

使用遗传算法优化登革热病毒信封蛋白的肽抑制剂

贺子豪^{1a,2,3}, 赵俊强^{1b,4}, 杨敏^{1a,2,3}, 张松杰^{1a,2,3}, 徐永涛^{1a,2,3}

(1.新乡医学院 a.生物医学工程学院;b.三全学院,河南 新乡 453000;

2.新乡市生物医学信息研究重点实验室,河南 新乡 453000;3.临床与生物医学大数据融合技术河南省工程实验室,河南 新乡 453000;4.新乡市智能医学虚拟重点实验室,河南 新乡 453000)

摘要:登革热病毒严重威胁了人类的健康.目前仍无有效的治疗措施.膜融合是登革热病毒入侵细胞的第一步.信封蛋白的构象变化在膜融合过程中起到关键作用.研发针对信封蛋白的肽抑制剂为治疗登革热提供了潜在的重要靶点.设计了优化针对信封蛋白的肽抑制剂的遗传算法,分别使用 3 种不同的选择算子(轮盘法、锦标赛法和并行法),对初始父代进行 300 代的优化.通过比较发现并行法的结果较好,命名最优的结果为 DN58ga.与此同时还获得了一些多肽衍生物.通过计算验证,这些多肽比父代表现出与登革热病毒信封蛋白更好的结合能力.这些新的肽抑制剂有可能打断融合过程中的蛋白-蛋白相互作用,并有望作为开发登革热病毒信封蛋白新型抑制剂的起点.

关键词:登革热病毒;膜融合;信封蛋白;肽抑制剂;遗传算法

中图分类号:R318.04

文献标志码:A

登革热长期威胁人类健康.目前全球每年大约有 0.5~1 亿例登革热病毒感染的案例,分布在 80 多个国家.登革热病毒是黄病毒属的主要成员,通过蚊媒传播.目前针对该病毒感染没有特效疗法.这主要是因为该病毒有 4 种不同的血清型,疫苗很难被研发,且易出现交叉感染^[1].该病毒引起的登革出血热、登革休克综合征等严重威胁着人类的生命健康.登革热病毒在入侵人体细胞时,信封蛋白会在酸性条件下的核内体中发生一系列的构象变化,导致病毒包膜与细胞膜融合,将登革热病毒 RNA 导入宿主细胞,从而感染细胞^[2].虽然膜融合的机制尚不明确,但针对病毒信封蛋白设计的肽抑制剂阻断膜融合可以使宿主细胞不受感染.这种与原配体竞争型的肽抑制剂已经被验证是有效的,其中比较成功的例子就是针对 HIV 病毒的恩夫韦肽(T-20),已被 FDA 批准可用于治疗艾滋病^[3].目前主要使用的登革热病毒抑制剂是针对病毒 RNA 复制的病毒多聚酶抑制剂,而最具有吸引力的抗病毒治疗策略是病毒进入阻断剂的研发.这种阻断剂主要分为 2 类:1)针对信封蛋白的抑制剂;2)针对宿主细胞上登革病毒受体的抑制剂^[4].本文研究的是针对信封蛋白的抑制剂.徐永涛等通过计算方法设计了针对登革热病毒信封蛋白的肽抑制剂^[5],其中 DS36wt 和 DN58wt 是直接来自信封蛋白上截取的序列,DS36opt 和 DN58opt 是分别通过对 DS36wt 和 DN58wt 进行蒙特卡洛优化而得到的.信封蛋白在膜融合过程中,会发生构象变化形成三聚体发卡结构,此结构对膜融合有重要意义,如图 1 所示.从信封蛋白自身衍生的抑制剂可以影响三聚体发卡结构的形成,打断膜融合过程^[6].目前通过色氨酸荧光猝灭实验^[7]和构象能量分析^[8]都已表明 DN58opt 与 E 蛋白有良好的结合能力,有望阻止膜融合过程,成为有效的肽抑制剂.

本文进一步优化了 DN58opt,使其与信封蛋白具有更好的结合力,从而影响信封蛋白的构象变化,阻止病毒感染细胞.挑选通过蒙特卡洛算法衍生出的多肽作为父代,用遗传算法优化.使用了 3 种不同的选择算子分别对 DN58opt 进行优化,并将结果与 DN58opt 进行比较,发现结合能力有一定提高.这种启发式的算法

收稿日期:2019-01-29;修回日期:2019-05-10.

基金项目:国家自然科学基金(201603180);中国国家留学基金委(201708410492);新乡医学院科研启动经费(505060;505144);新乡医学院生物医学工程学院开放课题(2018-BME-KFKT-01;2018-BME-KFKT-10;2018-BME-KFKT-11);河南省高等学校青年骨干教师培养计划(2017GGJS219).

作者简介:贺子豪(1995-),男,河南安阳人,新乡医学院硕士研究生,研究方向为生物信息学,E-mail: hezihao311@163.com.

通信作者:徐永涛(1982-),男,副教授,研究方向为生物信息学,E-mail: yxu@xxmu.edu.cn.

相比之前的使用人工神经网络从蛋白库中筛选高活性的多肽具有耗时短、效率高等优点^[9].此方法得出的多肽可以对实验提供一定的指导,为针对登革热病毒信封蛋白的肽抑制剂的设计提供了新的方向.

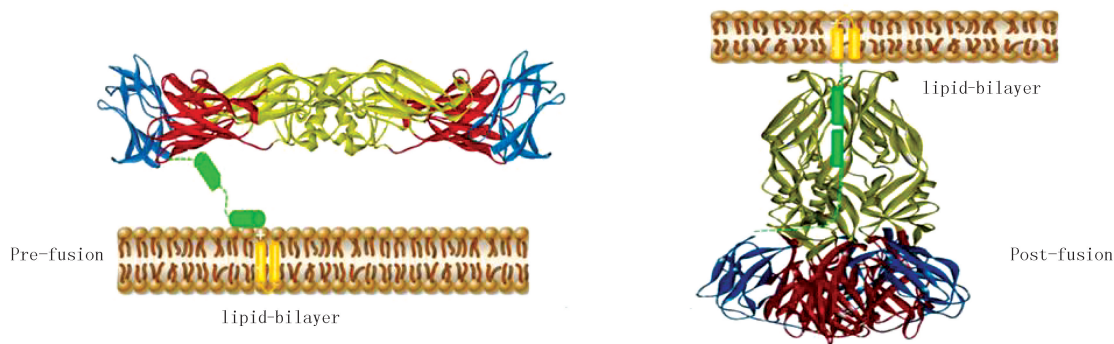


图 1 登革热病毒 E 蛋白膜融合前后构象变化

Fig.1 Conformational changes of dengue virus E protein membrane before and after fusion

1 材料与方法

1.1 材料

3 种选择算子的遗传算法脚本编写采用 SHELL 语言,使用 32 核计算机在 Linux 环境下运行.登革热病毒信封蛋白使用 PDB code:10AN 的晶体结构.使用 AMBER14 软件中的 sander 模块进行能量最小化^[10].使用 scwrl4 软件进行氨基酸替换^[11].蛋白稳定性测评通过 residue-specific all-atom probability discriminatory function(RAPDF) 打分评估^[12],得分越低说明结合的越稳定.多肽序列重复度图谱由 Clustal Omega 得出^[13].20 个父代如表 1 所示,其中得分最低(结合能力最强)的是 DN58opt(-34.99).

表 1 遗传算法父代及其 RAPDF 得分

Tab.1 Parents of Genetic Algorithm and its RAPDF score

氨基酸序列	RAPDF 得分	氨基酸序列	RAPDF 得分
DN58opt: TWWCFYFCRRHHPFWFFYRHN	-34.99	QAWVIVLCHNHNRLRLQWQ	-34.66
EHWWFWFCHHDHHLVFIHCHH	-34.90	KKWMLIICHHHRTVYIRIEQE	-34.60
NLWCFWFINREREYWFIYQHE	-34.84	KQWVIVLCHNRRNRLRLQWQ	-34.60
DWVLLLLHTTHHPFWFFWRYSR	-34.81	DYYIFIFSHRERMLMFVIEHR	-34.59
KYWIFIFCHPERMLMFVIEHR	-34.80	KYWIFIFCHPENMLYFKIEDP	-34.57
RYWVIVLCHNHNRLRLQWQ	-34.78	KIWVIVLCHNRRNRLRLQWQ	-34.57
KIWVIFSHRERMLMFVIEHR	-34.77	RKWMLIFCHHHDTVYIKIEDV	-34.56
KYWVIVLCHNHNRLRLQWQ	-34.75	RRWMFIFCHPHDMVYIKIEDP	-34.56
FAWVIVLCHNHNRLRLQWQ	-34.70	KYWVIVLCHNDNRNRLRLQWQ	-34.56
KYWIFIFCHPERMLYFYIEDP	-34.69	KKWMLIICHHHRTVYIKIEQK	-34.55

1.2 方法

1.2.1 遗传算法思路

遗传算法(GA)不同于一般的传统优化算法,是通过模拟自然界遗传机制和生物进化论从而解决寻优问题,具有自组织,自适应,自学习的特点,以及较强的鲁棒性^[14].本文的目的是优化由 21 个氨基酸组成的肽抑制剂 DN58opt.当遗传算法应用于多肽的优化时,从生物学角度可以找到良好的对应关系.如染色体相当于多肽,基因型代表多肽的序列,基因代表氨基酸,基因在染色体中的位置相当于氨基酸在多肽中的位置,生物的表现型就相当于因此多肽而生成的蛋白质结构^[15].这充分说明了使用 GA 优化多肽的合理性.GA 流程如下:1)设置种群规模 N ,交叉概率(P_c),突变概率(P_m),终止条件.2)通过适应度函数计算各个个体的适应度.3)种群进化:a)选择,用选择算子选出 $M/2$ 对父代($M \geq N$);b)交叉,对所选择的 $M/2$ 对父代,每对按交叉概率执行交叉形成 M 个中间个体;c)变异,对 M 个中间个体分别独立依变异概率执行变异,生成 M 个候

选个体;d) 选择子代,从 M 个候选个体中按适应度选出 N 个个体组成新的种群.4)若满足终止条件则终止,否则重复第 3 步.

1.2.2 编码表示

在 GA 寻优问题中,首要解决的就是编码表示问题.传统 GA 一般使用二进制编码,格雷码编码,实数编码.本文采用的是非数值编码^[16],多肽的氨基酸序列作为编码表示.

1.2.3 适应度函数的选择

GA 之所以会一代比一代表现更好,是因为通过适应度函数对每一代的结果进行了筛选.如同自然界中的优胜劣汰,只有表现良好的个体才有资格作为下一代的父代,继续进行 GA.目前有 RAPDF 法^[12]和 dDFIRE 法^[17]2 种评定蛋白质稳定的打分函数.研究表明对于 α 螺旋结构的信封蛋白适用 dDFIRE 法,而 β 折叠结构的信封蛋白更适用 RAPDF 法打分.登革热病毒的信封蛋白属于 β 折叠结构,所以本文采用 RAPDF 打分函数作为适应度函数^[5].RAPDF 是基于条件概率代表原子距离的函数

$$S(\{d_{ab}^{ij}\}) = - \sum \ln \frac{P(d_{ab}^{ij} | C)}{P(d_{ab}^{ij})} \alpha - \ln P(C | \{d_{ab}^{ij}\}),$$

其中, $P(C)$ 为随机挑选的结构属于“正确”集的概率; $P(d_{ab}^{ij} | C)$ 为在正确结构下, a 型 i 原子和 b 型 j 原子之间的距离为给定距离的概率; $P(d_{ab}^{ij})$ 为 a 型 i 原子和 b 型 j 原子之间的距离为给定距离的概率,无论正确与否; $P(C | d_{ab}^{ij})$ 为 a 型 i 原子和 b 型 j 原子之间的距离为给定距离,此结构是正确结构的概率.

本文通过测试信封蛋白的稳定性来评估多肽抑制剂与信封蛋白的结合能力,使用 RAPDF 打分函数对信封蛋白稳定性进行评估.DN58wt 在信封蛋白的 374~402 位置,但 395~402 这部分的序列还没有被破解^[18],所以用 374~394 这段序列表示 DN58wt.不断地将通过遗传算法生成的新的多肽抑制剂在此位置(374~394)进行替换,分数越低就说明由 GA 生成的多肽与信封蛋白结合的越稳定,打断登革热病毒信封蛋白构象变化的可能性越大,从而阻止膜融合.

1.2.4 选择算子

选择算子是根据适应度函数来进行操作的.在选择哪些个体可以进行交叉,突变的过程中起到关键作用.GA 会通过适应度函数选择优秀个体,但好的选择算子也不会一味选取最优个体.那样会减少基因库的多样性,会使结果快速收敛至局部最优解,无法得到全局最优解.本文选取 3 种不同的选择算子进行试验.

1.2.4.1 轮盘法选择算子

选择某个个体的概率是根据该个体的适应度与当前群体中其他成员的适应度的和的比值来确定的.由于 RAPDF 得分越小越稳定,为便于计算,采用取绝对值的方法.将当前种群所有的 RAPDF 值的绝对值相加得到总和.每个个体适应度的绝对值决定了被选择的概率.在进行轮盘法寻优时,设置 $M=40$,可以尽可能保证遗传信息的多样性.因为轮盘法对适应度不高的个体比较友好,所以有可能影响种群整体进化的方向,导致种群整体的适应度很难提高甚至下降.为了避免这种情况的发生,加入了精英主义思想^[16],每一代从上一代的高分个体中选取一些来当新的父代,考虑种群数量,本文只选取最优个体加入新的父代.

1.2.4.2 锦标赛法选择算子

每次从当前种群随机选取 Y 个个体,然后选择 Y 个个体中适应度最高的个体.将这一步骤重复 M 次(M 为每代种群的个体数量).本文设置 $Y=2$,从父代中随机选择 2 个个体,对比各自的适应度,适应度高的个体被选中.这种方法的优点是收敛速度快,性能优秀的那一部分群体被充分挖掘,但是容易陷入局部最优解.锦标赛法对适应度低的个体很不友好, M 设置过高低分的个体完全被淘汰,种群内无法保持信息多样性,所以在此法中设置 $M=30$.由于高分个体极易被选中,因此无需加入精英主义思想.

1.2.4.3 基于种群交流的选择方法

自然界中,同一物种存在多个种群,但通常在地理位置上相隔较远,那么即使某个种群中存在优秀的基因,也会因为距离较远所以无法交换.而并行法选择算子就是使各种群内的优秀基因相互交换.其做法是用 2 个或 2 个以上的选择方法,对整体进行综合选择.本文将轮盘法和锦标赛法相结合,这种方法也叫并行遗传算法.魏等人^[19]在 Schaffer 函数中对比了轮盘法、锦标赛法和并行法,结果表明了并行法的全局搜索能力更

强,结合了轮盘法与锦标赛法各自的优点.并行法的关键是在迭代一定次数后进行一次信息交换,流程如图 2 所示.并行法中包含 2 个种群:轮盘法 $M=40$,加入精英主义思想;锦标赛法 $M=20$.

1)对最优的 20 个父代进行排序并编号,将数据分为奇数组(针对编号)和偶数组(针对编号).2)对于采用轮盘法选择算子的种群,奇数组为父代.对于采用锦标赛选择算子的种群,偶数组为父代.3)随后在 2 种群中分别独立运行.4)设置每 10 代进行一次种群间的信息交换,首先取得各自种群最优的前 10 个子代,将各自子代的前 5 名交换种群,然后取交换后的结果分别作为各自的父代继续运行.5)交换信息 30 次(300 代),终止运行.

1.2.5 交叉算子与变异算子

从选择算子选出的个体,要依据与 P_c 和 P_m 来判断是否进行交叉和变异.本文采用了单点交叉与单点变异的方法^[20].这 3 个算子是遗传算法必不可少的部分,交叉算子体现了全局搜索能力,变异算子体现了局部搜索能力,使用合适算子对获得全局最优解至关重要.已有研究表明用遗传算法优化碱基序列可以获得成功,并建议迭代次数 >50 ^[21].由于多肽中的氨基酸序列比 DNA 中的碱基序列的排列更为复杂,结合实验室现有设备,本文采用设置迭代次数终止遗传算法,设置迭代次数为 300 代.

2 结 果

本文设计了 3 种采用不同选择算子的遗传算法,分别是 1)轮盘法;2)锦标赛法;3)并行法.将 20 个父代分别提交到 3 个不同的算法中,进行 300 代的计算.各个算法的第 300 代得分较优的前 5 个多肽序列如表 2 所示.轮盘法和并行法的第 300 代结果较 DN58opt(-34.99)均有一定提高.而锦标赛法的第 300 代结果中,只有最优的结果与 DN58opt 保持一致,其他结果均没有 DN58opt 得分优秀.其中由并行法得到的最后一代的最优结果在 3 种方法最后一代最优结果中得分最低为 -35.81,相比 DN58opt 与信封蛋白结合得更稳定,有可能成为新的信封蛋白肽抑制剂.

表 2 各选择算子第 300 代遗传算法的序列及 RAPDF 得分

Tab.2 Sequence and RAPDF score of the 300th GA obtained by different select operator

并行法结果	RAPDF 得分	轮盘法结果	RAPDF 得分	锦标赛法结果	RAPDF 得分
EYWCWFCHHHHHFVFFYRHE(DN58ga)	-35.81	EYWCWFCCRREEQFWFFYRHH	-35.64	RYWWFWFCHHDHHLVFIYQHE	-34.99
HYWCWFCHHHHHFCFFYRHP	-35.78	RYWCWFCCRREEQFWFFYRHP	-35.58	EHWWFWFCHHDHHLVFIYIEHR	-34.98
HYWCWFCHHHHHFVFFYRHE	-35.76	EYWCWFCCRREEQFWFFYRHP	-35.56	EHWWFWFCHHDHHLVFRLLQWQ	-34.95
HYWCWFCHHHHHFCFYRHP	-35.74	RYWCWFCCRREEQFWFFRHH	-35.55	KYWIFWFCHHDHHLVFIYQHE	-34.93
ELWCWFCHHHHHFVFFYRHE	-35.73	EYWCWFCCRREGQFWFFYRHP	-35.54	KYWIFIFCHHDHHLVFIIEHR	-34.91

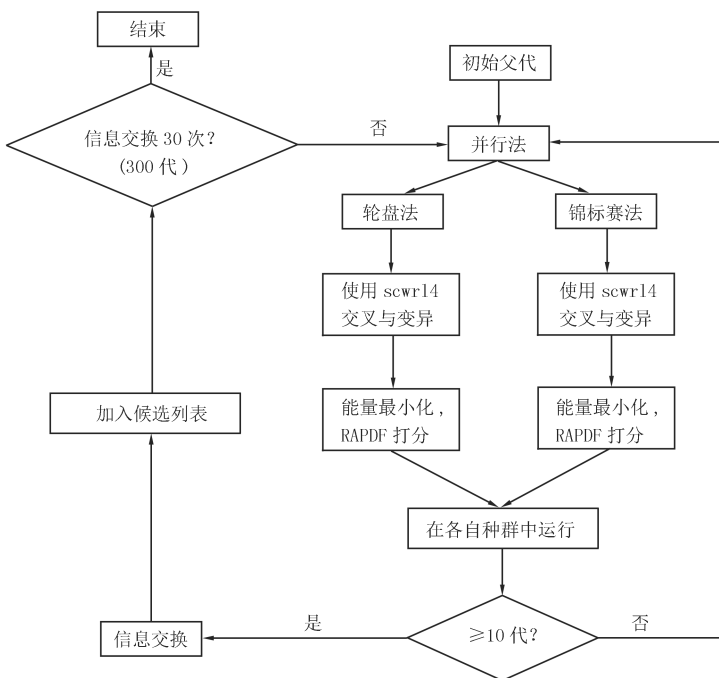


图 2 并行法选择算子的遗传算法流程图

Fig.2 Flow chart of GA for parallel selection operator

除了要求提升活性,计算耗时也是评价计算设计药物方法性能的一个重要指标.在 GA 中,收敛速度决定了计算所需的时间.3 种方法的收敛情况和最终结果如图 3 所示.a)轮盘法在前 60 代过程中,每 10 代均有明显提升,在第 60 代时适应度为 -35.55 ,在此后的 70~170 代适应度只有略微提升,第 170 代的适应度仅降低到 -35.56 ,在第 230 代时出现了轮盘法的最优结果 -35.64 .b)锦标赛法在第 11 代的适应度为 -35.24 ,之后直到 300 代,适应度一直在附近波动,但始终没有低于此结果.c)并行法在第 40 代的适应度为 -35.66 ,最终整体最优结果为 -35.81 .此外,查看了并行法中 2 个种群各自优化的情况,发现,在第 23 次信息交换(第 221~230 代)轮盘法种群中出现了最优值 -35.81 ,随后的第 24 次信息交换(第 231~240 代)锦标赛种群也出现 -35.81 .初始父代与 3 种算法的结果的序列重复度图谱如图 4 所示,色块覆盖区域越广泛表明序列的相似性越高.图 4(a)显示 20 个初始父代的相似程度较低,有较大的信息挖掘价值;而经历了 300 代轮盘法和并行法后的结果(图 4(b)和图 4(d))中有大范围的色块分布,表明序列相似度高;图 4(c)是锦标赛法第 300 代的结果,色块覆盖区域较小,序列相似程度较低.

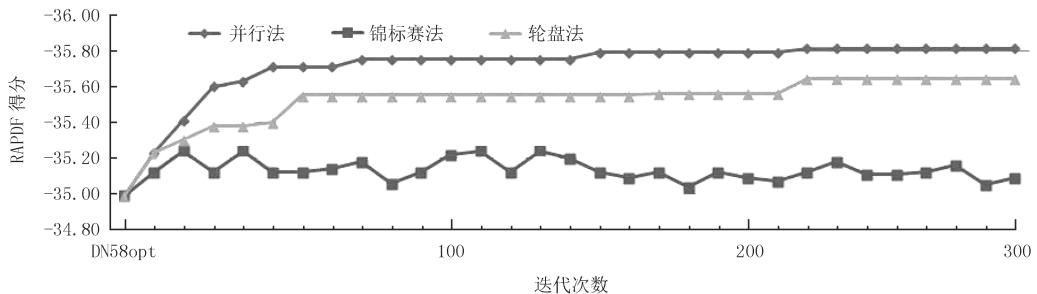


图3 3种不同选择算子 GA 收敛情况

Fig.3 GA convergence of 3 different selection operators

3 讨论

DN58opt 是由 DN58wt 通过蒙特卡洛计算方法优化得到的^[5],并已经得到实验验证^[7-8].这种计算指导实验的抑制剂研发模式,不但加快了研发速度而且节省了大量人力财力,目前国内通过计算设计小分子抑制剂的研究有所增加,但是计算设计肽抑制剂还是非常少见.将计算测试抑制剂活性与遗传算法结合,并采用 3 种选择算子用 SHELL 语言编写脚本,分别进行 300 代遗传算法.现在从收敛速度、最优值、信息挖掘程度对 3 种方法进行分析.

3.1 轮盘法 GA

从图 3 可知,在 300 代遗传算法的前 60 代中,初始提升速度较为缓慢,在 60 代左右陷入局部最优解(-35.55),持续到 170 代左右略微变化至 -35.56 ,最终在第 230 代跳出局部最优解,并收敛到 -35.64 ,即轮盘法最优结果.轮盘法 GA 中,每 10 代的最优结果较 DN58opt 均有提高,虽然过程中陷入局部最优解,但最终没有收敛到局部最优解.这种选择方法不但可以大概率的保存性能优秀的子代,而且适应度低的个体也有机会被选中,保证了种群的多样性,使得 DN58opt 被逐步优化.轮盘法最优序列如表 2 所示,得分为 -35.64 .图 4(b)色块覆盖区域比图 4(a)更加广泛.这说明信息被充分挖掘,同时也表明了优秀的多肽之间有一定相似性.

3.2 锦标赛法 GA

如图 3 所示,在 300 代遗传算法过程中,每 10 代的最优结果波动较大,且始终没有收敛,在 11 代搜索到了局部最优解 -35.24 ,氨基酸序列为 EHWWFWFCHHHPHFWWFWRYYR,对比 DN58opt 只有略微提高,而表 2 中没有 -35.24 的结果.证明在算法过程中由于某次选择淘汰了这个序列,导致接下来陷入局部最优.文献[19]通过增加迭代次数,在 Schaffer 函数测试中,锦标赛法最终也找到了全局最优值.对于函数寻优问题而言,其还有机会跳出局部最优解,但对于本文优化多肽这类非数值编码优化问题,可能由于复杂度的提升导致短时间内无法跳出局部最优解.由于设备、时间等原因设置遗传算法进行 300 代.300 代的锦标赛法结果并不理想.从图 4(c)看出其第 300 代的结果色块分布并不广泛,说明锦标赛法对信息挖掘能力较弱.

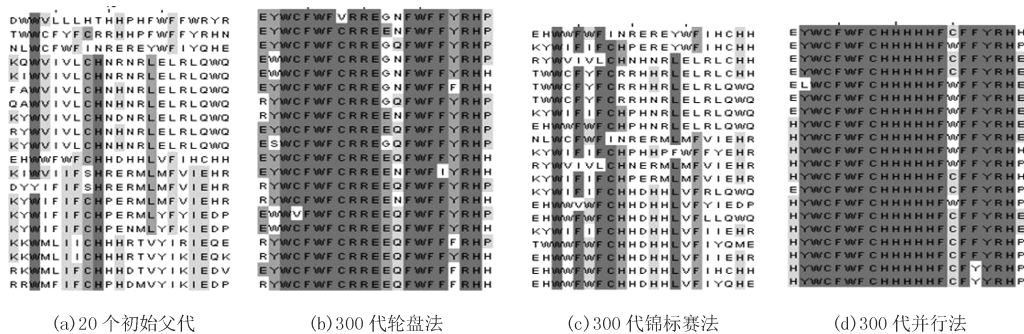


图 4 初始代与使用不同选择算子计算的氨基酸序列相似性对比

Fig. 4 The similarity of amino acid sequences between initial parent and results obtained by different selection operators

3.3 并行法 GA

对于并行法 GA, 设置了探测子种群和开发子种群. 探测子种群用于在优化过程中不断提供新的个体, 克服过早收敛到局部最优. 开发子种群用于局部寻优, 并把好的个体保留下来^[22]. 从上面 2 种选择算子的结果可看出, 轮盘法具有很好的探测能力, 极大地保持了种群信息的多样性, 生成新的个体克服早熟问题, 因此将其设置为探测子种群. 锦标赛法具有存优的功能, 使优秀个体不易流失, 因此将其设置为开发子种群. 如图 3 所示, 每 10 代的最优结果较 DN58opt 均有提高, 最终收敛到 -35.81. 多肽序列如表 2 所示, 将此多肽命名为 DN58 ga, 相比 DN58opt 有较大提高. 图 4(d) 表示第 300 代结果的相似性程度, 色块覆盖了大部分区域, 相似性有极大地提高, 再次说明了优秀的序列有一定的相似性, 且并行法对信息挖掘的能力较强.

综上所述, 发现轮盘法 GA 全局搜索能力较强, 收敛速度较慢, 不易陷入局部最优解. 锦标赛法 GA 全局搜索能力一般, 但局部搜索速度较快, 存优性能较强, 与 Schaffer 函数测试得出的结论一致^[19]. 为了得到针对本系列肽抑制剂寻优更有效的选择算子, 通过分析比较这 2 种算法的结果, 将二者有机结合, 并根据它们的特性分别设置为探测子种群和开发子种群, 编写了并行法选择算子. 从并行法结果来看, 这种算子兼备二者的优点, 收敛速度快, 全局寻优能力强, 得到了本文中最优的结果 DN58ga(-35.81).

在进行轮盘法和并行法遗传算法过程中, 除了通过 3 种选择算子得到的最优结果, 还衍生出来了一系列 RAPDF 得分优秀的多肽, 如表 2 所示. 这些多肽较 DN58opt 活性(计算机测评)均有一定提高, 若经过进一步实验验证, 有可能成为新的登革热病毒肽抑制剂.

4 结 论

本文设计编写了优化登革热病毒 E 蛋白肽抑制剂的遗传算法, 对 DN58opt 进行优化生成了 DN58ga, 将 RAPDF 分数从 -34.99 优化到了 -35.81, 同时生成的一系列衍生物也具有较好的 RAPDF 得分, 但真实活性需要进一步实验验证. 这些发现有望为研制新的登革热病毒的肽抑制剂提供思路. 这种方法相比之前的计算机设计肽抑制剂的手段具有耗时短, 产出多的优点. 因为本文遗传算法模式固定, 也可以应用在优化其他包膜病毒(比如 HIV 病毒, HBV 病毒, 埃博拉病毒, 流感病毒)的信封蛋白肽抑制剂之中.

参 考 文 献

[1] McArthur M A, Sztein M B, Edelman R. Dengue vaccines: recent developments, ongoing challenges and current candidates[J]. Expert Review of Vaccines, 2013, 12(8): 933-953.

[2] Harrison S C. Viral membrane fusion[J]. Nature Structural & Molecular Biology, 2008, 15: 498-507.

[3] Teissier E, Penin F, Pêcheur E I. Targeting cell entry of enveloped viruses as an antiviral strategy[J]. Molecules, 2010, 16(1): 221-250.

[4] 张高红, 郑永唐. 抗登革病毒化学药物研究进展[J]. 中国药理学杂志, 2017(10): 809-813.

[5] Xu Y, Rahman N A B D, Othman R, et al. Computational identification of self inhibitory peptides from envelope proteins[J]. Proteins: Structure, Function, and Bioinformatics, 2012, 80(9): 2154-2168.

[6] Xu Y, Yu S, Zou J W, et al. Identification of peptide inhibitors of enveloped viruses using support vector machine[J]. PloS One, 2015,

- 10(11):e0144171.
- [7] Baharuddin A, Hassan A A, Othman R, et al. Dengue Envelope Domain III-Peptide Binding Analysis via Tryptophan Fluorescence Quenching Assay[J]. Chemical and Pharmaceutical Bulletin, 2014, 62(10):947-955.
- [8] Amir-Hassan A, Lee V S, Baharuddin A, et al. Conformational and energy evaluations of novel peptides binding to dengue virus envelope protein[J]. Journal of Molecular Graphics and Modelling, 2017, 74:273-287.
- [9] Fjell C D, Jessen H, Cheung W A, et al. Optimization of antibacterial peptides by genetic algorithms and cheminformatics[J]. Chemical biology & drug design, 2011, 77(1):48-56.
- [10] Chen S, Zhang X, Li F, et al. Theoretical study of binding affinity for diamidine with DNA[J]. Structural Chemistry, 2016, 27(2):681-696.
- [11] Canutescu A A, Shelenkov A A, Dunbrack R L. A graph-theory algorithm for rapid protein side-chain prediction[J]. Protein Science, 2003, 12(9):2001-2014.
- [12] Shen M Y, Sali A. Statistical potential for assessment and prediction of protein structures[J]. Protein Science, 2006, 15(11):2507-2524.
- [13] Sievers F, Higgins D G. Clustal omega[J]. Curr Protoc Bioinformatics, 2013, 48:1-16.
- [14] 顾玮. 遗传算法概述[J]. 办公自动化, 2016(7):34-35.
- [15] Belda I, Madurga S, Llorà X, et al. ENPDA: an evolutionary structure-based de novo peptide design algorithm[J]. Journal of Computer-Aided Molecular Design, 2005, 19(8):585-601.
- [16] 葛继科, 邱玉辉, 吴春明, 等. 遗传算法研究综述[J]. 计算机应用研究, 2008, 25(10):2911-2916.
- [17] Yang Y, Zhou Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures[J]. Proteins Structure Function & Bioinformatics, 2010, 72(2):793-803.
- [18] Modis Y, Ogata S, Clements D, et al. Structure of the dengue virus envelope protein after membrane fusion[J]. Nature, 2004, 427(6972):313.
- [19] 魏全新, 刘贤锋, 黄锵, 等. 遗传算法选择方法的比较分析[J]. 通讯和计算机(中英文版), 2008, 5(8):61-65.
- [20] 颀斌, 杨扬, 旷毅. 基于多目标遗传算法的云服务部署优化方法[J]. 华中科技大学学报(自然科学版), 2016, 44(z1):80-83.
- [21] Kamphausen S, Hölte N, Wirsching F, et al. Genetic algorithm for the design of molecules with desired properties[J]. Journal of Computer-Aided Molecular Design, 2002, 16(8/9):551-567.
- [22] 张庆科, 杨波, 王琳, 等. 基于 GPU 的现代并行优化算法[J]. 计算机科学, 2012, 39(4):304-310.

Optimization of peptide inhibitors targeting E protein of dengue virus using genetic algorithms

He Zihao^{1a,2,3}, Zhao Junqiang^{1b,4}, Yang Min^{1a,2,3}, Zhang Songjie^{1a,2,3}, Xu Yongtao^{1a,2,3}

(1a. School of biomedical engineering; Sanquan College Xinxiang Medical University, Xinxiang 453000, China; 2. Xinxiang key laboratory of biomedical information research, Xinxiang 453000, China; 3. Henan engineering laboratory of combinatorial technique for clinical and biomedical big data, Xinxiang 453000, China; 4. Xinxiang key laboratory of Medical Artificial Intelligence and Virtual Reality, Xinxiang 453000, China)

Abstract: Dengue virus is a serious threat to human health, but there is no effective treatment. Membrane fusion is the first step of viral infection. The conformational changes of envelope (E) protein play a key role in the process of membrane fusion. Therefore, the development of peptide inhibitors targeting E protein provides new ideas for the treatment of dengue fever. In our study, a genetic algorithm for optimizing DN58opt, which is an effective inhibitor targeting E protein of dengue virus peptide was designed. Three different selection operators: roulette method, championship method and parallel method were used to optimize DN58opt for 300 generations. The results of parallel method were better than other method, and DN58ga was obtained. At the same time, some polypeptide derivatives were received. Through computer evaluation, these polypeptides showed better binding affinities than DN58opt. The novel peptide inhibitors have potential to disrupt the protein-protein interaction in the fusion process and may serve as starting points for development of novel inhibitors for viral E proteins of dengue virus.

Keywords: Dengue virus; membrane fusion; envelope protein; peptide inhibitor; genetic algorithm

[责任编辑 杨浦 王凤产]