., Swersky K., Wang Z., Adams R. P., de Freitas N., *Proc. IEEE*, **2016**, *104*(1), 148—175
[54] Jones D. R., Schonlau M., Welch W. J., *J. Global. Optim.*, **1998**, *13*(4), 455—492
[55] Yuan F., Sun Y. Y., Liang F., Liu N., Jin L., Luo Y., Tan L. X., Xi Y. Z., *Chin. Med. Biotechnol.*, **2008**, *3*(5), 343—349(袁方, 孙

玉英, 梁飞, 刘楠, 金荔, 骆媛, 谭刘欣, 奚永志. 中国医药生物技术, **2008**, *3*(5), 343—349)
[56] Blount S. L., Liu X., McBride J. D., *Cells*, **2024**, *13*(20), 1740
[57] Karuppiah V., Sangani D., Whaley L., Pengelly R., Uluocak P., Carreira R. J., Hock M., Cristina P. D., Bartasun P., Dobrinic P., Smith N., Barnbrook K., Robinson R. A., Harper S., *Sci. Rep.*, **2024**, *14*(1), 26851
[58] Sidney J., Peters B., Frahm N., Brander C., Sette A., *BMC Immunol.*, **2008**, *9*, 1
[59] Xia Z., Huynh T., Kang S. G., Zhou R., *Biophys. J.*, **2012**, *102*(6), 1453—1461
[60] Bradley P., *eLife*, **2023**, *12*, e82813
[61] Chen J., Sawyer N., Regan L., *Protein Sci.*, **2013**, *22*(4), 510—515
[62] Bogan A. A., Thorn K. S., *J. Mol. Biol.*, **1998**, *280*(1), 1—9
[63] Zhou R., Das P., Royyuru A. K., *J. Phys. Chem. B*, **2008**, *112*(49), 15813—15820
[64] Wang J., Mao J., Li C., Xiang H., Wang X., Wang S., Wang Z., Chen Y., Li Y., No K. T., Song T., Zeng X., *J. Cheminform.*, **2024**, *16*(1), 142
[65] Jurtz V., Paul S., Andreatta M., Marcatili P., Peters B., Nielsen M., *J. Immunol.*, **2017**, *199*(9), 3360—3368
[66] Rossjohn J., Gras S., Miles J. J., Turner S. J., Godfrey D. I., McCluskey J., *Annu. Rev. Immunol.*, **2015**, *33*, 169—200
[67] Harndahl M., Rasmussen M., Roder G., Dalgaard Pedersen I., Sørensen M., Nielsen M., Buus S., *Eur. J. Immunol.*, **2012**, *42*(6), 1405—1416
[68] Stone J. D., Kranz D. M., *Front. Immunol.*, **2013**, *4*, 244
[69] Mariuzza R. A., Wu D., Pierce B. G., *Front. Immunol.*, **2023**, *14*, 1303304
[70] Chen L., Flies D. B., *Nat. Rev. Immunol.*, **2013**, *13*(4), 227—242
[71] Wang L., Wu Y., Deng Y., Kim B., Pierce L., Krilov G., Lupyan D., Robinson S., Dahlgren M. K., Greenwood J., Romero D. L., Masse C., Knight J. L., Steinbrecher T., Beuming T., Damm W., Harder E., Sherman W., Brewer M., Wester R., Murcko M., Frye L., Farid R., Lin T., Mobley D. L., Jorgensen W. L., Berne B. J., Friesner R. A., Abel R., *J. Am. Chem. Soc.*, **2015**, *137*(7), 2695—2703
[72] Ross G. A., Lu C., Scarabelli G., Albanese S. K., Houang E., Abel R., Harder E. D., Wang L., *Commun. Chem.*, **2023**, *6*(1), 222
[73] Chronister W. D., Crinklaw A., Mahajan S., Vita R., Koşaloğlu-Yalçın Z., Yan Z., Greenbaum J. A., Jessen L. E., Nielsen M., Christley S., *Front. Immunol.*, **2021**, *12*, 640725
[74] Karniadakis G. E., Kevrekidis I. G., Lu L., Perdikaris P., Wang S., Yang L., *Nat. Rev. Phys.*, **2021**, *3*(6), 422—440

(Ed.: Y, K, M)