分子对接研究现状

基于多目标差分进化的分子对接算法研究

**摘要：**2015 博士

如今越来越多生物大分子的三维结构得到了精确测定。随着计算机技术的发展，在现代基于结构的药物设计中，计算机辅助药物设计的方法已经得到广泛应用。计算机辅助药物设计不仅可以节约新药研发的成本，而且可以缩短新药研发的时间。在计算机辅助药物设计中，分子对接是其重要组成部分。分子对接是通过计算机模拟，预测两个或多个分子形成稳定复合物的结合模式。分子对接在理解分子生物功能，预测药物毒性，虚拟药物筛选等方面都有广泛应用。在本文中我们主要分析蛋白质与小分子的对接。解决分子对接有两方面内容，一是空间搜索算法，二是打分函数。空间搜索是找出配体小分子在受体蛋白质结合位点附近最优的结合模式。空间搜索需要考虑小分子移动，转动，二面角扭转以及柔性对接中可变残基的空间变化。目前空间搜索有分子动力学模拟，几何匹配以及智能算法。如何快速有效地进行空间搜索是分子对接的研究重点。打分函数作为评价对接模式好坏的标准，在对接过程中起重要作用。打分函数不仅要能正确评估对接结果，还要能区分不同对接结果之间的差异。目前打分函数主要有基于力场的打分函数，基于经验的打分函数和基于知识的打分函数。影响打分函数的因素有范德华相互作用，氢键相互作用以及静电相互作用等。打分函数与空间搜索是分子对接密不可分的两个部分，打分函数作为空间搜索的指引标准，决定了对接结果的好坏，而空间搜索则需要能遍历到全局最优解，同时避免陷入局部最优解。目前有很多分子对接软件可供使用，其中既有免费的开源软件，也有很多商业分子对接软件。虽然目前分子对接软件已经有很多，但每年也持续有新的分子对接软件面世。这说明分子对接软件仍然面临诸多挑战。有研究显示，目前主流的分子对接软件的准确率只有70%左右，仍然有很大的提升空间。分子对接需要解决的问题是提高打分函数的准确性，同时采用更快速准确的空间搜索算法。在本文中，我们提出了一种采用多目标差分进化的分子对接算法。与以往打分函数综合所有分子相互作用得到一个总打分值不同，多目标优化是将不同的分子相互作用项作为多个优化目标，同时优化所有的分子相互作用。一直以来，传统打分函数的难点是如何分配不同能量项的权值，得到一个总打分函数。因为打分函数不可避免地受经验影响，要得到准确的权值是非常困难的。通过多目标优化避免了因权值的分配对对接结果的影响，可以更准确地评价对接结果。我们将分子间范德华相互作用，氢键相互作用，静电相互作用以及小分子内部的范德华相互作用作为分子对接优化的多个目标。针对分子对接问题，我们对这些能量项采用了改进的计算方法，在保证准确性的前提下，提高对接的速度。在空间搜索方面，我们采用差分进化算法进行搜索。作为一种进化优化算法，差分进化采用遗传进化的策略优化目标。但与其他进化算法不同的是，差分进化采用差分向量选取行的个体，速度更快更准确，而且差分进化采用实数编码，比遗传算法等采用二进制编码的进化算法计算量更少，速度更快。在分子对接过程中会产生大量的三维结构，特别是采用进化算法这种群体优化策略会产生很多不同的分子空间结构。我们提出了一种新的几何分类算法，并将其应用到分子对接中。几何分类算法是一种自顶向下递归的分类算法，以数据集合中边缘的个体为种子节点进行分类。种子节点的选取是根据距离矩阵计算能组成最大几何体的顶点。相比其他分类算法，几何分类算法对数据的处理更具有一致性。我们将几何分类算法应用到分子对接过程中和最终结果分析中。在对接过程中，将当前群体结构分类，提取具有代表性的结构，删除冗余的结构，提高了空间搜索的效率。实验结果表明，几何分类算法在分类结构数据中有很好的效果。基于以上对接方法，我们开发了一款具有完整功能的分子对接软件。我们将其与目前主流的分子对接软件在标准测试集上进行了比较，试验结果显示我们的方法可以准确地进行分子对接。

[文章提出了新的算法，同时对进行了算法的实现，程序的编制，以及需要和根据其他算法编制的软件进行对比，这是一个比较难的方面，需要自己编写程序进行对比实验，这里需要专门研究其他相关的软件的使用]

#### 基于群体智能的分子对接算法研究与实现

#### 摘要:

分子对接作为一种药物设计工具,在药物发现领域发挥着越来越重要的作用。它能够充分利用计算机优化算法,实现计算机辅助药物虚拟筛选。分子对接的目标是搜索庞大的对接构象,从而找到自由能最低的构象(也就是最稳定的构象)。分子对接问题包含两个方面：一是打分函数,用来表示分子模型,有效地指引搜索；另一个是搜索算法,对对接产生的庞大构象进行搜索；这样分子对接问题就转化成数学上的优化问题。本文的主要从搜索算法入手,全面分析群体智能算法,将经典的群体智能算法变种引入到分子对接平台(AutoDock)中,构建了用群体智能算法处理分子对接问题的方法体系。之后,本文分别以PSO、 ABC和AFSA的变种为搜索算法,实现了分子对接程序FIPSDock、 ABCDock和AFSADock,并将它们与4个经典的分子对接算法进行了比较,即GOLD, DOCK, FlexX和用LGA实现的AutoDock.本文的对接实验包括同源对接和交叉对接两部分,在多个方面对经典的分子对接软件和本文实现的FIPSDock、 ABCDock、 AFSADock进行了综合比较。实验结果显示,基于群体智能算法的分子对接程序在处理分子对接问题上,相对于基于拉马克遗传算法的AutoDock,有更快的收敛速度,有更加稳定的表现和更高的精度；结果还显示,FIPSDock、 ABCDock和AFSADock在搜索能力方面胜于使用相同力场的AutoDock和SODOCK,在精度上优于四种经典的分子对接算法。

#### 求解蛋白质分子对接问题的遗传算法研究

摘要

蛋白质-配体对接是基于结构的药物开发中最重要的方法之一，也是大规模虚拟筛选的一种重要方法。通过建立蛋白质-配体对接模型，研究蛋白质受体与配体的相互作用，发现和设计一种更有效、更理想的药物具有非常重要的意义。分子对接的过程是寻找带有最小能量值的蛋白质和配体的构象。在蛋白质-配体对接过程中，打分函数和搜索算法是两个重要的组成部分。打分函数是一个用来评价对接构象能量值的力场,它用来评估计算机模拟预测出的配体与受体的结合构象。搜索算法是找出受体和配体之间的最佳结合模式。一些算法已被证明是非常有效的用来求解决蛋白质-配体对接的问题，一些研究人员已经改进了这些方法并提高了这些对接方法的效率。例如，模拟退火算法（simulated annealing，SA）、标准遗传算法（genetic algorithm，GA）、拉马克遗传算法（Lamarckian genetic algorithm，LGA）、SODOCK和人工蜂群算法（artificial bee colony, ABC）。然而，开发更高效、更可靠的搜索算法仍然是分子对接问题的一个挑战。 本文中，提出了三个改进的遗传算法求解蛋白质-配体对接问题。第一个方法是基于交叉精英保留的遗传算法（genetic algorithm with crossover elitist preservation，CEPGA），它利用交叉精英保留机制保存上一代种群精英的个体，以使遗传算法的交叉操作更有效率。第二个是基于历史信息指导的遗传算法（running history information guided genetic algorithm，HIGA），这个方法使用历史信息的指导策略记录所有被评价解的访问位置以防止解的重访问。第三个是基于进化方向指导的遗传算法（population evolution direction guided genetic algorithm，EDGA），这个新算法应用种群进化方向指导策略保证算法的搜索方向不断的得到优化，以使遗传算法得到更优的解。 AutoDock 是一款由美国Scripps 研究所开发的一款分子对接软件。它是一个免费和开放源代码的软件，也是应用最广泛的蛋白质-配体对接程序。这款软件产生网格的结合位点，然后用搜索算法来找到最佳结合的受体和配体的构象，最后用打分函数评价这个构象。AutoDock 的4.2.6版本是本文所采用的实验平台，并且AutoDock 4.2.6中的半经验自由能力场是被应用作为本文实验的打分函数去评价分子对接的构象。 为了证明我们提出的三种改进的遗传算法的优越性。本文通过来自RCSB蛋白质数据库的测试用例（PDB）（http://www.rcsb.org/pdb），将这三种新的方法与GA、LGA、SODOCK和ABC在求解分子对接问题上进行比较分析。结果表明改进后的遗传算法求得的最小值、中值和均值有所降低，从整体上提高了解的质量。

#### 分子对接构象搜索优化策略与算法研究

摘要

21世纪新药研究的热点集中于先导化合物的发掘与设计,其中使用计算机辅助药物设计是先导化合物设计的重要方法之一。由于在先导物发现中的优异表现及高效应用,分子对接在药物研究的诸多方面得到了广泛的运用和推广。在分子对接方法中,首先将配体小分子放在受体的活性位点上,在这个过程中,有很多可选的匹配模式,需要按照能量互补、几何互补、化学环境互补等原则,找出这若干模式中,能量得分最低,也即为最稳定最佳匹配模式。确定对接分子之间的结合强度和最佳的结合位置是分子对接所面临的重要问题,评价函数和构象搜索是分子对接的两个重要研究领域。由于配体和受体分子可以自由的旋转和移动,同时,两个分子构象也在改变,使得可能存在大量的结合模式,无法利用简单的搜索方法来完成这个艰巨的任务,必须引入其他高效的优化方法。目前常用遗传算法及其衍生优化算法来进行分子对接构象的搜索。构象搜索算法的研究及其模型的优化是本文的主要研究内容,经典遗传算法是基础,故本文介绍了分子对接相知识、对接原理之后,对经典遗传算法进行了详细阐述,随后着重于构象搜索算法的研究,并从数学角度出发提出了基于拉马克遗传算法的分子对接模型,通过模拟个体在生命周期内获得的适应度提高以表现型的形式传递到下一代,可以弥补经典遗传算法在局部搜索方面能力的不足。模型中,引入了信息熵、精英保留策略及多种群,加速进化的同时保证了种群的多样性。然而,拉马克学习的不足在于不能对生物个体有效的区分,造成局部搜索资源的浪费,本文对所提出的模型加以改进,对进行拉马克学习的个体进行班德文学习,以挖掘个体的学习潜能,使得拉马克学习能更有针对性。利用学习潜能和适应度一起作为评价个体的指标,亦即给拉马克学习赋予了个体区分的能力。本文的难点在于模型的研究、建立及优化。目前常用的对接软件包括DOCK、AUTODOCK及FlexX等,这些软件各有优劣,AUTODOCK4.2是一个开源软件,本文在其源码的基础上,进行修改,这样可以使得其他参数均一致的情况下,与原AUTODOCK4.2程序实验结果进行对比,得以验证本文所提出的模型确实能有效运用于分子对接中。

#### 基于适应度的分子对接优化算法评价研究

摘要

分子对接（molecular docking）是指两个分子相互识别、寻找最佳结合模式的过程。分子对接已成为结构生物学和分子生物学的重要组成部分。小分子蛋白质靶点对接（protein-ligand docking）是研究最全面及应用最广的一类对接模型。它是计算机辅助药物设计（computer-aided drug design）的理论基础。在小分子和蛋白质靶点对接领域，全局优化算法用来优化小分子针对蛋白质分子的结合模式。这些优化算法对对接的准确性起着至关重要的作用。目前为止已有数十种优化算法应用到分子对接领域。但是，我们对这些优化算法的性能还没有一个清晰的认识。主要的原因是一些对比研究有时采用了不恰当的评价标准。尽管分子对接的目的是预测两个分子之间的结合模式，我们看似可以直接用结合模式准确性来评价不同算法的性能。但是结合模式的准确性更多地依赖于采用的打分函数（scoring function）而不是所采用的优化算法。所以本文中，为了对优化算法的性能有一个公正的和清晰的呈现，我们将在各个领域都适用的评价方法——适应度，作为优化算法唯一的评价标准。因为使用这一标准要求不同的优化算法需要具有相同的目标函数，我们采取的方案是将不同的算法移植到同一对接软件。同时为了对算法的性能有一个全面的比较，我们构造的适应度评价方法结合了优化的准确性、鲁棒性和效率。本文我们选取了六个优秀的优化算法进行对比，包括两个基于进化算法的优化算法（LGA，DockDE）和四个基于粒子群的优化算法（SODock，varCPSO,varCPSO-ls，FIPSDock）。因为不同的对接分子对对应着不同的目标函数，并且有些算法对小分子包含旋转键的个数比较敏感，所以我们选取了一个包含多种类别的蛋白质和拥有大跨度旋转键的小分子的复合物集合作为测试数据集。实验结果表明，varCPSO 和 varCPSO-ls具有最快的收敛速度。但因为他们的很难获得到较低的能量值，两种算法的准确性较差。SODock显示出非常棒的准确性和鲁棒性。并且当处理一些类药性分子（小分子旋转键小于10）时，它在优化效率上拥有明显的优势。FIPSDock表现出出众的鲁棒性，但是在优化准确性上距SODock还有一些差距。总的来说，四个基于粒子群的优化算法在各方面都优于两个基于进化的算法，尤其是针对具有较高旋转键的小分子。同时针对各个算法的特点，我们对它们所表现性能的原因做了深入的讨论。最后我们给出了各个算法的对接模式准确性结果。SODock虽然具有非常高的适应度准确性，但它的对接模式准确性却不如varCPSO、varCPSO-ls和FIPSDock。我们提取SODock对接模式较差而其他三种基于PSO的算法（varCPSO、varCPSO-ls和FIPSDock）对接模式较好的结果后发现， SODock在这些实例中都全部获得了最低的能量。所以结果再次说明了，相比对接模式准确性，基于适应度准确性的优化算法评价方法更合理。我们基于适应度的评价方法对其它对接优化算法的评价具有一定的参考价值。