

一种新型氨基酸拓扑结构信息矢量及在肽定量构效关系研究中的应用

李志良^{①②}, 李根容^{①②}, 舒茂^{①②}, 孙家英^{①②}, 杨善彬^{①③}, 梅虎^{①④*}, 张梦军^{①②③},
周萍^{①②}, 吴世容^{①②}, 陈国华^{①②}, 吕凤林^{①③}, 吕廷亭^{①②③}

① 重庆大学化学化工学院/生物工程学院, 重庆 400044;

② 化学生物传感与计量学国家重点实验室, 长沙 410012;

③ 第三军医大学医学检验系, 重庆 400037;

④ 新加坡国立技术学院生命科学技术中心, 新加坡 139651

* 联系人, E-mail: meihu@126.com

收稿日期: 2007-05-20; 接受日期: 2007-09-06

国家“863”项目(2006AA02Z123)、霍英东基金(98-7-6)、国家新药基金(97-8-6)、化学生物传感与计量学国家重点实验室基金(2005012)、重庆直辖市应用基础基金(01-3-6)及重大自主创新基金科技攻关项目(04-9-9)资助

摘要 从表征分子的 66 个结构参量和拓扑指数出发, 经主成分分析(PCA)得到一种包含 134 个氨基酸体系的特征描述子: 氨基酸拓扑与结构信息矢量(VTSA). 将其应用于 58 个血管紧张素转化酶抑制剂(ACEI)和 88 个弹性蛋白酶模拟底物催化动力体系(ESCK)的定量构效关系(QSAR)或定量序效模拟(QSAM)研究中, 结合遗传偏最小二乘(GPLS)、支持向量机回归(SVMR)以及本文设计的免疫神经网络(INN)技术, 成功建立了上述两个肽类似物样本集定量预测模型, 并取得优于已有文献报道的结果: ACEI, $R_{cu}^2 \geq 0.82$, $R_{cv}^2 \geq 0.77$, $E_{rmse} \leq 0.44$ (GPLS+SVM); ESCK, $R_{cu}^2 \geq 0.84$, $R_{cv}^2 \geq 0.82$, $E_{rmse} \leq 0.20$ (GPLS + INN).

关键词

氨基酸拓扑结构信息矢量(VTSA)

定量构效关系(QSAR)

定量序效模拟(QSAM)

分子结构与拓扑特征描述子(MSCD)

血管紧张素转化酶抑制剂(ACEI)

弹性蛋白酶模拟底物催化动力体系(ESCK)

众所周知, 肽是生物体内非常重要的活性物质, 具有激素、酶抑制剂或底物、生长促进剂、抑制剂、神经递质、免疫调节剂以及抗生素等多种生物功能, 因此有着广阔药物研发和应用前景. 在新药创制和开发环节中, 定量构效关系(QSAR)一直是药物化学和药理科学研究的一个中心问题, 同时也是药物设计学的重要基础. 近年来, 已在相关领域开始采用 QSAR 研究肽及类似物化学结构与生物活性间的内在联系, 并取得一定成效^[1~4]. 由于肽类物质分子结构

巨大且空间构象具很大柔性, 因此其 QSAR 所面临最艰巨任务是如何将其分子结构特性提取为一组特征参数, 并与生物活性建立起优良统计学模型. 早在 20 世纪 60 年代 Anfinsen 等^[5]研究表明生物大分子三维结构信息包含于其一级序列当中, 该点启示给肽分子结构表达提供了一个可行途径, 即通过对组成序列单个氨基酸残基作结构表征, 进而构成整体肽链描述矢量. 在该思想指引下, Sneath^[6]首先使用 20 种标准天然氨基酸的几个半定量实验参数对氨基酸结

构进行表征, 并成功用于后叶催产素类似物活性预测中; 之后基于实验或计算所得单个氨基酸描述子相继提出, 其代表如Hellberg等^[7,8]采用正交变换技术对氨基酸 29 种物理化学性质处理产生Z标度(Z scale)以及Collantes等^[9]使用量子化学途径计算的残基侧链全向表面积(ISA)和电荷指数(ECI). 然而, 由于非蛋白或非编码氨基酸在自然界普遍存在以及药物设计中氨基酸结构改造或多肽修饰所起到的重要作用, 特别是近年来模拟肽学发展, 业已认识到仅仅对 20 种标准编码氨基酸所组成的编码肽进行QSAR研究难以满足相关学科领域发展要求. Jonsson等^[10]首先在 55 种氨基酸(20 种标准编码氨基酸和 35 种合成非编码氨基酸)的 12 个实验性质基础上得到第一种可用于非编码氨基酸表征的多元特征矢量. 之后Sandberg等^[11]将Z标度推广到包括编码氨基酸在内的 87 种常见氨基酸. 但是由于非编码氨基酸多样性以及实验参数收集困难, 使得对其发展一组统一而高效的描述子体系变得极为艰难, 因此该类研报仅见上述两例. 很显然, Jonsson和Sandberg工作还很不完善, 一方面所收集的非编码氨基酸个数非常有限; 另则对每个氨基酸使用初始变量受实验条件限制因而数量不多且个别氨基酸参数或缺.

结构拓扑描述子是完全不同于传统物理化学参数的结构表征方法, 它基于分子图中拓扑不变量将化合物转化为一组具数学抽象意义的特征参数, 从而在纯理论层面上实现分子结构表征. 自 1947 年Winer^[12]提出 W 指数以来, 迄今已出现 200 余种分子拓扑描述子, 其代表如Hosoya指数(Z)^[13]、Randic指数(χ)^[14]、Balaban指数(J)^[15]、Kier-Hall连接指数($^m\chi^v$)^[16]等. 实践证明, 化合物分子拓扑描述子与多种物化性质及生物活性显著相关, 因而在QSPR/QSAR中得到广泛应用^[17-20]. 鉴此, 考虑到拓扑描述子仅从分子二维图形连接方式即可求得, 不受实验条件限制, 且能充分表达分子结构特征, 本文将它应用于 134 种常见氨基酸参数表征中, 通过主成分分析(PCA)技术对 66 个通用拓扑描述子进行变量压缩及信息提取, 最终每一种氨基酸得到 6 个显著主成分得分作为整体分子拓扑特征综合结构信息描述子. 为便于使用, 这里暂将其命名为氨基酸拓扑结构信息矢量(VTSA).

1 原理和方法

1.1 氨基酸数据集

从多篇文献^[10,11,21-23]收集 134 种氨基酸类型及分子结构, 参见网络版附表 1. 其中 1~20 号为 20 种标准编码氨基酸; 另外包括最近发现的两种非标准编码的 21 和 22 号氨基酸——硒代半胱氨酸^[24]和吡咯赖氨酸^[25]; 其余 112 种是一些较常见天然非编码氨基酸及人工修饰非蛋白氨基酸, 如鸟氨酸、羟脯氨酸、高丝氨酸等.

1.2 氨基酸拓扑结构信息矢量(VTSA)

Cho等^[26]曾用Focus-2D途径来实现五肽血管舒缓激肽促进剂(BPP)虚拟肽库设计, 即使用Molconn-X软件计算肽链每个位点多拓扑表达矢量, 以此组合成整个化合物结构描述子并参与遗传算法(GA)种群迭代寻优过程, 用以获得包含最优残基类型的序列特征. 以此为出发点, 考虑提出氨基酸综合拓扑描述子, 其在低维矢量基础上包含目标分子绝大多数拓扑信息, 并适应于大量非天然氨基酸体系. 采用分子图形学软件包Chemoffice 2005 搭建 134 种氨基酸分子连接结构, 并以MDL文件形式(*.mol)将构建完成分子进行保存; 进而使用描述子计算软件Chem3D 8.0、ChemSW 5.0、Codessa 2.7 和Dragon 5.2 对每个氨基酸生成 66 个通用分子拓扑描述子(拓扑描述子名称参见网络版附表 2, 具体数据见网络版附表 3). 需说明的是由于拓扑描述子仅涉及分子连接方式而与三维构象无关, 因此在搭建分子结构时毋需对其作几何优化. 此处考虑到拓扑描述子反映分子整体结构特征, 不同描述子间可能存在较大信息重叠, 且直接使用 66 个参数来表征肽链中单个氨基酸将会导致QSAR模型难以实现且解释困难, 故结合经典数据信息提取技术——主成分分析(PCA)压缩变量维数. 先按列对原始变量矩阵 $X_{134 \times 66}$ 作数据标准化处理, 继用PCA得到 6 个显著主成分(按SPSS 13.0 默认标准, 特征值大于 1 者), 分别解释原始变量矩阵 66.60%, 10.29%, 7.52%, 4.33%, 2.40%和 1.56%方差, 累积解释方差为 92.70%. 对每个氨基酸所对应 6 个显著主成分得分, 它由氨基酸 66 个原始变量与每个主成分得分系数乘积计算而来, 当用来替代原始变量矩阵时仅损失 7.29%信息. 这里称 134 种氨基酸 6 个显著

主成分得分为氨基酸拓扑结构信息矢量(VTSA), 其值参见网络版附表 4. 主成分分析使用统计学软件包 SPSS 13.0 实现.

当使用 VTSA 对一个肽类似物序列进行表征时, 所包含全部氨基酸残基依次由相应 VTSA 描述子所代表. 例如对一个二肽序列第 1 个位点可由对应氨基酸 6 个 VTSA 所描述, 同样第 2 个位点也由其 6 个 VTSA 表征, 这样就可产生 12 个 VTSA 参数($2 \times 6 = 12$), 并作为这个二肽类似物整体分子结构变量. 以此类推长度为 n 的肽链可被表征为 $n \times 6$ 个描述变量.

1.3 建模方法

1.3.1 免疫神经网络(INN)

神经网络(NN)是一种通过模拟生物神经信号处理系统发展起来的机器学习方法, 它被广泛用于模式识别、函数拟合、过程控制、图形映射等众多领域. 但通常基于梯度递减的误差反向传播(BP)策略进行网络连接权值调整会导致ANN在训练过程中陷入局部最优, 且学习过程缓慢^[27]. 为改善ANN性能, 近来开始尝试将一些进化计算(EC)方法应用于ANN权值

训练, 并已取得一定成效^[28~30]. 免疫算法(IA)作为晚近发展的一种优秀EC策略已在众多领域表现出明显优势^[31], 由于该法借鉴复杂生物免疫系统的基本概念, 因此能避免许多传统算法容易早熟、收敛性差等诸多缺点. 此处借鉴克隆选择^[32]、免疫网络^[33]等当今常见IA算法, 针对ANN连接权值优化问题设计一种新型免疫神经网络(INN), 并作为BPNN替代工具用于肽类似物QSAR建模中. INN计算流程参见图 1, 限于篇幅这里只能对该算法思想进行概括性介绍. ①抗原识别及参数设置: 对待优化神经网络结构进行分析和确认, 并设定初始种群大小 N , 克隆选择比例 α , B细胞克隆数目 M , 清除率 η , 新成员比例 λ 等参数; ②编码初始B细胞抗体群: 随机生成 N 个初始实数串(B细胞抗体), 串长即为神经网络权值数目, 串中每个基因对应一个连接权值; ③克隆选择: 从整个群体中选出 $\alpha \cdot N$ 个高亲和性个体作为待操作对象. 亲和性 A 由下式计算: $A_i = 1/\exp(E_i)$, 其中 E 为个体所对应神经网络对训练集样本计算结果的均方根误差(E_{rmse} RMSEE); ④B细胞克隆: 将克隆选择所得高亲和性个体群按亲和性大小比例复制成为 M 个个体

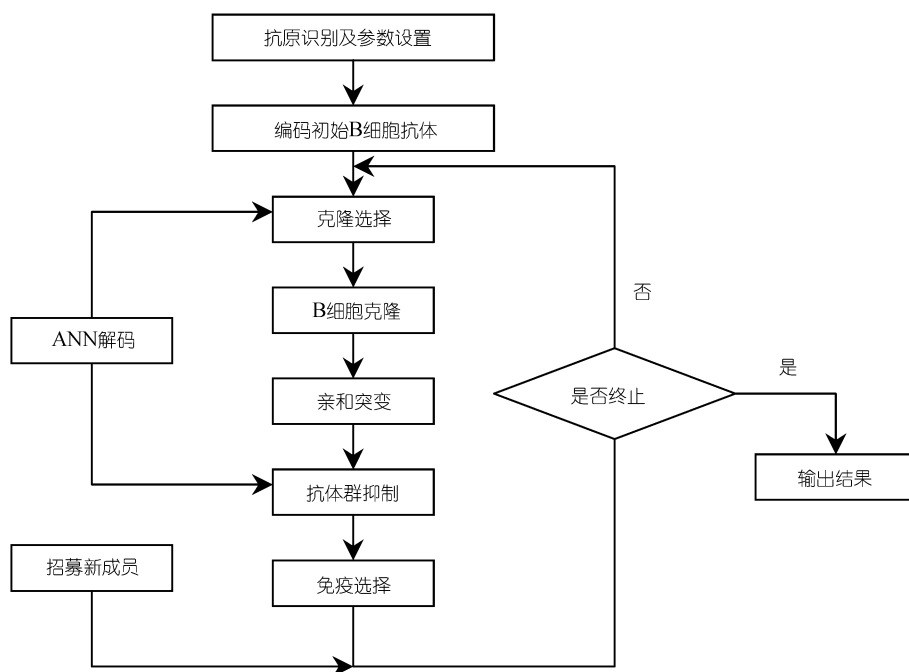


图 1 免疫神经网络 INN 的算法流程

①抗原识别及参数设置→②编码初始 B 细胞抗体群→③克隆选择→④B 细胞克隆→⑤亲和突变→⑥抗体群抑制→⑦免疫选择→⑧招募新成员→⑨ANN 解码器→完成/输出结果.

($M > \alpha \cdot N$); ⑤亲和突变: 将上述 M 个个体按亲和性大小反向随机变异, 即对每一个 B 细胞抗体 $W(W = \{w_1 w_2 w_3 \dots w_L, w_i \in M, 1 \leq i \leq L\})$ 的基因 w_i 实行如下改变: $w_i' = w_i + \sigma_i \cdot N(0, 1)$, 其中 $\sigma_i = [1 - f(W)]^{1/2}$; $N(0, 1)$ 代表服从标准正态分布随机数; $f(W)$ 为 B 细胞抗体 W 亲和性 A , 取值在 $[0, 1]$ 范围; ⑥抗体群抑制: 对亲和突变产生 B 细胞抗体群中个体按亲和性及浓度进行排序, 并将处于劣势(亲和性低且浓度偏大)的 $\eta \cdot M$ 个体加以定向清除. 有关个体 W_i 浓度 $C(W_i)$ 由下式计算:

$$C(W_i) = 1 / \{\sum_{j=1, j \neq i}^M |W_i - W_j|\}, \quad (1)$$

其中 $|W_i - W_j|$ 代表 i, j 两个个体在目标空间上 Euclid 距离; ⑦免疫选择: 将此时所得 $(1 - \eta) \cdot M$ 个个体与母体种群混合, 并以亲和性及浓度综合指标 Z 作为个体适应度, 在此基础上以转轮盘赌的方式随机选取 $(1 - \lambda) \cdot N$ 个个体. 这里 $Z_i = f(W) \cdot \exp[-C(W_i)]$; ⑧招募新成员: 在目标空间上随机产生 $\lambda \cdot N$ 个体与免疫选择得到 $(1 - \lambda) \cdot N$ 个个体共同组成最终包含 N 个个体的种群进入下一代繁殖; ⑨ANN 解码器: 将待优化神经网络连接权值与 B 细胞抗体进行相互转换, 同时计算个体亲和性. 可看到, 与 GA 显著不同之处在于 IA 在寻优过程中不但考虑目标函数相关信息, 同时也兼顾种群内部差异特性; 另外 IA 中 B 细胞克隆、抗体群抑制、招募新成员等算子等都为 GA 所不具有, 而实验证明引入这些算子有助于增强算法综合性能. 由此可认为 IA 在算法前期全局搜索能力及后期局部收敛性上都应优于 GA. INN 采用 C 语言自编写应用程序 focus_inn.exe 实现.

1.3.2 其他

遗传偏最小二乘 (GPLS) 和支持向量机回归 (SVMR) 技术是近年来发展起来的有效数学工具, 现已在 QSAR 研究领域得到广泛应用, 限于篇幅其详细原理和基础算法请参见有关文献^[34, 35].

2 结果及分析

2.1 血管紧张素转化酶抑制剂

2.1.1 概述

肾素血管紧张素系统(RAS)在调节人体血压中起着关键作用. 由肝脏产生血管紧张素原经肾素催化断裂为非活性十肽血管紧张素 I, 再经血管紧张素转

化酶(ACE)催化裂解为具极强血管收缩功能的八肽血管紧张素 II, 因此 ACE 就成为降血压药物作用靶点^[36]. ACE 抑制剂通过模拟结合底物血管紧张素 I 活性部位结构特征从而达到竞争抑制 ACE 有效生物活性的目的. 58 个二肽 ACE 抑制剂是一组经典 QSAR 样本集, 曾被众多文献用于测试新型氨基酸描述子有效性^[2, 9, 37~42]. 该组样本一级序列及生物活性引自文献^[37], 活性由 pIC_{50} 表示(网络版附表 5).

2.1.2 遗传偏最小二乘回归建模

首先使用 VTSA 描述子对 58 个二肽序列加以结构表征, 每个分子产生 12 个变量, 进而使用 GPLS 技术对初始变量进行筛选建模. 相关参数设置: 初始群体大小: 100; 最大遗传代数: 200; 收敛标准: 80%; 变异概率: 0.5%; 交叉互换: 2 点; 复制方式: 转轮盘赌; 交互检验: 留一法(LOO); 数据预处理: 标准化; 重复试验次数: 10; 其他均为默认设置. 筛选结果: 主成分数(A): 2; 适应度(e_{rmsecv} , RMSECV): 0.468; 最优变量子集: $VTSA_4^1, VTSA_5^1, VTSA_1^2, VTSA_2^2$ (上标代表所属残基位置, 下标代表该残基 VTSA 描述子编号). 由于遗传算法是一种随机搜索算法, 其运算结果具一定不确定性; 为避免偶然因素影响, 重复进行 10 次计算. 由于该组样本初始变量范围较小(仅 12 个变量), 所以 10 次运算中有 9 次都获得相同最佳变量子集, 唯一一次不同结果也极为相近(仅一变量不同, 且模型适应值非常接近), 由此可认为 GPLS 找到了全局最优变量子集. 最终所得 PLS 模型共使用 2 个显著主成分(数目由 LOOCV 确定)参与建模, 分别解释自变量 X 和因变量 Y 的 62.31% 及 81.98% 方差, 交叉检验所解释 Y 变量方差为 77.95%. 图 2 为 58 个样本在 PLS 模型前两个主成分上得分分布散点图, 其中 pIC_{50} 大于 4 的样本以圆圈标记; pIC_{50} 为 3~4 的样本以三角标记; pIC_{50} 小于 3 的样本以菱形标记. 可看到二肽按其活性大小从该图左下角到右上角呈依次递增分布, 具相近药效化合物较好地聚在一起, 由此可见 GA 筛选得到 4 个 VTSA 描述子子集的前两个主成分已基本能表现出该组样本活性分布特征. 另外所有样本中仅 2 及 5 号样本略微超出 95% 置信度 Hotelling T^2 椭圆圈, 经分析表明这两个二肽 C 端除都有一个很大色氨酸残基外, 其 N 端氨基酸残基分

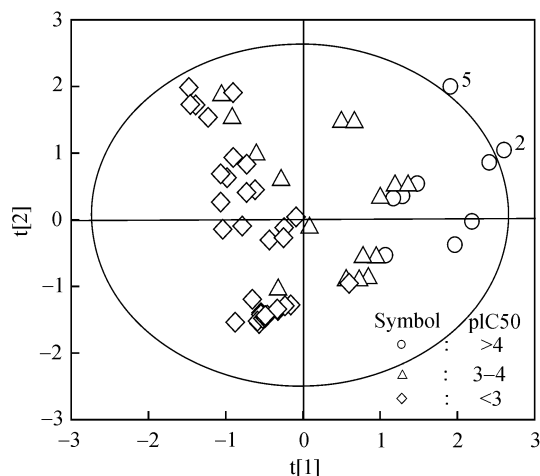


图2 在PLS模型中58个ACEI样本前两个主成分得分分布散点图

别为较大体积的异亮氨酸和精氨酸, 因此在整个样本集中它们具最大分子拓扑构形从而表现在得分分布散点图上略为异常. 图3给出PLS模型变量投影重要性指标(VIP)^[46], 可看到参与建模的4个变量中 $VTSA_1^2$ 和 $VTSA_2^2$ 具最大显著性, 它们同属于二肽第二个位点氨基酸残基, 由此可推测该位点氨基酸类型对整个化合物活性有较大影响. 通过观察可发现类似规律, 即当第二个氨基酸残基有较大拓扑体积时(如色氨酸W, 酪氨酸Y, 苯丙氨酸F), 二肽往往具较高生物活性; 反之亦然. GPLS对58个样本活性计算值与实测值相关参见图4.

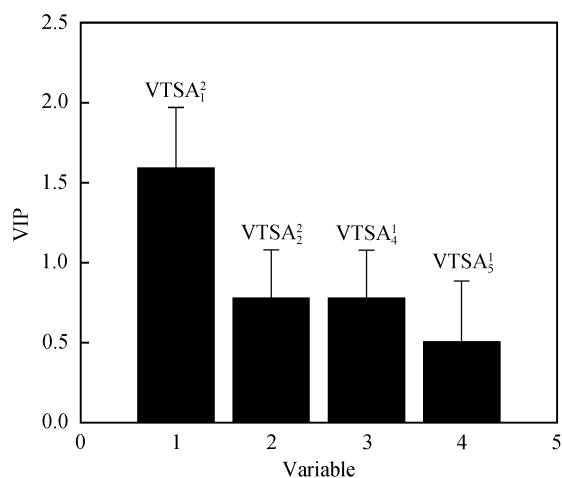


图3 PLS模型的变量投影重要性指标

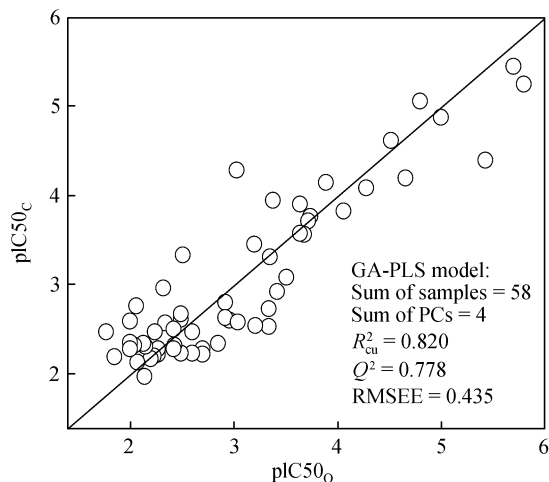


图4 以GPLS模型对58个ACE抑制剂生物活性计算值与实验观测值相关

2.1.3 支持向量机建模

进一步使用SVM技术对该组样本进行QSAR研究. 考虑到上文GPLS筛选4个变量仅与样本生物活性呈较明显线性相关, 而其余变量可能包含着非线性联系, 因此文中使用所有原始变量参与SVM建模. 首先对自变量空间进行 $[-1, +1]$ 规格化处理, 并以交叉检验均方根误差(e_{rmse} , RMSECV)作为模型质量判定标准(因SVM建模速度相对较慢, 故这里交叉检验为留1/5法), 经初步尝试表明在3种常见核函数(线性核、多项式核及径向基核)中线性核函数具较优效果, 再通过网格逐步搜索最终确定惩罚参数 C 和不灵敏度 ϵ 分别设置为100及0.1. 所得模型复相关系数 R_{cu}^2 、交叉检验 Q_{cu}^2 及均方根误差(E_{rmse} , RMSEE)为0.850, 0.802和0.378. 可看到, 由于SVM引入结构风险最小化(SRM)原则来指导建模, 从而有效地将期望风险降至最低, 使得模型泛化能力得到显著提高, 虽然业已采用较严格留1/5法作交叉验证, 但所得预测结果 Q_{cu}^2 仍较满意. 另在建立SVM模型时发现对于该组ACE抑制剂样本集而言线性核函数明显优于其他两种非线性核函数, 表明二肽VTSA描述子能较好地与其活性呈线性相关, 这与上文GPLS模型结论一致. 图5是SVM模型对ACE抑制剂活性计算值与实验观测值相关情况, 其中绝大多数样本点大致都均匀分布于过原点45°斜线附近, 仅19号化合物正向残差偏大, 对其结构分析表明该化合物由两个较大

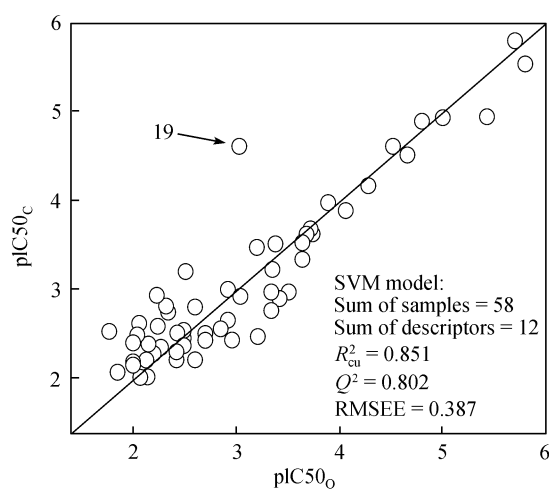


图5 以 SVMR 模型对 58 个 ACE 抑制剂生物活性计算值与实验观测值相关

氨基酸残基组成(异亮氨酸 I 和苯丙氨酸 F), 其分子整体拓扑构形远大于处于同等级活性的其他二肽, 因此模型将其 pIC_{50} 计算偏高. 另外收集已有文献对这 58 个 ACE 抑制剂样本集 QSAR 研究报道, 并将其与本模型同列于表 1 中以供比较. 其中 1~4 为当今一些主流氨基酸描述子建模情况; 5~8 是本实验室前期对该组样本研究结果; 9 和 10 是本文所得建模结果. 可看到与其他模型相比使用 VTSA 描述子建立的两个模型都具较优统计量, 特别是 SVM 模型, 其整体质量明显优于文献报道.

表 1 关于 ACE 抑制剂 QSAR 模型比较

编号	描述子	方法	$m^a)$	$A^a)$	$R_{cu}^{2\ b)}$	$Q_{cu}^{2\ c)}$	$E_{rmse}^d)$
1	Z scale ^[37]	PLS	6	2	0.770	0.723	—
2	t score ^[38]	PLS	14	1	0.744	—	0.50
3	ISA-ECI ^[9]	PLS	4	2	0.700	—	—
4	MS-WHIM ^[2]	PLS	6	2	0.708	0.637	0.54
5	VMEE ^[39]	SMR-MLR	10	2	0.741	0.710	0.503
6	SSIA ^[40]	SMR-PLS	4	1	0.789	0.772	0.471
7	VHSE ^[41]	SMR-PLS	5	1	0.770	0.745	0.480
8	VSTV ^[42]	PLS	6	1	0.789	0.767	0.462
9	VTSA ^{e)}	GPLS	4	2	0.820	0.778	0.431
10	VTSA ^{e)}	SVM	12	—	0.850	0.802 ^{f)}	0.378

a) m : 描述子变量数, A : 主成分数; b) R_{cu}^2 : 累计复相关系数 R^2 ; c) Q_{cu}^2 : 累计交互验证 R_{cu}^2 ; d) E_{rmse} (RMSEE): 估计的均方根误差; e) 本文结果; f) 留组法交互验证(5 组)

2.2 弹性蛋白酶模拟底物

2.2.1 概述

弹性蛋白酶是以能降解不溶性弹性蛋白为特征

的肽链内切酶, 属丝氨酸蛋白酶类并广泛存于动物胰脏中. 众多研究表明人体中弹性蛋白酶与其天然底物平衡失调是导致慢性阻塞性肺疾病(COPD)重要原因^[43]. 近年来, 人工合成高效低毒弹性蛋白酶抑制剂成为开发治疗肺气肿药物研究热点. 1993 年 Nomizu 等^[44]首次采用 Free-Wilson/Fujita-Ban 法对一组包含 88 个弹性蛋白酶模拟底物样本集进行初步构效研究; 之后 Kimura 等^[45]在利用 Z 标度对该组样本序列表征的基础上使用非线性偏最小二乘回归(QPLS)成功建立数个较优预测模型; 基此 Sandberg 等^[11]将 Z 标度适用范围推广到包含 20 种标准编码氨基酸在内的 87 种常见氨基酸, 并利用该组样本集对扩展 Z 标度进行验证, 亦取得较优结果.

2.2.2 遗传偏最小二乘回归建模

该 88 个人工合成弹性蛋白酶模拟底物序列组成为: 琥珀酰三肽对硝基苯胺即 *Suc-A-B-Ala-pNa* (*Suc*: 琥珀酰succinyl; *pNa*: 对硝基苯胺 *p*-nitroanilide), 其中 A, B 两位点为可变残基. 通过分光光度法观测催化产物对硝基苯胺产率测得化合物与受体酶间 3 个反应动力学常数: $\lg(1/K_m)$, $\lg(k_{cat})$ 和 $\lg(k_{cat}/K_m)$ (参见网络版附表 6)^[44]. Sandberg 等^[11]使用 D-optimal 技术将该组样本集分成一个包含 32 个样本训练集和一个包含 56 个样本测试集; 为便比较本文中沿袭上述划分方式 (参见网络版附表 6: 1~32 号为训练集, 33~88 号为测试集). 由于该组底物仅含两个可变残基(A 和 B), 故这里只需使用 VTSA 描述子对这两处残基进行表征即可. 进而采用 GPLS 将表征得到 12 个原始 VTSA 变量与生物活性进行筛选建模(GA 参数设置同 2.1 节, 变量矩阵 $Y_{32 \times 3}$ 含 3 个目标因子)多次实验结果表明直接使用 VTSA 描述子建模难以取得较优的效果. Kimura^[45]和 Sandberg 等^[11]研究指出 A、B 两个残基位点存在着显著非线性因素, 以至他们在建模之前对描述子作二次项和交叉项变换. 鉴此本文中对 12 个 VTSA 变量采用同样处理手段, 最终产生一个由 90 个变量组成的变量空间, 其中 $V_1 \sim V_{12}$ 为 12 个原始 VTSA 描述子; $V_{13} \sim V_{24}$ 为 VTSA 描述子二次项; $V_{25} \sim V_{90}$ 为 VTSA 描述子交叉项. 在此基础上使用 GPLS 得到训练集样本最佳变量子集为: V_{10} , V_{12} , V_{16} , V_{17} , V_{20} , V_{21} , V_{22} , V_{23} , V_{27} , V_{32} , V_{46} , V_{54} , V_{70} , V_{79} , 可看到该子集中绝大多数

为二次项和交叉项. PLS建模得到 3 个显著主成分, 其累积解释 $\mathbf{Y}_{32 \times 3}$ 变量矩阵 95.1%方差, 交叉检验解释 $\mathbf{Y}_{32 \times 3}$ 矩阵方差为 83.4%(图 6). 图 7 是 14 个变量在 PLS 模型前两个主成分上载荷分布情况, 可看到不同类型变量大致聚成几个区域, 其中原始项(V_{10}, V_{12})和交叉项($V_{27}, V_{32}, V_{46}, V_{54}, V_{70}, V_{79}$)分处于第三及一象限; 而二次项($V_{16}, V_{17}, V_{20}, V_{21}, V_{22}, V_{23}$)多集中在第二、四象限. 另外, 在第二个主成分空间上催化常数 $\lg k_{\text{cat}}$ 与 $\lg k_{\text{cat}}/K_m$ 处于相同方向, 而米氏常数 $\lg(1/K_m)$ 则与之相反, 这较好表达它们三者之

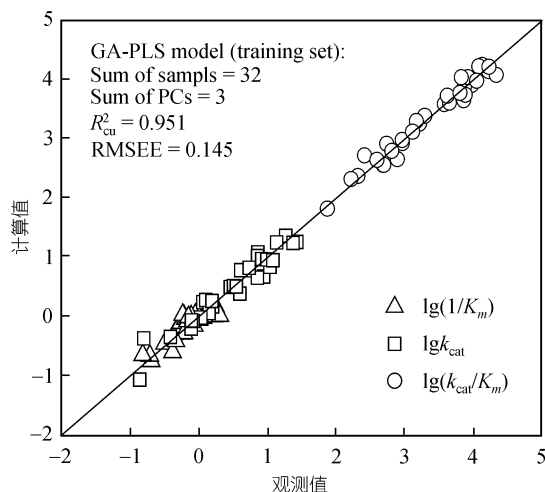


图 6 以 GPLS 模型对 32 个训练集样本活性计算值与实验观测值相关

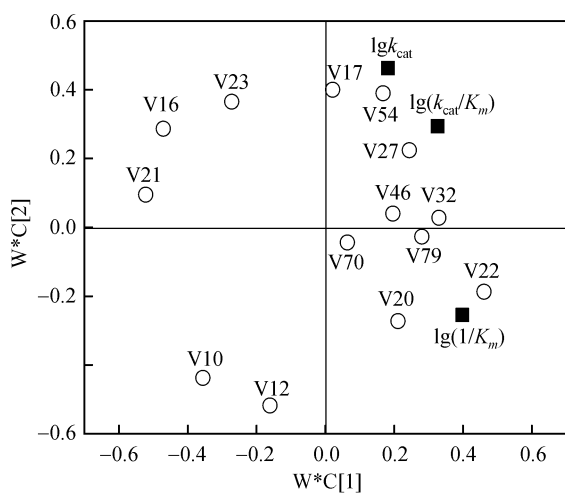


图 7 用 14 变量子集对 GPLS 模型前两个主成分载荷贡献散点图

间函数关系. 使用上述模型对 56 个测试集样本 $\lg(1/K_m)$, $\lg(k_{\text{cat}})$ 及 $\lg(k_{\text{cat}}/K_m)$ 预测均方根误差 E_{rmsep} (RMSEP)分别为 0.200, 0.215 和 0.291(图 8); 而 Sandberg 等^[11]采用扩展 Z 标度对这 56 个样本 $\lg(k_{\text{cat}})$ 和 $\lg(k_{\text{cat}}/K_m)$ 预测的 E_{rmsep} 分别为 0.23 和 0.30(在该文献中没有对 $\lg(1/K_m)$ 进行建模), 由此可见本文模型较优.

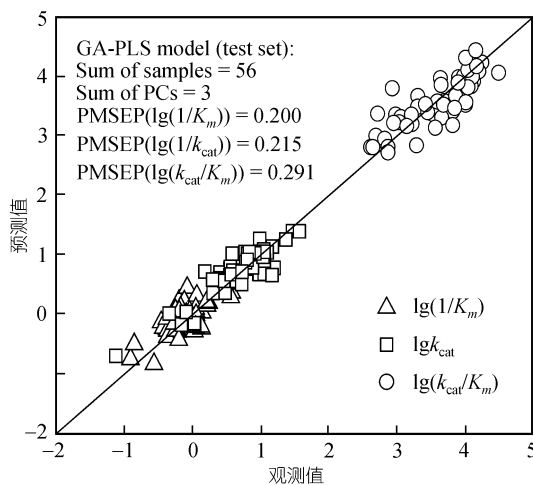


图 8 以 GPLS 模型对 56 个测试集样本活性预测值与实验观测值相关

2.2.3 免疫神经网络建模

考虑到非线性方法中 SVM 只能针对单因素建模, 即一个模型中只含一个因变量. 虽对多因素问题可采用分别建模加以解决, 但从可解释性及便于使用角度出发更希望在一个模型中包含所有目标. 因此选用 INN 技术来实现对弹性蛋白酶模拟底物 3 个参数模拟. ① 样本集划分: 考虑到神经网络对训练集样本数量要求较为苛刻, 故这里重新将这 88 个底物划分成一个包含 60 个样本训练集、一个包含 10 个样本监控集和一个包含 18 个样本测试集(参见网络版附表 6, 训练集: 1~60 号样本; 监控集: 61~70 号样本; 测试集: 71~88 号样本), 其中监控集用于防止网络训练过程中出现过拟合; ② 变量预处理: 从降低模型复杂度和提高网络训练速度等方面考虑, 使用 PCA 对初始 90 维自变量进行空间压缩和信息提取, 最终选取前 17 个特征值大于 1 的显著主成分作为 INN 输入变量, 其累积解释原始变量矩阵 92.81%方差. 考

考虑到主成分得分单位尺度较为一致, 故这里不再作数学处理而直接用于 INN 建模; ③ 网络拓扑结构确定: 由于神经网络隐含层节点数目至今还没一个统一确定标准, 按照通常规定模型中待定参数个数(权值数目)不能超过训练样本数, 因此很容易计算得到隐含层节点必须满足 $\text{hidden} < 2.71$, 据此确定网络拓扑结构为 $17(+1\text{bias})-2(+1\text{bias})-3$ (即 IA 搜索目标空间维数为 45); ④ 网络传递函数: 隐含层和输出层激活函数分别用 Sigmoid 和 Linear; ⑤ 参数设定: 初始种群大小 $N = 100$; 克隆选择比例 $\alpha = 0.6$; B 细胞克隆数目 $M = 120$; 清除率 $\eta = 0.1$; 更新成员比例 $\lambda = 0.05$; 最大繁殖代数 $C = 500$. 同时采用监控集来防止网络出现过拟合, 即将每次迭代后种群历史最优个体解码成相应神经网络, 并计算该网络对监控集样本 3 个目标平均误差大小, 当计算误差随迭代而开始显著增大时即停止进一步训练. 由于 INN 是种随机算法, 每次训练得到结果都不固定; 为避免偶然因素, 对上述网络重复进行 50 次训练, 并以最优结果作最终确定的网络模型. 该模型训练过程中种群历史最优个体在训练集和监控集上误差(3 个目标上平均均方根误差 E_{rms} , RMS)变化曲线见图 9, 从中可看到训练集误差曲线总是呈递减改变, 但当网络训练到第 213 次时监控集误差开始增加, 由此可认为此时网络出现过拟合现象而应停训. 该模型对测试集 18 个样本 $\lg(1/K_m)$, $\lg(k_{\text{cat}})$ 和 $\lg(k_{\text{cat}}/K_m)$ 预测均方根误差 E_{rmsep} (RMSEP)分别为 0.211、0.253 及 0.202, 略优于 GPLS 模型. 但考虑到 INN 输入变量经较多数学处理, 可能损失部分信息, 且没对原始变量作任何挑选, 因此总

体上 INN 对该组样本模型构造能力优于 GPLS. 另外使用 BPNN 作对比研究发现多次训练虽亦能取得与 INN 相近建模结果, 但该算法极易陷入局部极优解、训练速度较慢且模型稳定性差, 其综合性能在一定程度上逊色于 INN. 需说明的是由于 INN 结构较为复杂, 待设参数较多, 如何达到这些参数最优配比成为该算法待解决问题. 另外虽 Kimura^[45] 和 Sandberg^[11] 在模型划分方式、因变量个数等方面与本文有所不同, 但为便于参考还将其结果与本建模值同列于表 2. 其中 Kimura^[45] 模型并没划分测试集而直接使用整个 88 个样本进行建模; 而 Sandberg^[11] 在研究该组样本时没加入 $\lg(1/K_m)$, 其因变量矩阵仅含 $\lg(k_{\text{cat}})$ 和 $\lg(k_{\text{cat}}/K_m)$

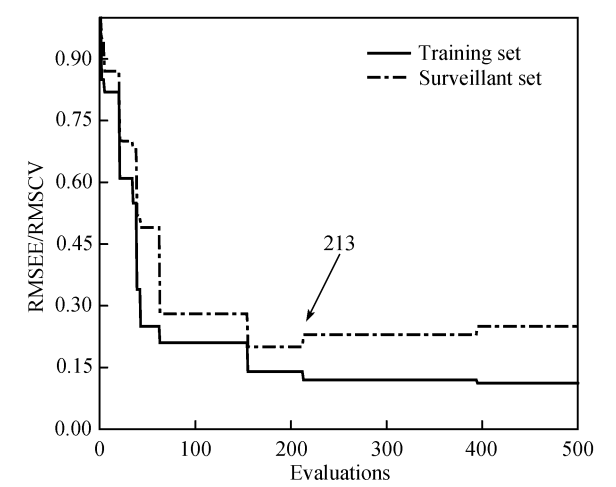


图 9 在 INN 训练过程中群体历史最优个体在训练集和监控集上误差变化曲线

表 2 关于弹性蛋白酶模拟底物 QSAR 模型比较

序号	活性	描述子	方法	$m^a)$	$A^a)$	训练集	$R_{\text{cu}}^{2b)}$	$Q^2c)$	$E_{\text{rmse}}^d)$	测试集	$E_{\text{rmsep}}^e)$
1	$\lg k_{\text{cat}}$	Z scale ^[11]	PLS	17	4	32	0.930	0.780	0.29	56	0.23
	$\lg(k_{\text{cat}}/K_m)$								0.30		0.30
2	$\lg(1/K_m)$	Z scale ^[45]	QPLS	6	2	98	0.542	0.410	0.21	0	—
	$\lg k_{\text{cat}}$				4		0.842	0.748	0.21		—
	$\lg(k_{\text{cat}}/K_m)$				3		0.754	0.629	0.29		—
3	$\lg(1/K_m)$	VTSA ^{g)}	GPLS	14	3	32	0.952	0.833	0.145	56	0.200
	$\lg k_{\text{cat}}$										0.215
	$\lg(k_{\text{cat}}/K_m)$										0.291
4	$\lg(1/K_m)$	VTSA ^{g)}	INN	17	—	60	0.853	—	0.125, 0.192 ^{g)}	18	0.211
	$\lg k_{\text{cat}}$						0.918	—	0.172, 0.200 ^{g)}		0.253
	$\lg(k_{\text{cat}}/K_m)$						0.947	—	0.154, 0.189 ^{g)}		0.202

a) m : 描述子变量数; A : 主成分数; b) R_{cu}^2 : 累计复相关系数 R^2 ; c) Q_{cu}^2 : 累计交互验证 R_{cu}^2 ; d) E_{rmse} (RMSEE): 估计的均方根误差; e) E_{rmsep} (RMSEP): 预测的均方根误差; f) 本文结果; g) 监控集的预测均方根误差

两项. 通过对比发现, 在 VTSA 描述子基础上通过两种不同建模途径得到结果整体上优于文献报道, 由此认为 VTSA 描述子同样能较好用于非编码氨基酸 QSAR 体系.

3 结束语

定量构效关系(QSAR)方法学研究与开发包括分子描述子设计与整合及建模方法研发和改善, 这两者缺一不可而相辅相成. 对前者文中尝试将分子拓扑学引入到非编码肽结构表征及建模中, 结合主成分分析技术得到一种通用型氨基酸结构信息描述子: VTSA, 使用该描述子有效解决传统方法实验参数收

集困难、氨基酸分子多样性受限等诸多问题; 对后者文中设计一种新型免疫神经网络算法, 并在肽类似物 QSAR 研究中取得初步成功. 同时结合当今一些较有效的建模方法如 GPLS 和 SVM, 对 58 个二肽血管紧张素转化酶抑制剂及 88 个弹性蛋白酶模拟底物进行较深入构效分析, 结果表明 VTSA 不仅能对纯编码氨基酸所组成肽链进行表征, 同样适用于非编码氨基酸体系结构描述, 并皆取得优于已有文献结果, 从而表明基于结构与拓扑性质的综合信息描述子 VTSA 能较好反映肽及其类似物结构与活性之间内在联系, 并有望在 QSAR 研究特别是肽类体系中得到广泛应用.

致谢 生物力学与组织工程教育部重点学科、化学医药工程及生物组学信息重庆市重点实验室周鹏博士、郑小林、李声时、蔡绍哲、李晟励、杨力、李圣和、王庆云等教授, 许志玲、田菲菲、梁桂兆、刘振德、李苑、蒋稼欢、高剑坤等博士进行有关讨论或给予有益协助, 谨致谢忱!

参考文献

- 1 Kidera A, Konishi Y, Oka M. A statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J Protein Chem*, 1985, 4: 23—55[DOI]
- 2 Zaliani A, Gancia E. MS-WHIM scores for amino acids: A new 3D-description for peptide QSAR and QSPR studies. *J Chem Inf Comput Sci*, 1999, 39: 525—533[DOI]
- 3 Liu S S, Cai S X, Yin C S. A novel MHDV descriptor for dipeptide QSAR studies. *J Chin Chem Soc*, 2001, 48: 253—260
- 4 Raychaudhury C, Banerjee A, Bag P. Topological shape and size of peptides: Identification of potential allele specific helper T cell antigenic sites. *J Chem Inf Comput Sci*, 1999, 39: 248—254
- 5 Anfinsen C B, Haber E, Sela M. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA*, 1961, 47: 1309—1318[DOI]
- 6 Sneath P H. Relations between chemical structure and biological activity in peptides. *J Theor Biol*, 1966, 12: 157—195[DOI]
- 7 Hellberg S, Sjostrom M, Wold S. The prediction of bradykinin potentiating potency of pentapeptides. An example of a peptide quantitative structure-activity relationship. *Acta Chem Scand*, 1986, 40: 135—140[DOI]
- 8 Hellberg S, Sjostrom M, Skagerberg B. Peptide quantitative structure-activity relationships, a multivariate approach. *J Med Chem*, 1987, 30: 1126—1135[DOI]
- 9 Collantes E R, Dunn W J. Amino acid side chain descriptors for quantitative structure-activity relationship studies of peptide analogues. *J Med Chem*, 1995, 38: 2705—2713[DOI]
- 10 Jonsson J, Eriksson L, Hellberg S. Multivariate parameterization of 55 coded and non-coded amino acids. *Quant Struct-Act Relat*, 1989, 8: 204—209[DOI]
- 11 Sandberg M, Eriksson L, Jonsson J. New chemical descriptors for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J Med Chem*, 1998, 41: 2481—2491[DOI]
- 12 Winer H. Structural determination of paraffin boiling point. *J Am Chem Soc*, 1947, 69: 2636—2641[DOI]
- 13 Hosoya H. Topological index. A new proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull Chem Soc*, 1971, 44: 2332—2339[DOI]
- 14 Randic M. On characterization of molecular branching. *J Am Chem Soc*, 1975, 97: 6609—6615[DOI]
- 15 Balaban A T. High discrimination distance-based topological index. *Chem Phys Lett*, 1982, 89: 399—404[DOI]
- 16 Kier L B, Hall L H. *Molecular Connectivity in Structure-Activity Analysis*. New York: J Wiley & Sons, 1986

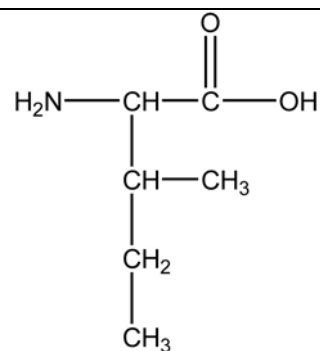
- 17 Bonchev D, Mekenjan O, Protic G. Application of topological indices to gas chromatographic data: Calculation of the retention indices of isomeric alkylbenzenes. *J Chromatogr*, 1979, 176: 149—156[DOI]
- 18 Buydens L, Massart D L, Geerlings P. Prediction of gas chromatographic retention indexes with topological, physicochemical, and quantum-chemical parameters. *Anal Chem*, 1983, 55: 738—744[DOI]
- 19 Call D J. Molecular connectivity in chemistry and drug research. *Environ Toxicol Chem*, 1995, 4: 10—19
- 20 Basak A C, Gute B D, Grunwald G D. Comparative study of topological and geometrical parameters in estimating normal boiling points and octanol/water partition coefficient. *J Chem Inf Comput Sci*, 1996, 36: 1054—1060[DOI]
- 21 Bock A, Forchhammer K, Heider J. Selenocysteine: The 21st amino acid. *Mol Microbiol*, 1991, 5: 515—520[DOI]
- 22 Atkins J F, Gesteland R. The 22nd amino acid. *Science*, 2002, 296: 1409—1410[DOI]
- 23 阎爱新, 田桂林, 叶蕴华. 非蛋白氨基酸对生物活性多肽的修饰及其构效关系的研究进展. *有机化学*, 2000, 20: 299—305
- 24 Chambers I, Frampton J, Goldfarb P. The structure of the mouse glutathione peroxidase gene: The selenocysteine in the active site is encoded by the termination codon TGA. *EMBO J*, 1986, 5: 1221—1227
- 25 Srinivasan G, James C M, Krzycki J K. Pyrrolysine encoded by UGA in archaea: Charging of a UAG-decoding specialized tRNA. *Science*, 2002, 296: 1459—1462[DOI]
- 26 Cho S J, Zheng W, Tropsha A. Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. *J Chem Inf Comput Sci*, 1998, 38: 259—268[DOI]
- 27 何永勇, 褚福磊, 钟秉林. 基于进化计算的神经网络设计与实现. *控制与决策*, 2001, 16: 257—262
- 28 Montana D J, Davis L. Training feedforward neural networks using genetic algorithms. In: *Proceedings of the International Joint Conference on Neural Network (IJCNN)*, 1989
- 29 De Weijer A P, Lucasius C B, Buydens L M C. Using genetic algorithms for an artificial neural network model inversion. *Chemom Intell Lab Syst*, 1993, 20: 45—55[DOI]
- 30 van den Bergh F, Engelbrecht A P. Training product unit networks using cooperative particle swarm optimizers. *Proceedings of the Third Genetic and Evolutionary Computation Conference*, 2001
- 31 王磊, 潘进, 焦李成. 免疫算法. *电子学报*, 2000, 28: 74—78
- 32 De Castro L N, Von Zuben F J. The clonal selection algorithm with engineering applications. *Genetic and Evolutionary Computation Conference*, Las Vegas USA, 2000: 36—37
- 33 Toma N, Endo S, Yamada K. The immune distributed competitive problem solver with major histocompatibility complex and immune network. *IEEE International Conference on Systems, Man, and Cybernetics*, Nashville TN USA, 2000, 3: 1865—1870
- 34 Rogers D, Hopfinger A J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J Chem Inf Comput Sci*, 1994, 34: 854—866[DOI]
- 35 张学工. 关于统计学习理论与支持向量机. *自动化学报*, 2000, 26: 32—42
- 36 Hassell C H. The Design and synthesis of new triazolo-, pyrazolo-, and pyridazo-pyridazine derivatives as inhibitors of angiotensin converting enzyme. *J Chem Soc Perkin Trans I*, 1984, 23: 155—162[DOI]
- 37 Hellberg S, Eriksson L, Jonsson J, Lindgren F, Sjöström M, Skagerberg B, Wold S, Andrews P. Minimum analogue peptide sets (MAPS) for quantitative structure-activity relationships. *Int J Pept Protein Res*, 1991, 37: 414—424
- 38 Cocchi M, Johansson E. Amino acids characterization by GRID and multivariate data analysis. *Quant Struct Act Relat*, 1993, 12: 1—8[DOI]
- 39 Li S Z, Fu B H, Wang Y Q. On structural parameterization and molecular modeling of peptide analogues by molecular electronegativity-edge vector (VMEE): Estimation and prediction for biological activity of pentapeptides. *J Chin Chem Soc*, 2001, 48: 937—944
- 40 周鹏, 吴世容, 李志良. 一种基于三维原子场相互作用矢量的新型氨基酸结构信息描述子. *科学通报*, 2006, 51(1): 34—39
- 41 Mei H, Zhou Y, Li, S Z. A new descriptor of amino acids and its application in peptide QSARs. *Peptide Science*, 2005, 80: 775—786
- 42 梅虎, 周原, 李志良. 一种新的氨基酸描述子及其在肽 QSAR 中的应用. *物理化学学报*, 2004, 20: 821—825
- 43 Snider G L. Animal models of emphysema. *Am Rev Respir Dis*, 1986, 133: 149—150
- 44 Nomizu M, Iwaki T, Yamashita T, Inagaki Y, Asano K, Akamatsu M, Fujita T. Quantitative structure-activity relationship (QSAR) study of elastase substrates and inhibitors. *Int J Pept Protein Res*, 1993, 42: 216—226
- 45 Kimura T, Miyashita Y, Funatsu K. Quantitative structure-activity relationships of the synthetic substrates for elastase enzyme using nonlinear partial least squares regression. *J Chem Inf Comput Sci*, 1996, 36: 185—189[DOI]
- 46 张恒喜, 郭基联, 朱家元. 小样本多元数据分析方法及应用. 西安: 西北工业大学出版社, 2002, 30—31

附表 1 所收集的 134 种氨基酸类型与分子结构

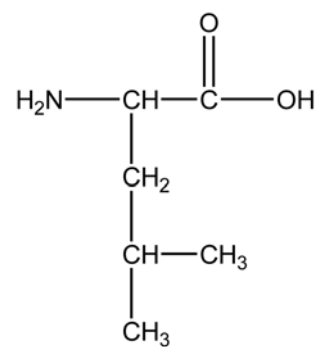
No.	Abbrev	Name	Structure
1	Ala	alanine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_3 \end{array} $
2	Arg	arginine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH} \\ \\ \text{C}=\text{NH} \\ \\ \text{NH}_2 \end{array} $
3	Asn	asparagine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{NH}_2 \end{array} $
4	Asp	aspartic acid	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{OH} \end{array} $

5	Cys	cysteine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array} $
6	Gln	glutamine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{NH}_2 \end{array} $
7	Glu	glutamic acid	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{OH} \end{array} $
8	Gly	glycine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{H} \end{array} $
9	His	histidine	$ \begin{array}{c} \text{O} \\ \parallel \\ \text{H}_2\text{N}-\text{CH}-\text{C}-\text{OH} \\ \\ \text{CH}_2 \\ \\ \text{Imidazole ring} \end{array} $

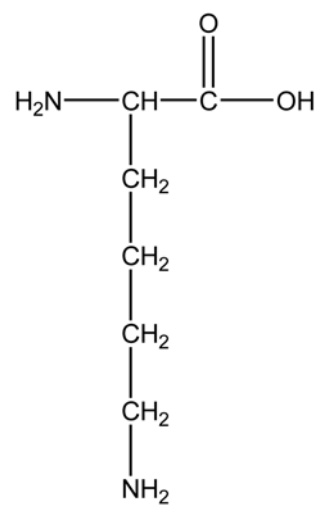
10 Ile isoleucine



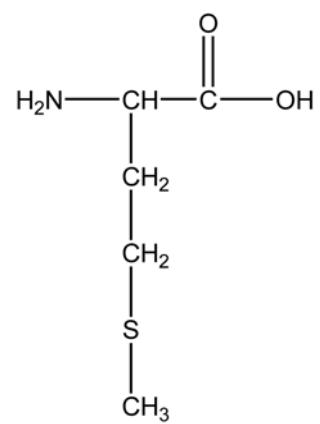
11 Leu leucine



12 Lys lysine

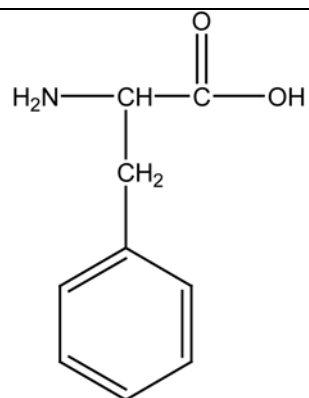


13 Met methionine



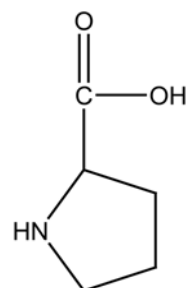
14 Phe

phenylalanine



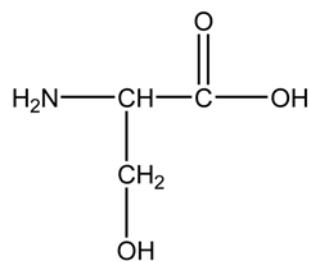
15 Pro

proline



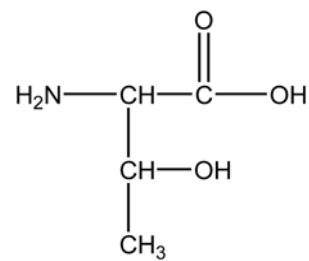
16 Ser

serine



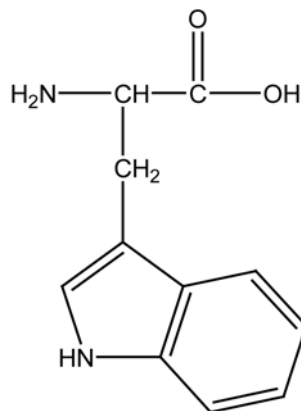
17 Thr

threonine



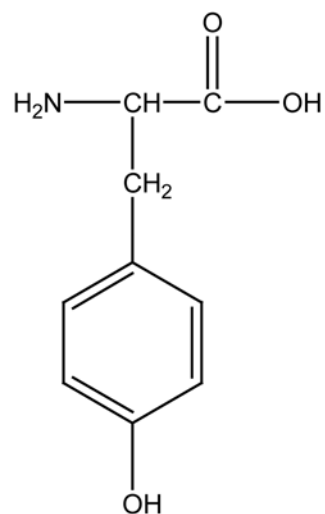
18 Trp

tryptophan



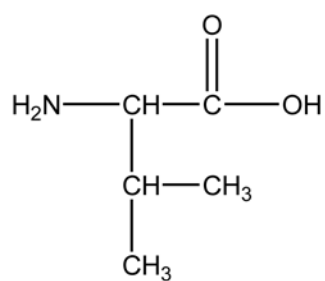
19 Tyr

tyrosine

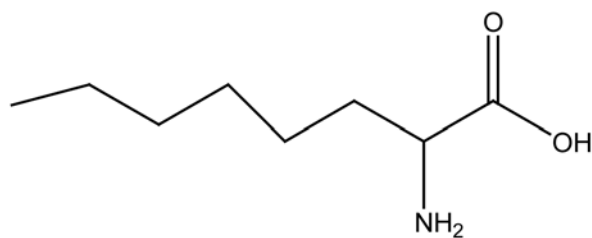


20 Val

valine

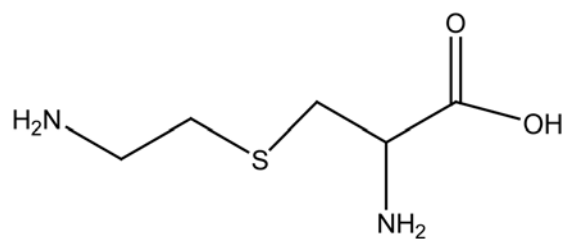


21 Acp

 α -aminocaprylic acid

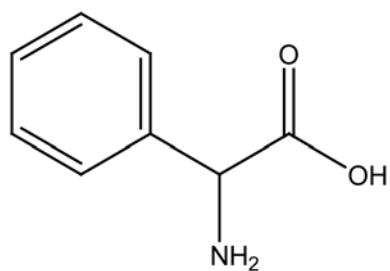
22 Aec

(S)-2-aminoethyl-L-cysteine·HCl



23 Afa

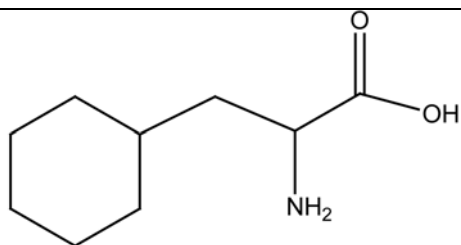
aminophenylacetate



24	Aib	α -aminoisobutyric acid	
25	Ail	alloisoleucine	
26	Alg	L-allylglycine	
27	Aba	α -aminobutyric acid	
28	Aph	<i>p</i> -aminophenylalanine	
29	β -Ala	β -alanine	
30	Brp	<i>p</i> -bromophenylalanine	

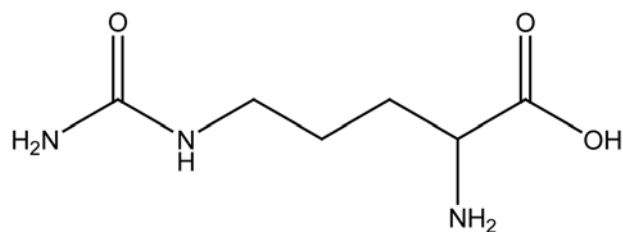
31 Cha

cyclohexylalanine

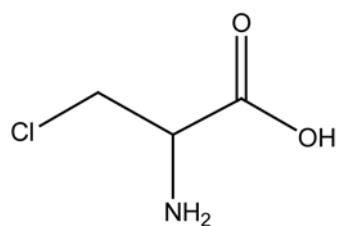


32 Cit

citrulline

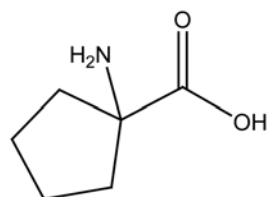


33 Cla

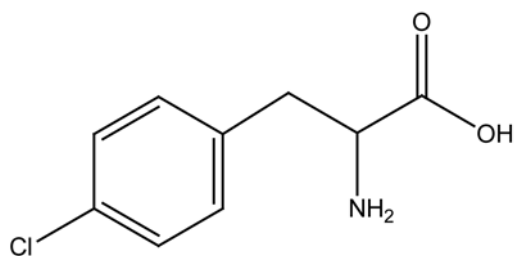
 β -chloroalanine

34 Cle

cycloleucine

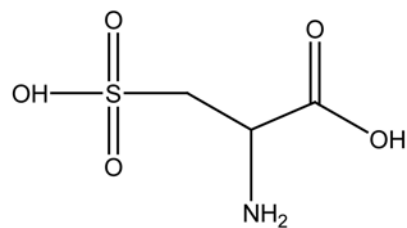


35 Clp

p-chlorophenylalanine

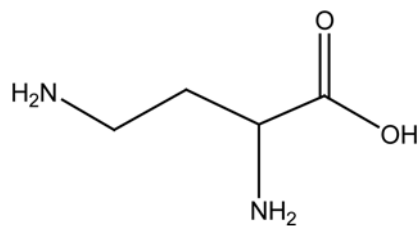
36 Cya

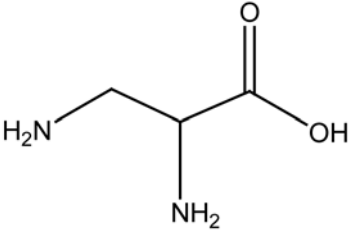
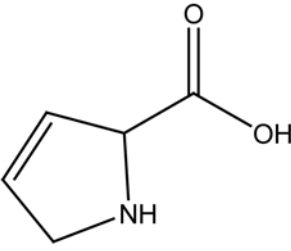
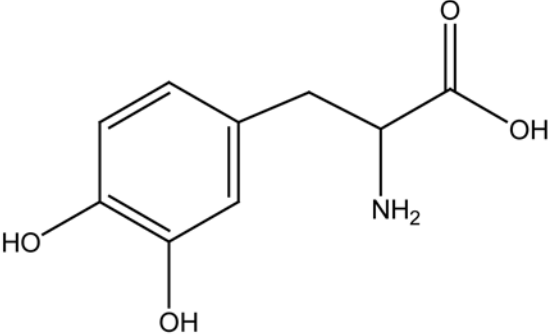
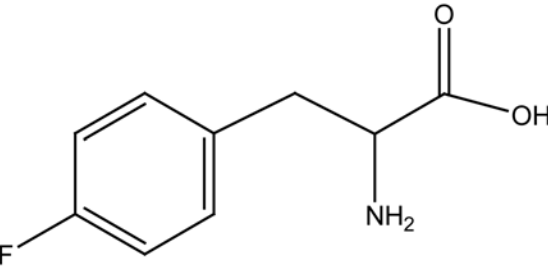
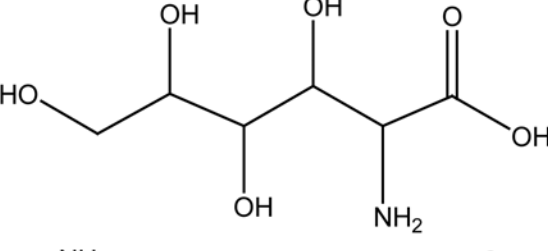
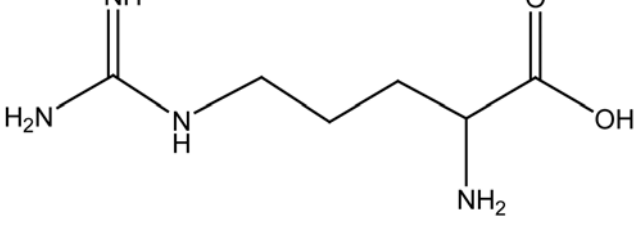
cysteic acid



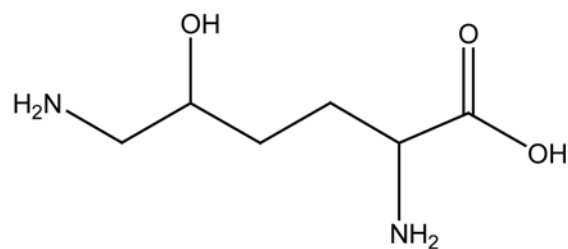
37 Dab

2,4-diaminobutyric acid

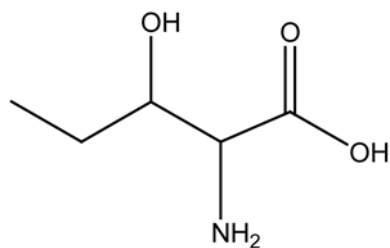


38	Dap	2,3-diaminopropionic acid	
39	Dhp	3,4-dehydropyrroline	
40	Dha	3,4-dihydroxyphenylalanine	
41	Fph	<i>p</i> -fluorophenylalanine	
42	Gaa	D-glucoseaminic acid	
43	Hag	homoarginine	

44 Hly

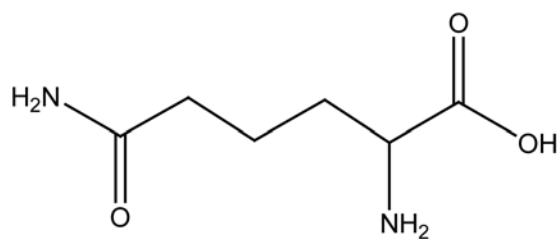
 δ -hydroxylysine-HCl

45 Hnv

DL- β -hydroxynorvaline

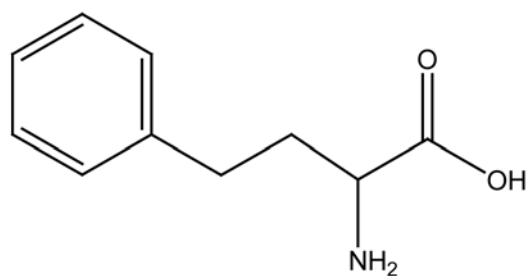
46 Hog

homoglutamine



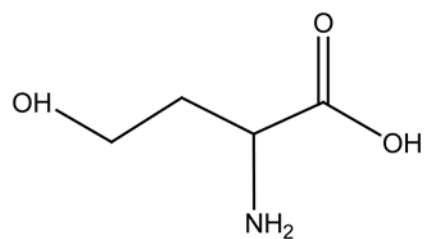
47 Hop

homophenylalanine



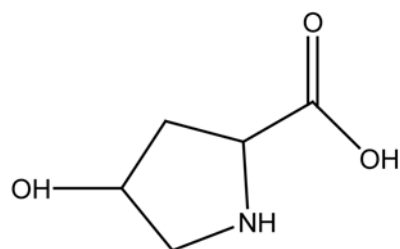
48 Hos

homoserine

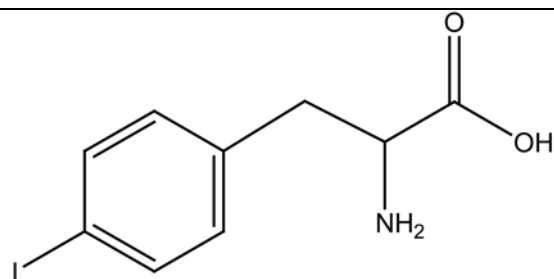


49 Hpr

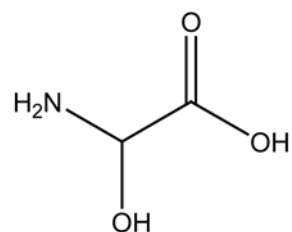
hydroxyproline



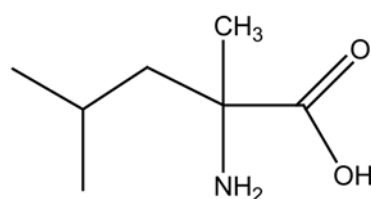
50 Iph *p*-iodophenylalanine



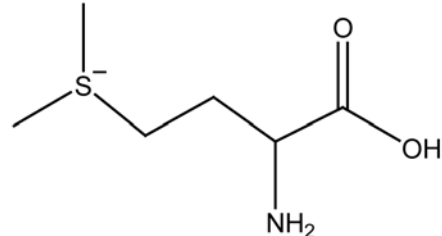
51 Ise isoserine



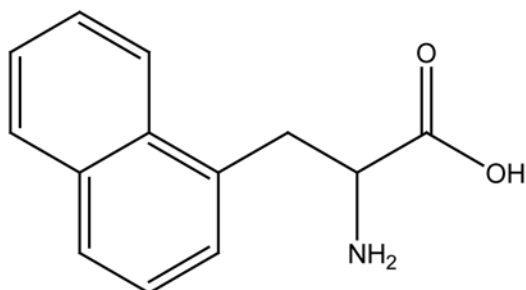
52 Mle α -methylleucine



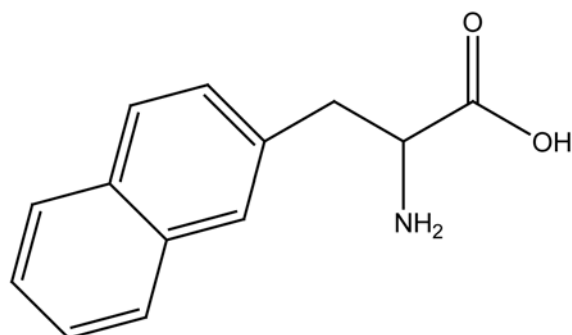
53 Msm DL-methionine-s-methylsulfoniumchloride



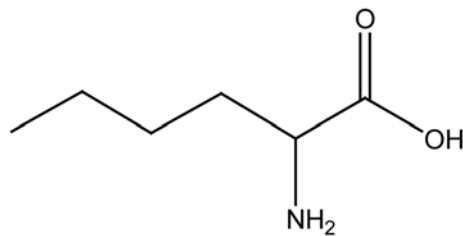
54 1Nala β -(1-naphthyl)alanine



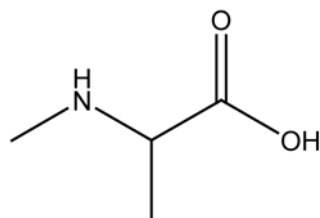
55 2Nala β -(2-naphthyl)alanine



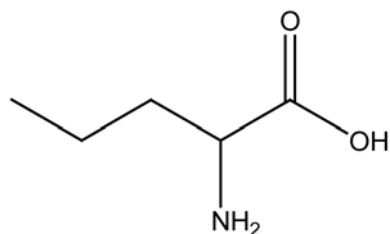
56 Nle norleucine(or 2-aminohexanoic acid)



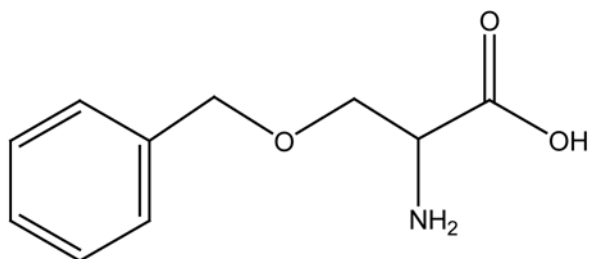
57 Nma *N*-methylalanine



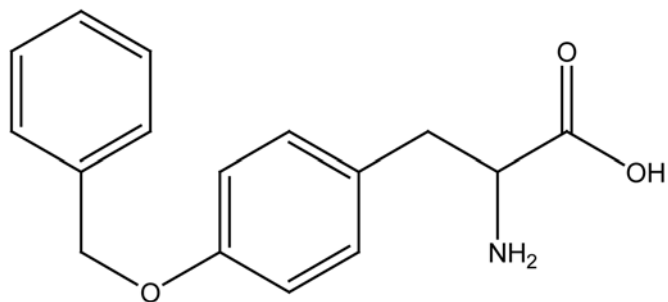
58 Nva norvaline(or 2-aminopentanoic acid)



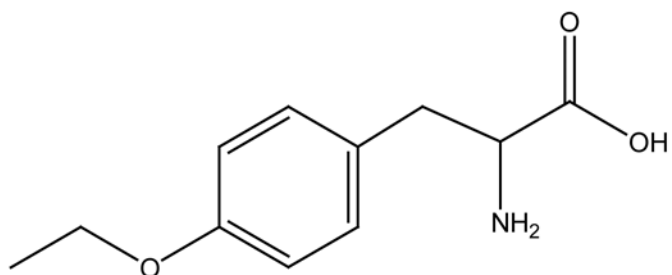
59 Obs *O*-benzylserine

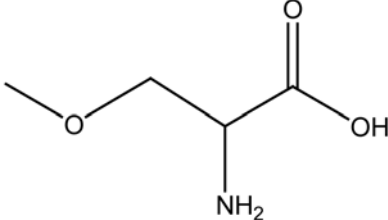
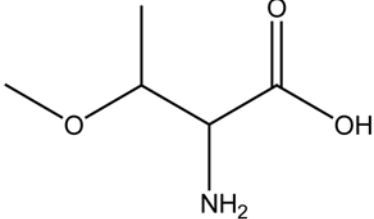
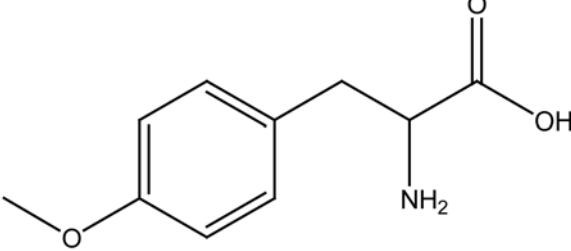
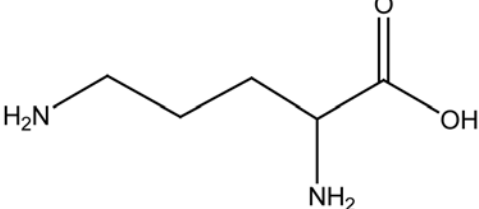
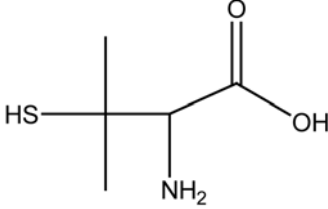
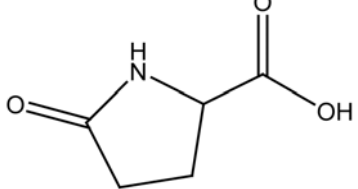
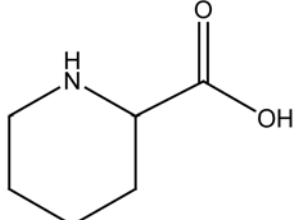


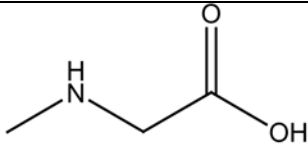
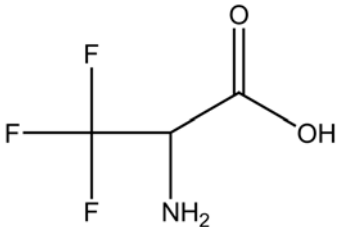
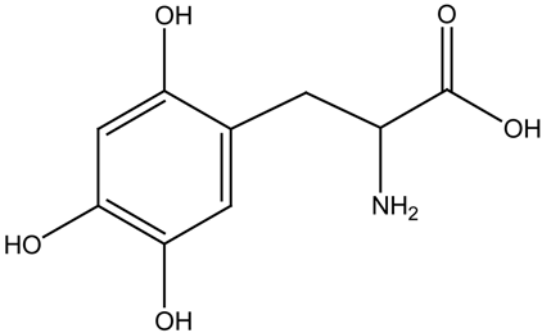
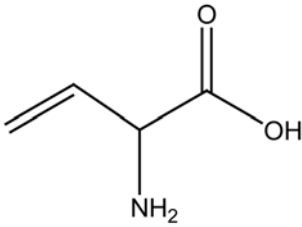
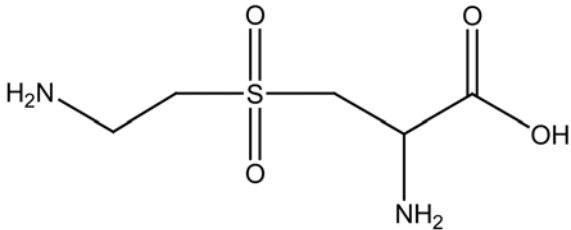
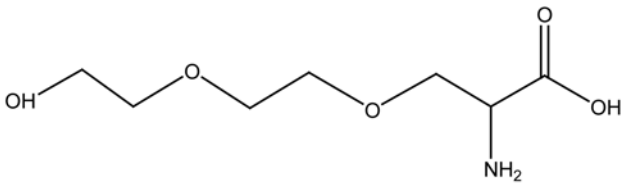
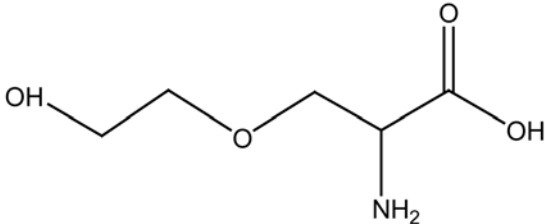
60 Obt *O*-benzyltyrosine

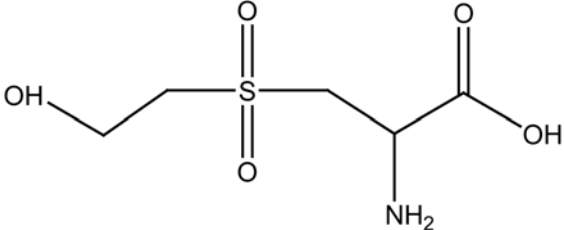
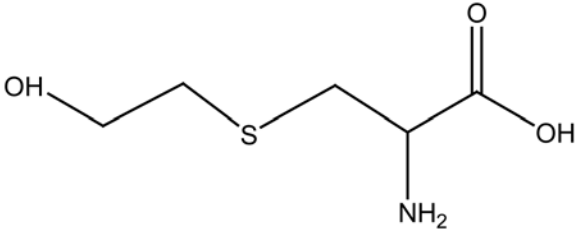
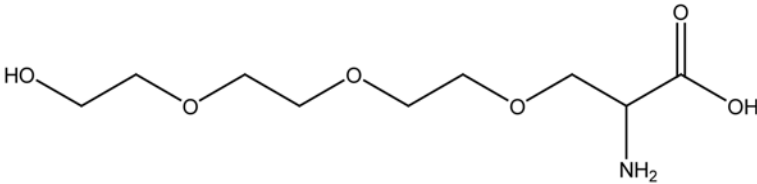
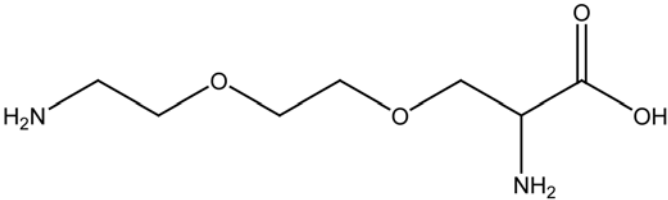
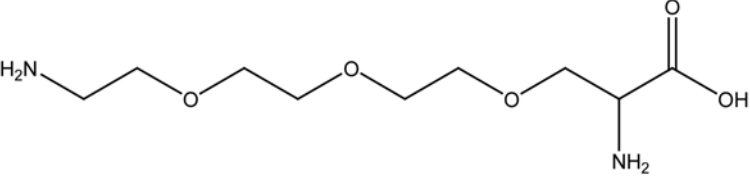
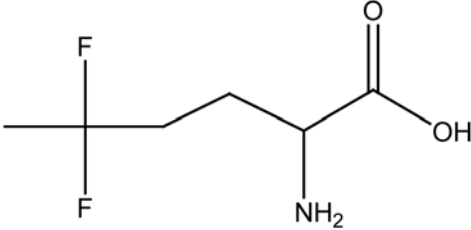
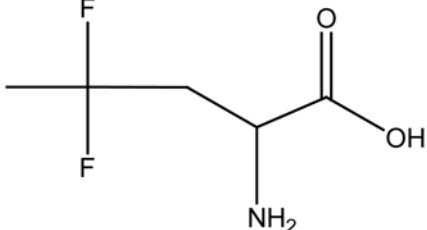


61 Oet *O*-ethyltyrosine

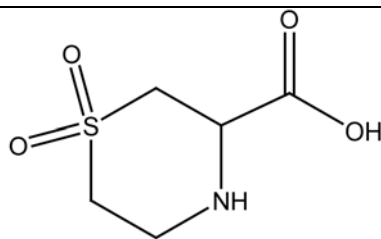


62	Oms	O-methylserine	
63	Omt	O-methylthreonine	
64	Omy	O-methyltyrosine	
65	Orn	ornithine	
66	Pen	penicillamine	
67	Pga	pyroglutamic acid	
68	Pip	pipecolic acid	

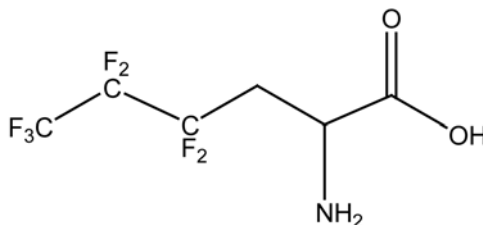
69	Sar	sarcosine	
70	Tfa	3,3,3-trifluoroalanine	
71	Thp	6-hydroxydopa	
72	Vig	L-vinylglycine	
73	Aas	(-)-(2R)-2-amino-3-(2-aminoethylsulfonyl) propanoic acid dihydrochloride	
74	Ahd	(2S)-2-amino-9-hydroxy-4,7-dioxanonanoic acid	
75	Aho	(2S)-2-amino-6-hydroxy-4-oxahexanoic acid	

76	Ahs	(-)-(2 <i>R</i>)-2-amino-3-(2-hydroxyethylsulfonyl)propanoic acid	
77	Ahp	(-)-(2 <i>R</i>)-2-amino-3-(2-hydroxyethylsulfanyl)propanoic acid	
78	Ahd	(2 <i>S</i>)-2-amino-12-hydroxy-4,7,10-trioxadodecanoic acid	
79	Dad	(2 <i>S</i>)-2,9-diamino-4,7-dioxanonanoic acid	
80	Dat	(2 <i>S</i>)-2,12-diamino-4,7,10-trioxadodecanoic acid	
81	Dfn	(<i>S</i>)-5,5-difluoronorleucine	
82	Dfv	(<i>S</i>)-4,4-difluoronorvaline	

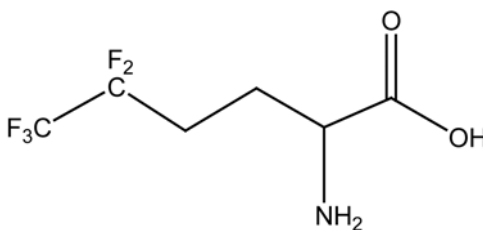
83 Dtc (3*R*)-1,1-dioxo-[1,4]thiaziane-3-carboxylic acid



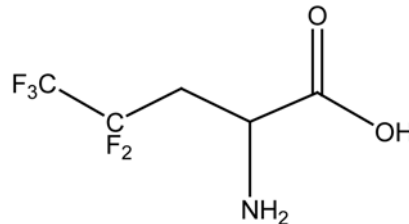
84 Hfn (S)-4,4,5,5,6,6,6-heptafluoronorleucine



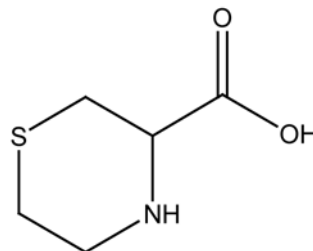
85 Pfn (S)-5,5,6,6,6-pentafluoronorleucine



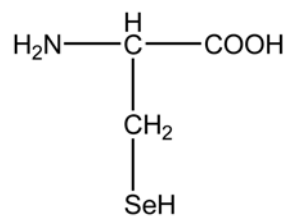
86 Pfv (S)-4,4,5,5,5-pentafluoronorvaline



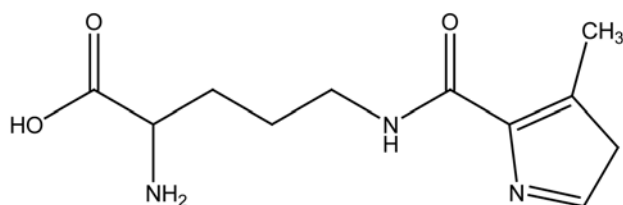
87 Tca (3*R*)-1,4-thiazinane-3-carboxylic acid



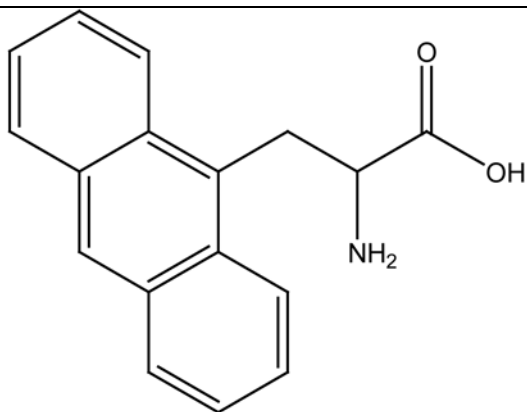
88 Sec selenocysteine



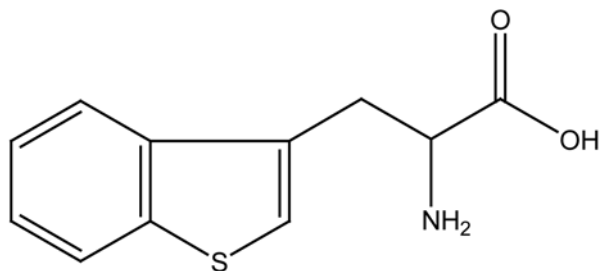
89 Pyl pyrrolysine



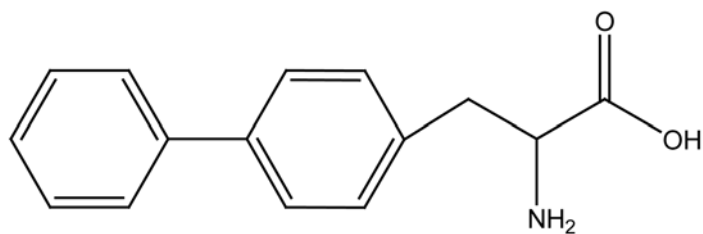
90 Ath β -(9-anthracenyl)alanine



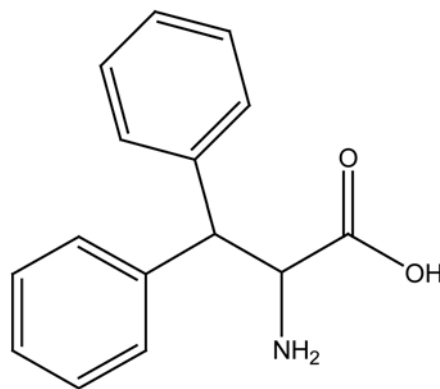
91 Bal β -(3-benzothiienyl)alanine



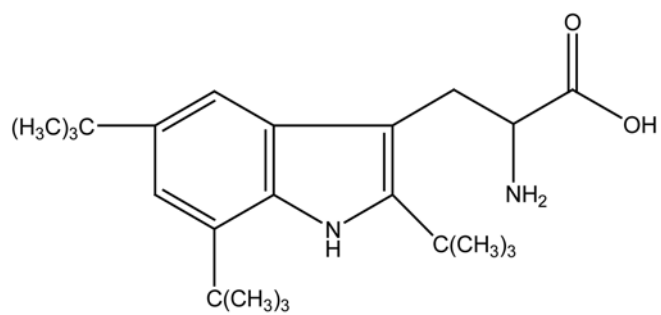
92 Bip β -(4,4'-biphenyl)alanine

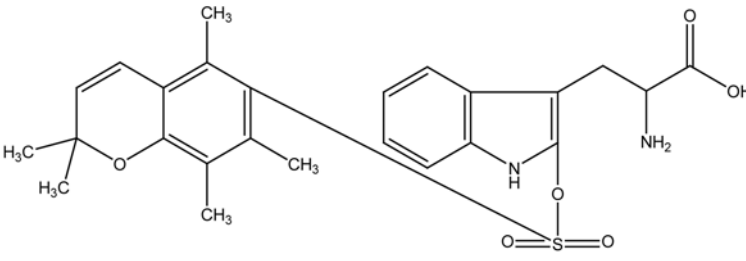
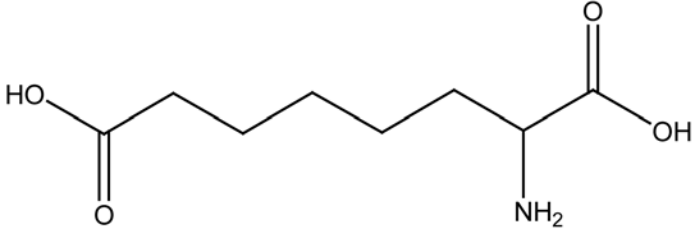
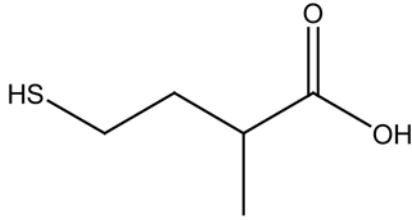
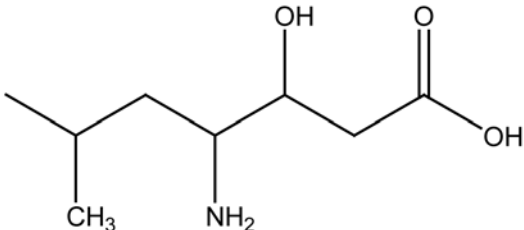
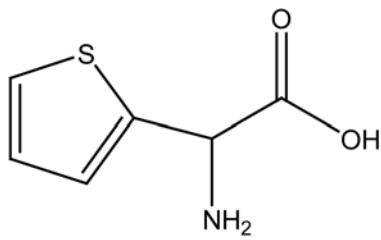
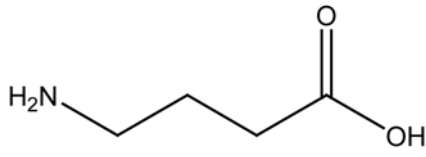
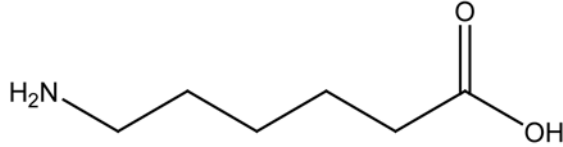


93 Dip β,β -diphenylalanine

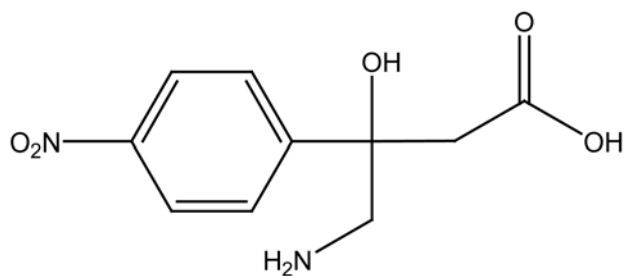


94 Tbt β -[3-(2,5,7-tri-tert-butyl-indolyl)]alanine

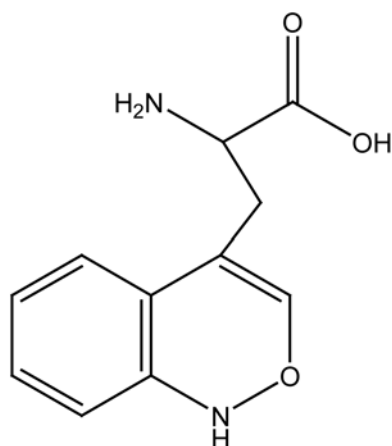


95	Tpc	β -{3-[2-(2,2,5,7,8-pentamethyl-chroman-6-sulfonyl)-indolyl]} alanine	
96	Asu	aminosuberic acid	
97	Hcy	homocysteine	
98	Sta	Statine	
99	Thi	β -(2-Thienyl)alanine	
100	γ -Abu	L- γ -aminobutyric acid	
101	Aca	ϵ -aminocaproic acid	
102	Ach	1-aminocyclohexane-1-carboxylic acid	

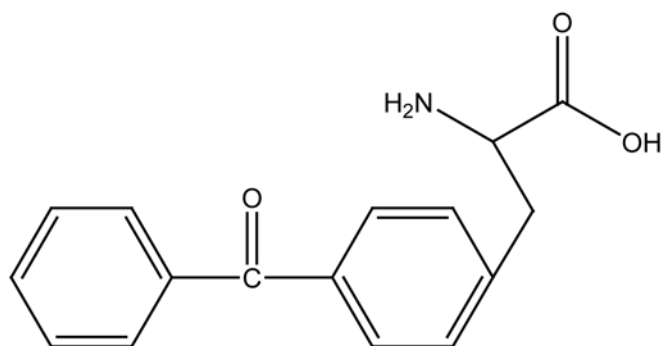
103 Afb β -amino- β -phenyl-p-nitro-L-butyric acid



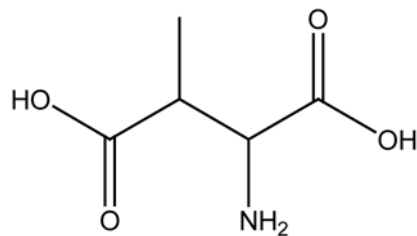
104 Aoq α -amino- β -[4-(1,2-dihydro-2-oxo-quinoliny)] propionic acid



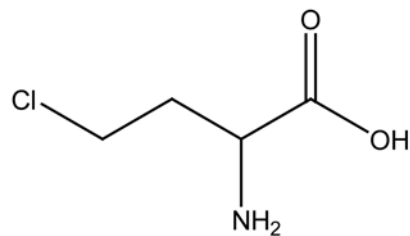
105 Bpa 4'-benzoylphenylalanine

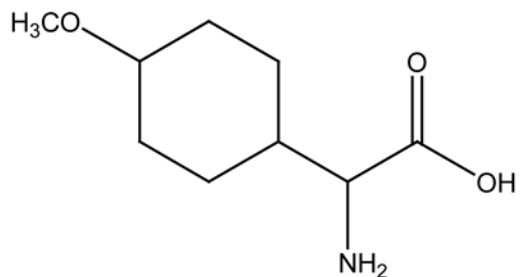
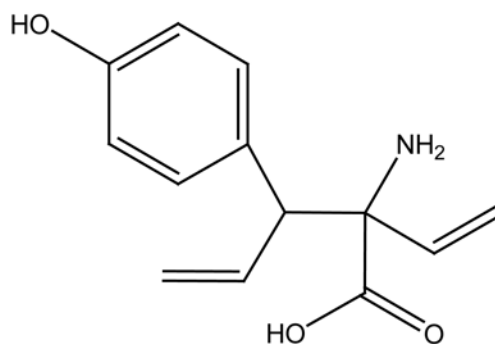


106 Mas β -methyl aspartic acid

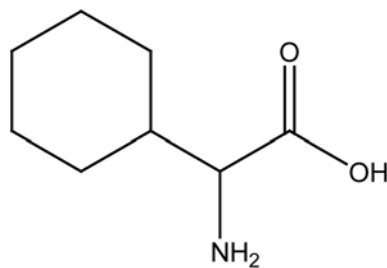


107 Ceg 2-chloroethylglycine

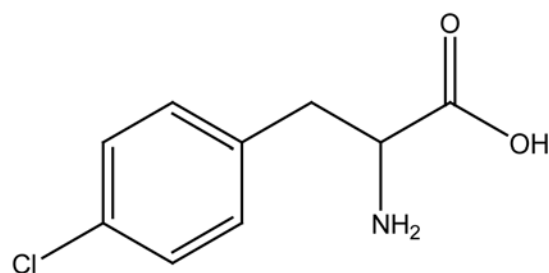
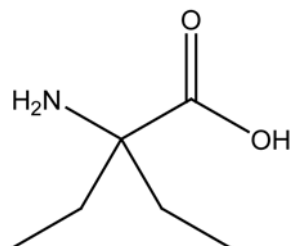


108 Cha β -cyclohexyl(p-methoxyl)-L-alanine109 Dty α,β -divinyltyrosine

110 Chg 2-L-cyclohexylglycine

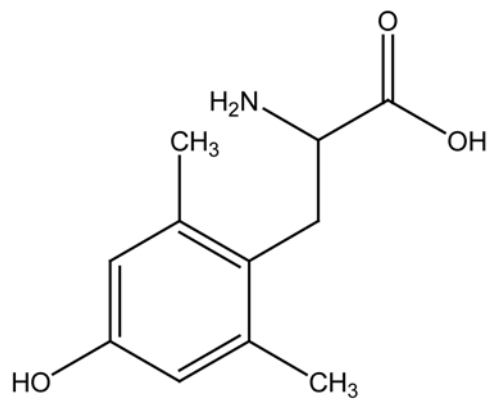


111 Cpa 4-chlorophenylalanine

112 Deg α,α -diethyl glycine

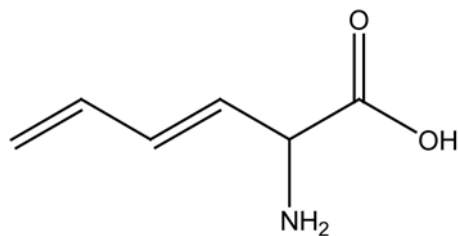
113 Dmt

2',6'-dimethyltyrosine



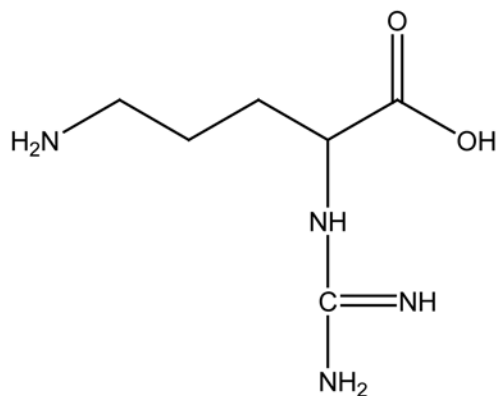
114 Dvg

divinyl glycine



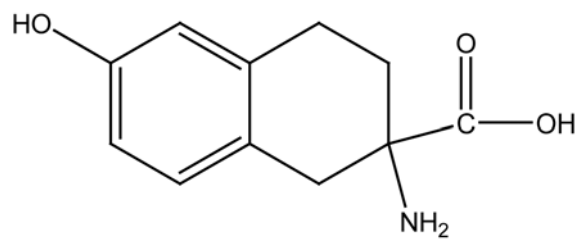
115 Gav

2-guanidine-5-amino-L-n-valeric acid



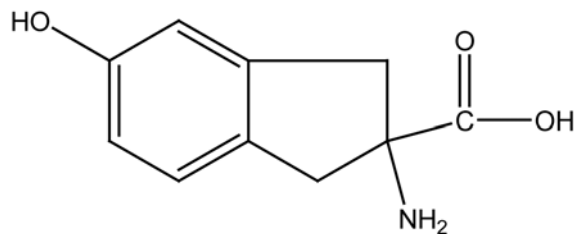
116 Hat

2-amino-6-hydroxytetralin-2-carboxylic acid

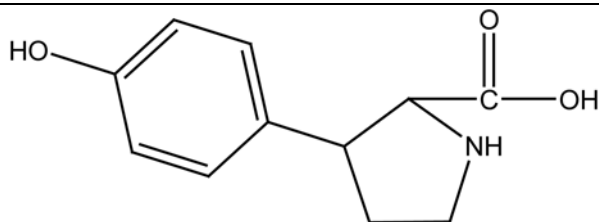


117 Hai

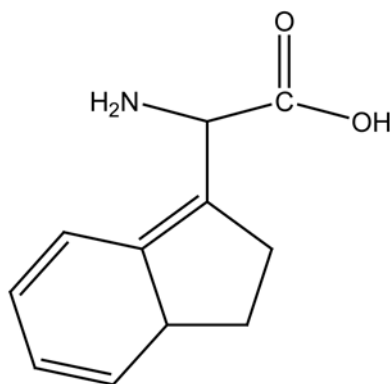
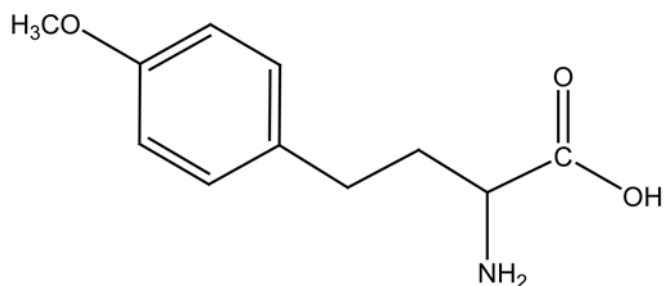
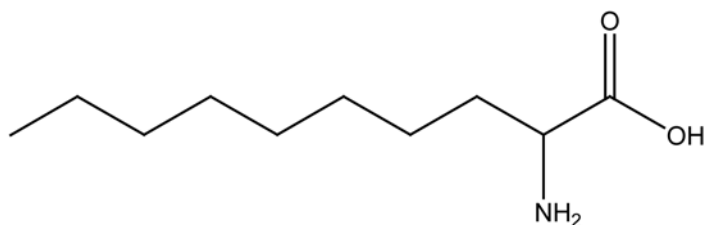
2-amino-5-hydroxyindan-2-carboxylic acid



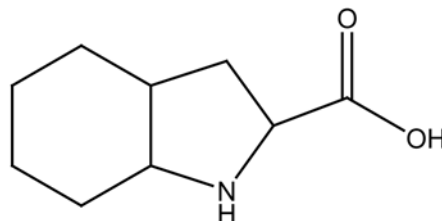
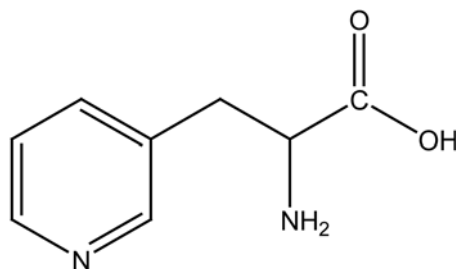
118 Hpp 3-(4'-hydroxyphenyl)proline



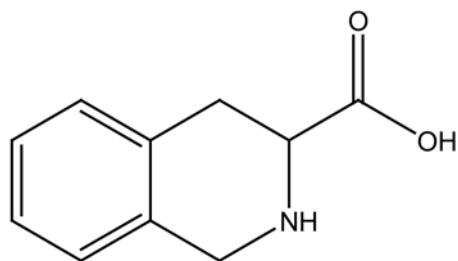
119 Ing 1-indanylglycine

120 Mhp *p*-methoxyhomophenylalanine121 Oct *n*-octylglycine

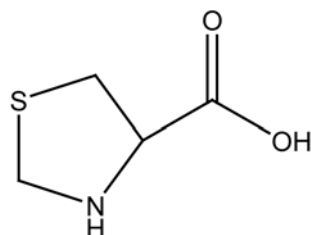
122 Oic octahydroindole-2-carboxylic acid

123 Pal β -pyridylalanine

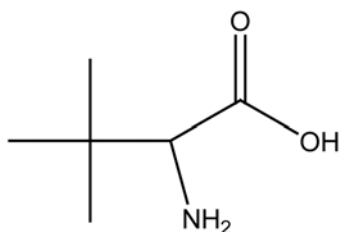
124 Tic 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid



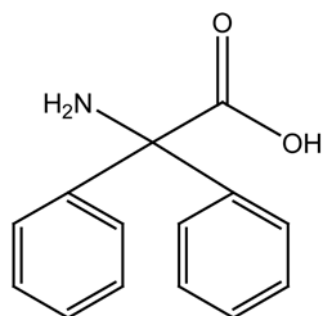
125 Thz L-4-thiazolidine carboxylic acid



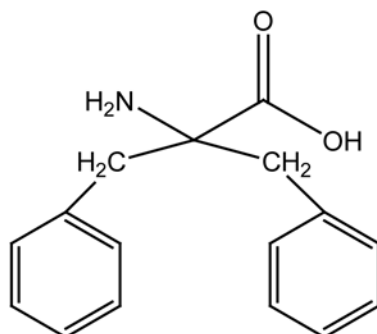
126 Tle L-tert-butylglycine

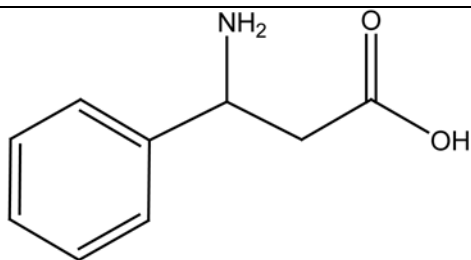
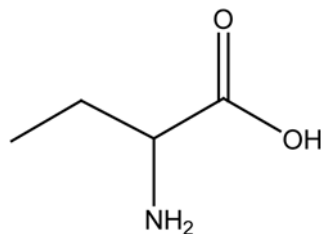


127 Dpg diphenylglycine



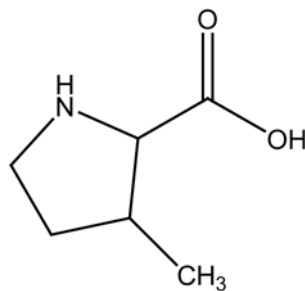
128 Dbz dibenzylglycine



129 β -Phe β -Phenylalanine130 α -Abu α -aminobutyric acid

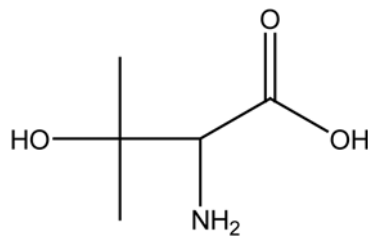
131 Mpr

3-methylproline



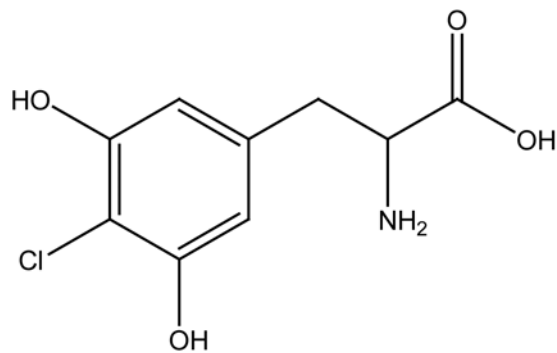
132 Hva

3-hydroxyvaline

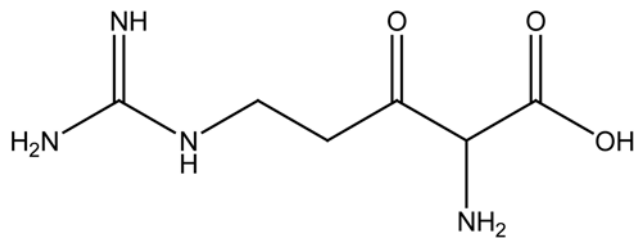


133 Dcp

3,5-dihydroxy-4-chloro-phenylalanine



134 Car

 β -carbonylarginine

附表 2 氨基酸 66 个通用化学结构与分子拓扑描述子

No.	Structural and topological variables	Software	Avr	Std
1	Balaban Index	Chem3D 8.0	66127.977	162161.719
2	Cluster Count	Chem3D 8.0	11.578	4.108
3	Diameter of side chain	Chem3D 8.0	6.593	2.273
4	Molecular Topological Index	Chem3D 8.0	1787.333	2525.753
5	Radius	Chem3D 8.0	3.541	1.157
6	Shape Attribute	Chem3D 8.0	9.674	4.081
7	2D Petitjean Shape Coefficient	Chem3D 8.0	0.861	0.156
8	Sum of Degrees	Chem3D 8.0	22.378	9.464
9	Sum of Valence Degrees	Chem3D 8.0	41.220	15.184
10	Wiener Index	Chem3D 8.0	254.252	343.400
11	KAPPA Shape Index 2	ChemSW 5.0	4.786	1.796
12	Connectivity Index 0	ChemSW 5.0	20.642	6.572
13	Connectivity Index 1	ChemSW 5.0	5.371	1.973
14	Connectivity Index 2	ChemSW 5.0	4.884	2.015
15	Connectivity Index 3	ChemSW 5.0	3.359	1.716
16	Connectivity Index 4	ChemSW 5.0	1.975	1.520
17	Valence Connectivity Index 0	ChemSW 5.0	18.153	6.437
18	Valence Connectivity Index 1	ChemSW 5.0	3.691	1.458
19	Valence Connectivity Index 2	ChemSW 5.0	2.752	1.351
20	Valence Connectivity Index 3	ChemSW 5.0	1.664	0.953
21	Valence Connectivity Index 4	ChemSW 5.0	0.949	0.758
22	Difference Index 0	ChemSW 5.0	2.365	0.867
23	Difference Index 1	ChemSW 5.0	1.682	0.757
24	Difference Index 2	ChemSW 5.0	2.133	0.970
25	Difference Index 3	ChemSW 5.0	1.695	0.924
26	Difference Index 4	ChemSW 5.0	1.027	0.859
27	Randic Index 0	Codessa 2.7	9.005	2.873
28	Randic Index 1	Codessa 2.7	5.372	1.973
29	Randic Index 2	Codessa 2.7	4.884	2.015
30	Randic Index 3	Codessa 2.7	3.359	1.716
31	Kier & Hall Index 0	Codessa 2.7	6.659	2.392
32	Kier & Hall Index 1	Codessa 2.7	3.727	1.483
33	Kier & Hall Index 2	Codessa 2.7	2.794	1.392
34	Kier & Hall Index 3	Codessa 2.7	1.700	0.984
35	Kier Shape Index 1	Codessa 2.7	9.687	2.864
36	Kier Shape Index 2	Codessa 2.7	4.208	1.624
37	Kier Shape Index 3	Codessa 2.7	3.557	1.766
38	Kier Flexibility Index	Codessa 2.7	3.612	1.544
39	Zagreb Index	Dragon 5.2	50.948	21.479
40	Quadratic Index	Dragon 5.2	5.526	3.832
41	Narumi Simple Topological Index (log)	Dragon 5.2	6.383	3.068
42	Narumi Harmonic Topological Indx	Dragon 5.2	1.678	1.497
43	Narumi Harmonic Geometric Indx	Dragon 5.2	1.712	0.174
44	Total Structure Connectivity Index	Dragon 5.2	0.431	0.102
45	Pogliani Index	Dragon 5.2	26.485	7.901
46	Log of Product of Row Sum (PRS)	Dragon 5.2	40.711	19.821
47	Average Vertex Distance Degree	Dragon 5.2	34.926	20.643
48	Mean Square Distance Index	Dragon 5.2	0.330	0.047
49	Schultz Molecular Topological Index	Dragon 5.2	941.852	970.611
50	Gutman Molecular Topological Index	Dragon 5.2	813.642	972.088
51	Xu Index	Dragon 5.2	11.551	3.795
52	Superpendentic Index	Dragon 5.2	91.663	280.189
53	Harary H Index	Dragon 5.2	17.248	7.617

54	Square Reciprocal Distance Sum Index	Dragon 5.2	28.068	14.874
55	First Mohar Index TI1	Dragon 5.2	-2.991	16.266
56	Second Mohar Index TI2	Dragon 5.2	2.666	0.913
57	Hyper-Distance-Path Index	Dragon 5.2	693.578	870.837
58	Detour Index	Dragon 5.2	348.444	435.631
59	Balaban Distance Connectivity Index (J)	Dragon 5.2	2.830	0.695
60	Maximal Electrotopological Negative Variation	Dragon 5.2	2.880	0.839
61	Maximal Electrotopological Positive Variation	Dragon 5.2	3.359	0.521
62	Molecular Electrotopological Variation	Dragon 5.2	18.892	8.083
63	E-state Topological Parameter	Dragon 5.2	221.514	1413.845
64	Kier Symmetry Index	Dragon 5.2	103.676	755.175
65	Balaban Centric Index	Dragon 5.2	26.948	15.951
66	Lopping Centric Index	Dragon 5.2	1.56474	8.765

附表 3 氨基酸 66 个通用化学结构与分子拓扑描述子具体数据

No.	Abbrev	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	Ala	1721	6	3	212	2	4.17	0.5	10	22	29	2.22	12.15	2.64	2.49	1.33	0	10.51	1.63	1.13	0.39	0	1.64	1.02
2	Arg	64589	12	8	1673	4	10.08	1	22	42	247	6.51	23.56	5.54	4.9	2.83	1.35	20.71	3.6	2.43	1.32	0.67	2.85	1.94
3	Asn	13687	9	5	624	3	7.11	0.67	16	36	96	3.92	15.44	4.04	3.85	1.98	1.02	12.7	2.3	1.62	0.74	0.3	2.74	1.73
4	Asp	13687	9	5	597	3	7.11	0.67	16	38	96	3.92	14.44	4.04	3.85	1.98	1.02	11.57	2.24	1.54	0.71	0.28	2.87	1.8
5	Cys	3821	7	4	301	2	5.14	1	12	23.56	46	3.06	12.86	3.18	2.63	1.78	0.47	11.56	2.4	1.49	0.91	0.23	1.31	0.78
6	Gln	24242	10	6	894	3	8.1	1	18	38	136	4.76	18.15	4.54	4.19	2.34	0.93	15.41	2.8	1.95	1.02	0.42	2.74	1.73
7	Glu	24242	10	6	859	3	8.1	1	18	40	136	4.76	17.15	4.54	4.19	2.34	0.93	14.28	2.74	1.88	0.99	0.4	2.87	1.8
8	Gly	712	5	3	130	2	3.2	0.5	8	20	18	2.25	9.28	2.27	1.8	0.82	0	7.64	1.19	0.6	0.17	0	1.64	1.08
9	His	19798	11	6	1155	3	9.09	1	22	42	165	4.13	17.27	5.2	4.61	3.32	2.04	14.82	3.16	2.21	1.3	0.72	2.45	2.04
10	Ile	13107	9	5	678	3	7.11	0.67	16	28	92	3.92	20.44	4.09	3.49	2.59	1.03	18.79	3.08	2.25	1.54	0.5	1.64	1.02
11	Leu	13687	9	5	705	3	7.11	0.67	16	28	96	3.92	20.44	4.04	3.85	1.98	1.02	18.79	3.02	2.57	1.04	0.61	1.64	1.02
12	Lys	25452	10	7	1008	4	8.1	0.75	18	32	143	5.76	21.98	4.68	3.72	2.4	1.05	19.92	3.37	2.23	1.29	0.63	2.07	1.31
13	Met	14521	9	6	694	3	7.11	1	16	26.67	102	4.84	18.28	4.18	3.36	2.15	0.86	17.15	4.04	2.71	1.59	0.86	1.13	0.14
14	Phe	30269	12	7	1573	4	10.08	0.75	24	42	212	4.89	19.97	5.7	4.96	3.57	2.22	17.6	3.72	2.63	1.62	0.96	2.37	1.98
15	Pro	3936	8	4	464	2	6.125	1	16	28	62	2.52	14.98	3.8	3.29	2.34	1.62	13.55	2.77	1.99	1.36	0.9	1.43	1.04
16	Ser	3821	7	4	301	2	5.14	1	12	28	46	3.06	12.86	3.18	2.63	1.78	0.47	10.66	1.77	1.13	0.51	0.08	2.2	1.41
17	Thr	7206	8	4	438	2	6.13	1	14	30	65	3.11	15.73	3.55	3.35	2.1	0.77	13.53	2.22	1.6	0.8	0.21	2.2	1.33
18	Trp	58498	15	8	2710	4	13.07	1	32	54	369	4.89	22.84	7.18	6.5	5.29	3.9	20.1	4.72	3.51	2.43	1.62	2.73	2.47
19	Tyr	44867	13	8	1876	4	11.08	1	26	48	268	5.02	20.84	6.09	5.58	3.98	2.36	17.97	3.86	2.82	1.71	0.97	2.87	2.24
20	Val	7206	8	4	478	2	6.13	1	14	26	65	3.11	17.73	3.55	3.34	2.1	0.77	16.09	2.54	2.11	1	0.29	1.64	1.02
21	Acp	42210	11	8	1427	4	9.09	1	20	32	194	6.69	25.69	5.18	4.07	2.65	1.23	24.05	4.17	2.79	1.69	0.91	1.64	1.02
22	Aec	25452	10	7	958	4	8.1	0.75	18	30.67	143	5.76	19.98	4.68	3.72	2.4	1.05	18.43	4.09	2.95	1.99	1.09	1.55	0.59
23	Afa	18715	11	6	1161	3	9.09	1	22	40	156	4.13	17.27	5.22	4.5	3.44	2.2	14.9	3.25	2.23	1.42	0.8	2.37	1.97
24	Aib	3493	7	3	310	2	5.14	0.5	12	24	42	1.85	15.08	2.94	3.52	1.73	0	13.43	1.97	2.03	0.55	0	1.64	0.98
25	Ail	13678	9	5	705	3	7.11	0.67	16	28	96	3.92	20.44	4.04	3.85	1.98	1.02	18.79	3.02	2.57	1.04	0.61	1.64	1.02
26	Alg	7751	8	5	513	3	6.13	0.67	14	28	70	3.94	15.57	3.68	3.01	1.88	0.79	13.5	2.27	1.47	0.74	0.28	2.07	1.41
27	Aba	3821	7	4	337	2	5.14	1	12	24	46	3.06	14.86	3.18	2.63	1.78	0.47	13.22	2.17	1.35	0.76	0.17	1.64	1.02
28	Aph	44867	13	8	1933	4	11.08	1	26	46	268	5.02	21.84	6.09	5.58	3.98	2.36	19.1	3.92	2.89	1.75	0.99	2.74	2.17
29	β-Ala	1899	6	4	229	2	4.17	1	10	22	32	3.2	11.99	2.77	2.18	0.87	0.58	10.35	1.69	0.93	0.36	0.12	1.64	1.08
30	Brp	44867	13	8	1819	4	11.08	1	26	44	268	5.02	19.84	6.09	5.58	3.98	2.36	18.49	4.61	3.69	2.21	1.23	1.36	1.48
31	Cha	30269	12	7	1573	4	10.08	0.75	24	36	212	4.89	25.97	5.7	4.96	3.57	2.22	24.33	4.68	3.68	2.63	1.88	1.64	1.01
32	Cit	64589	12	8	1619	4	10.08	1	22	44	247	6.51	22.56	5.54	4.9	2.83	1.35	19.62	3.55	2.38	1.31	0.66	2.94	1.98
33	Cla	3821	7	4	283	2	5.14	1	12	23.78	46	3.06	11.86	3.18	2.63	1.78	0.47	10.35	2.26	1.41	0.82	0.2	1.51	0.92
34	Cle	6505	9	4	608	2	7.11	1	18	30	81	2.32	17.91	4.15	4.02	3.14	1.76	16.26	3.17	2.67	1.92	1.22	1.65	0.98
35	Clp	44867	13	8	1819	4	11.08	1	26	43.78	268	5.02	19.84	6.09	5.58	3.98	2.36	17.66	4.2	3.21	1.94	1.09	2.19	1.89
36	Cya	22119	10	5	787	3	8.1	0.67	18	40.67	124	3.41	15.36	4.33	4.99	2.07	1.2	12.54	3.55	2.89	1.33	0.77	2.82	0.78
37	Dab	7751	8	5	488	3	6.13	0.67	14	28	70	3.94	16.57	3.68	3.01	1.88	0.79	14.5	2.37	1.52	0.79	0.3	2.07	1.31

38	Dap	3821	7	4	319	2	5.14	1	12	26	46	3.06	13.86	3.18	2.63	1.78	0.47	11.79	1.87	1.18	0.57	0.1	2.07	1.31
39	Dhp	3936	8	4	464	2	6.13	1	16	30	62	2.52	12.98	3.8	3.29	2.34	1.62	11.29	2.43	1.65	1.05	0.64	1.69	1.37
40	Dha	62340	14	8	2160	4	12.07	1	28	54	321	5.19	21.72	6.5	6.09	4.55	2.67	18.34	4	2.97	1.81	1.04	3.37	2.51
41	Fph	44867	13	8	1819	4	11.08	1	26	50	268	5.02	19.84	6.09	5.58	3.98	2.36	16.9	3.82	2.78	1.69	0.96	2.94	2.27
42	Gaa	78450	13	7	1534	4	11.08	0.75	24	52	254	5.67	23.59	5.91	5.21	4.25	1.98	19.74	3.55	2.6	1.58	0.69	3.86	2.36
43	Hag	64589	12	8	1673	4	10.08	1	22	42	247	6.51	23.56	5.54	4.9	2.83	1.35	20.71	3.6	2.43	1.32	0.67	2.85	1.94
44	Hly	39213	11	7	1204	4	9.09	0.75	20	38	180	5.63	22.85	5.07	4.31	2.95	1.08	20.23	3.44	2.41	1.4	0.61	2.62	1.64
45	Hnv	13107	9	5	632	3	7.11	0.67	16	32	92	3.92	18.44	4.09	3.49	2.59	1.03	16.24	2.76	1.84	1.12	0.42	2.2	1.33
46	Hog	40528	11	7	1236	4	9.09	0.75	20	40	186	5.63	20.85	5.04	4.55	2.58	1.18	18.12	3.3	2.3	1.26	0.62	2.74	1.73
47	Hop	46923	13	8	2076	4	11.08	1	26	44	280	5.67	22.68	6.2	5.3	3.89	2.31	20.31	4.22	2.97	1.9	1.09	2.37	1.98
48	Hos	7751	8	5	463	3	6.13	0.67	14	30	70	3.94	15.57	3.68	3.01	1.88	0.79	13.37	2.27	1.46	0.76	0.26	2.2	1.41
49	Hpr	6981	9	5	597	3	7.11	0.67	18	34	87	2.72	15.85	4.2	3.92	2.68	1.86	13.87	2.84	2.19	1.41	0.91	1.98	1.36
50	Iph	44867	13	8	1819	4	11.08	1	26	43.49	268	5.02	19.84	6.09	5.58	3.98	2.36	19.03	4.88	4	2.39	1.32	0.82	1.21
51	Ise	1721	6	3	188	2	4.17	0.5	10	26	29	2.22	10.15	2.64	2.49	1.33	0	7.96	1.31	0.78	0.25	0	2.2	1.33
52	Mle	21380	10	5	891	3	8.1	0.67	18	30	120	3.41	23.36	4.36	4.74	2.5	1.29	21.72	3.38	3.32	1.32	0.94	1.64	0.98
53	Msm	24242	10	6	970	3	8.1	1	18	27.67	136	4.76	22.15	4.54	4.19	2.34	0.93	21.15	5.27	4.79	2.2	1.21	1	-0.73
54	1Nala	78002	16	8	3253	4	14.07	1	34	56	434	5.56	24.54	7.68	6.86	5.54	4.11	21.76	5.13	3.79	2.64	1.79	2.78	2.55
55	2Nala	82281	16	9	3425	5	14.06	0.8	34	56	458	5.56	24.54	7.66	6.94	5.49	3.98	21.76	5.13	3.83	2.59	1.71	2.78	2.54
56	Nle	14521	9	6	748	3	7.11	1	16	28	102	4.84	20.28	4.18	3.36	2.15	0.86	18.63	3.17	2.09	1.19	0.55	1.64	1.02
57	Nma	3821	7	4	338	2	5.14	1	12	24	46	3.06	14.86	3.18	2.63	1.78	0.47	13.43	2.08	1.35	0.8	0.12	1.43	1.1
58	Nva	7751	8	5	513	3	6.125	0.67	14	26	70	3.94	17.57	3.68	3.01	1.88	0.79	15.92	2.67	1.73	0.92	0.43	1.64	1.02
59	Obs	70169	14	9	2596	5	12.07	0.8	28	50	361	6.48	23.39	6.7	5.66	4.13	2.54	20.72	4.3	2.95	1.73	0.95	2.67	2.4
60	Obt	273728	20	13	7271	7	18.05	0.86	42	70	988	8.44	31.37	9.65	8.43	6.52	4.5	28.03	6.39	4.57	2.94	1.8	3.34	3.26
61	Oet	93384	15	10	3007	5	13.07	1	30	52	420	6.55	26.26	7.13	6.13	4.51	2.96	23.64	4.83	3.23	2.01	1.23	2.62	2.3
62	Oms	7751	8	5	473	3	6.13	0.67	14	30	70	3.94	15.57	3.68	3.01	1.88	0.79	13.63	2.15	1.4	0.66	0.25	1.94	1.53
63	Omt	13107	9	5	634	3	7.11	0.67	16	32	92	3.92	18.44	4.09	3.49	2.59	1.03	16.5	2.6	1.8	1.16	0.35	1.94	1.49
64	Omy	65347	14	9	2388	5	12.07	0.8	28	50	337	5.78	23.55	6.63	5.75	4.39	2.67	20.93	4.25	3	1.93	1.1	2.62	2.38
65	Orn	14521	9	6	715	3	7.11	1	16	30	102	4.84	19.28	4.18	3.36	2.15	0.86	17.21	2.87	1.88	1.04	0.46	2.07	1.31
66	Pen	12267	9	4	591	2	7.11	1	16	27.56	86	2.72	18.65	3.85	4.4	2.37	1	17.35	3.01	3.62	1.3	0.41	1.31	0.85
67	Pga	6981	9	5	598	3	7.11	0.67	18	36	87	2.72	13.85	4.2	3.92	2.68	1.86	11.76	2.72	2	1.29	0.8	2.1	1.48
68	Pip	7074	9	5	657	3	7.11	0.67	18	30	88	3.24	17.69	4.3	3.64	2.59	1.8	16.26	3.27	2.34	1.61	1.06	1.43	1.04
69	Sar	1899	6	4	232	2	4.17	1	10	22	32	3.2	11.99	2.77	2.18	0.87	0.58	10.56	1.63	0.92	0.33	0.15	1.43	1.14
70	Tfa	12267	9	4	437	2	7.11	1	16	46	86	2.72	11.65	3.85	4.4	2.37	1	8.14	1.91	1.36	0.62	0.14	3.51	1.95
71	Thp	83360	15	8	2437	4	13.07	1	30	60	374	5.37	22.59	6.91	6.62	5.01	3.03	18.71	4.14	3.11	1.93	1.1	3.87	2.78
72	Vig	3821	7	4	337	2	5.14	1	12	26	46	3.06	12.86	3.18	2.63	1.78	0.47	10.79	1.79	1.1	0.54	0.1	2.07	1.39
73	Aas	55112	12	7	1426	4	10.08	0.75	22	42.67	211	4.89	21.78	5.39	5.34	2.94	1.89	18.64	3.73	2.9	1.61	0.91	3.14	1.66
74	Ahd	101858	13	10	2110	5	11.08	1	24	48	330	8.59	25.1	6.18	4.78	3.15	1.58	22.31	3.93	2.36	1.22	0.57	2.8	2.25
75	Aho	25452	10	7	916	4	8.1	0.75	18	38	143	5.76	18.98	4.68	3.72	2.4	1.05	16.49	2.85	1.74	0.85	0.35	2.5	1.83
76	Ahs	55112	12	7	1376	4	10.08	0.75	22	44.67	211	4.89	20.78	5.39	5.34	2.94	1.87	17.51	3.64	2.84	1.57	0.85	3.27	1.75
77	Ahp	25452	10	7	916	4	8.1	0.75	18	32.67	143	5.76	18.98	4.68	3.72	2.4	1.05	17.3	4	2.89	1.91	1.04	1.68	0.68
78	Ahd	301117	16	13	4063	7	14.06	0.86	30	58	634	11.48	31.23	7.68	5.84	3.9	2.11	28.13	5.01	2.97	1.59	0.79	3.09	2.67
79	Dad	101858	13	10	2185	5	11.08	1	24	46	330	8.59	26.1	6.18	4.78	3.15	1.58	23.44	4.02	2.42	1.24	0.58	2.66	2.16
80	Dat	301117	16	13	4180	7	14.06	0.86	30	56	634	11.48	32.23	7.68	5.84	3.9	2.11	29.26	5.1	3.03	1.62	0.8	2.96	2.58
81	Dfn	37513	11	6	1044	3	9.09	1	20	44	172	4.13	20.07	4.83	5.32	2.49	0.99	17.18	3.19	2.55	1.21	0.53	2.89	1.64
82	Dfv	22119	10	5	739	3	8.1	0.67	18	42	124	3.41	17.36	4.33	4.99	2.07	1.2	14.47	2.69	2.22	0.9	0.47	2.89	1.64
83	Dtc	17767	11	5	1029	3	9.09	0.67	22	40.67	148	3.16	17.48	5.01	5.26	3.19	2.33	14.98	3.63	3.02	1.94	1.28	2.5	1.38
84	Hfn	197331	16	7	1948	4	14.06	0.75	30	84	415	4.7	19.36	6.83	8.18	5.54	2.31	13.36	3.63	2.93	1.61	0.7	6	3.19
85	Pfn	115100	14	7	1620	4	12.07	0.75	26	68	319	4.68	19.57	6.08	6.89	4.39	1.3	14.81	3.51	2.69	1.5	0.6	4.75	2.57
86	Pfv	75443	13	6	1195	3	11.08	1	24	66	244	4.02	16.86	5.58	6.56	4	1.35	12.11	3.01	2.35	1.21	0.48	4.75	2.57
87	Tca	7074	9	5	613	3	7.11	0.67	18	28.67	88	3.24	15.69	4.3	3.64	2.59	1.8	14.78	3.99	3.07	2.28	1.65	0.91	0.31
88	Sec	3821	7	4	301	2	5.14	1	12	23.88	46	3.06	12.86	3.18	2.63	1.78	0.47	12.55	3.11	1.9	1.34	0.41	0.31	0.08
89	Pyl	244813	18	11	5368	6	16.06	0.83	36	62	762	8.23	32.54	8.52	7.41	5.62	3.68	29.48	6.01	4.34	2.78	1.63	3.07	2.51
90	Ath	162136	20	8	5637	4	18.05	1	44	70	744	6.41	29.11	9.66	8.77	7.42	5.94	25.91	6.54	4.95	3.62	2.64	3.2	3.12
91	Bal	58498	15	8	2661	4	13.07	1	32	50.67	369	4.89	21.84	7.18	6.5	5.29	3.9	19.83	5.49	4.31	3.33	2.49	2.01	1.69

92	Bip	154081	18 11	5118	6 16.06	0.83	38	62	683	6.96	27.56	8.66	7.64	6.05	4.3	24.91	5.79	4.21	2.84	1.81	3.04	2.87
93	Dip	129849	18 8	4374	4 16.06	1	38	62	575	6.96	27.96	8.7	7.48	6.05	4.73	24.86	5.78	4.16	2.96	2.07	3.1	2.92
94	Tbt	809808	27 9	12316	5 25.04	0.8	56	78	1623	7.68	56.95	12.03	14.16	8.99	6.72	54.37	9.72	10.82	5.18	3.53	2.58	2.31
95	Tpc	1630081	34 14	23933	7 32.03	1	74	118.67	3270	10.05	53	15.82	16.36	13.01	10.9	48.37	11.99	10.33	7.4	5.48	4.62	3.83
96	Asu	98896	13 9	2103	5 11.08	0.8	24	46	320	7.41	25.27	6.04	5.25	3.08	1.53	22.4	4.24	2.94	1.72	0.94	2.87	1.8
97	Hcy	7751	8 5	463	3 6.13	0.67	14	25.56	70	3.94	15.57	3.68	3.01	1.88	0.79	14.26	2.9	1.9	1.01	0.54	1.31	0.78
98	Sta	57471	12 7	1549	4 10.08	0.75	22	38	220	5.61	26.72	5.43	5.21	2.66	1.95	24.53	4.09	3.46	1.62	1.03	2.2	1.34
99	Thi	11699	10 5	830	3 8.1	0.67	20	34.67	118	3.41	14.56	4.72	4.15	3.19	2.03	12.96	3.61	2.81	2.07	1.31	1.6	1.11
100	γ -Abu	4321	7 5	371	3 5.14	0.67	12	24	52	4.17	14.7	3.27	2.54	1.14	0.61	13.05	2.19	1.29	0.59	0.25	1.64	1.08
101	Aca	16240	9 7	816	4 7.11	0.75	16	28	114	6.13	20.11	4.27	3.24	1.64	0.98	18.47	3.19	1.99	1.09	0.6	1.64	1.08
102	Ach	11010	10 5	832	3 8.1	0.67	20	32	111	2.94	20.61	4.65	4.37	3.39	1.94	18.97	3.67	3.02	2.17	1.4	1.64	0.98
103	Afb	152012	17 9	3442	5 15.06	0.8	35	68	532	6.25	25.05	7.89	7.46	5.38	4.02	20.87	4.74	3.58	2.37	1.42	4.18	3.14
104	Aoq	78002	16 8	3083	4 14.06	1	34	60	434	5.56	23.54	7.68	6.86	5.54	4.11	20.51	4.87	3.56	2.4	1.56	3.03	2.81
105	Bpa	253741	20 12	6703	6 18.05	1	42	72	916	7.85	29.54	9.58	8.5	6.87	4.79	25.82	6.25	4.54	3.02	1.89	3.71	3.33
106	Mas	21566	10 5	774	3 8.1	0.67	18	40	121	4	17.34	4.46	4.21	2.93	1.21	14.44	2.68	1.99	1.21	0.37	2.87	1.79
107	Ceg	7751	8 5	438	3 6.13	0.67	14	25.78	70	3.94	14.57	3.68	3.01	1.88	0.79	13.06	2.76	1.8	0.96	0.48	1.51	0.92
108	Cha	65347	14 9	2388	5 12.07	0.8	28	44	337	5.78	29.55	6.63	5.75	4.39	2.67	27.61	5.14	4.06	3.03	2.06	1.94	1.49
109	Dty	141349	17 8	3584	4 15.06	1	34	60	494	6.25	28.05	7.95	7.05	6	4.12	24.34	5.03	3.75	2.71	1.64	3.72	2.92
110	Chg	18715	11 6	1161	3 9.09	1	22	34	156	4.1	23.27	5.22	4.5	3.44	2.2	21.62	4.2	3.26	2.44	1.69	1.64	1.02
111	Cpa	44867	13 8	1819	4 11.08	1	26	43.78	268	5.02	19.84	6.09	5.58	3.98	2.36	17.66	4.2	3.21	1.94	1.09	2.19	1.89
112	Deg	12243	9 4	640	2 7.11	1	16	28	86	3.24	20.49	4.06	3.52	3.01	1.07	18.85	3.09	2.17	1.69	0.55	1.64	0.98
113	Dmt	82012	15 8	2618	4 13.07	1	30	52	368	5.37	26.58	6.91	6.65	4.79	3.44	23.82	4.69	3.71	2.42	1.6	2.77	2.22
114	Dvg	14521	9 6	748	3 7.11	1	16	32	102	4.84	16.28	4.18	3.36	2.15	0.86	13.95	2.46	1.48	0.74	0.28	2.33	1.72
115	Gav	57386	12 7	1511	4 10.08	0.75	22	42	220	6.51	23.56	5.57	4.78	2.65	1.8	20.71	3.61	2.38	1.32	0.77	2.85	1.96
116	Hat	54848	15 8	2473	4 13.07	1	32	54	346	4.11	24.05	7.01	6.99	5.61	3.92	21.23	4.84	4.09	2.96	1.99	2.82	2.17
117	Hai	38837	14 7	1996	4 12.07	0.75	30	52	280	3.54	21.34	6.51	6.66	5.25	3.83	18.53	4.34	3.77	2.63	1.84	2.82	2.17
118	Hpp	57367	15 8	2590	4 13.07	1	32	54	362	4.89	23.84	7.18	6.52	5.15	4.14	21.18	4.93	3.76	2.81	2.04	2.66	2.25
119	Ing	39680	14 7	2146	4 12.07	0.75	30	48	286	4.24	23.13	6.7	6.05	5.12	3.93	20.81	4.76	3.69	2.87	2.12	2.32	1.94
120	Mhp	95302	15 10	3043	5 13.07	1	30	52	428	6.55	26.26	7.13	6.09	4.71	2.76	23.64	4.75	3.33	2.22	1.24	2.62	2.38
121	Oct	101858	13 10	2438	5 11.08	1	24	36	330	8.59	31.1	6.18	4.78	3.15	1.58	29.46	5.17	3.5	2.19	1.27	1.64	1.02
122	Oic	19378	12 6	1412	4 10.08	0.5	26	38	188	3.4	23.55	5.77	5.28	4.22	3.32	22.12	4.76	3.94	3.16	2.51	1.43	1.01
123	Pal	30269	12 7	1533	4 10.08	0.75	24	44	212	4.89	18.97	5.7	4.96	3.57	2.22	16.47	3.57	2.51	1.5	0.86	2.5	2.13
124	Tic	28633	13 7	1785	4 11.08	0.75	28	46	238	4.02	20.26	6.27	5.62	4.52	3.52	18.16	4.3	3.24	2.38	1.63	2.1	1.97
125	Thz	3936	8 4	426	2 6.13	1	16	26.67	62	2.52	12.98	3.8	3.29	2.34	1.62	12.07	3.49	2.64	1.95	1.48	0.91	0.31
126	Tle	12267	9 4	635	2 7.11	1	16	28	86	2.72	20.65	3.85	4.4	2.37	1	19.01	2.84	3.18	1.2	0.37	1.64	1.02
127	Dpg	95785	17 8	3632	4 15.06	1	36	60	474	5.76	25.3	8.15	7.19	6.21	4.73	22.2	5.29	3.85	2.8	1.82	3.1	2.87
128	Dbz	179146	19 10	5480	5 17.05	1	40	64	714	7.14	30.72	9.1	8.23	6.35	4.66	27.62	6.2	4.79	3.15	2.13	3.1	2.9
129	β -Phe	29814	12 7	1548	4 10.08	0.75	24	42	209	4.89	19.97	5.7	5	3.3	2.64	17.6	3.72	2.62	1.62	1.01	2.37	1.98
130	α -Abu	3821	7 4	337	2 5.14	1	12	24	46	3.06	14.86	3.18	2.63	1.78	0.47	13.22	2.17	1.35	0.76	0.17	1.64	1.02
131	Mpr	6744	9 4	629	2 7.11	1	18	30	84	2.72	17.85	4.22	3.82	2.78	2.16	16.42	3.18	2.54	1.8	1.17	1.43	1.04
132	Hva	12267	9 4	591	2 7.11	1	16	32	86	2.72	18.65	3.85	4.4	2.37	1	16.46	2.56	2.47	1.03	0.3	2.2	1.29
133	Dcp	84274	15 8	2397	4 13.07	1	30	55.78	378	5.37	21.59	6.91	6.61	5.05	3.13	18.4	4.48	3.47	2.25	1.22	3.19	2.43
134	Car	89562	13 8	1902	4 11.08	1	24	50	290	6.45	22.43	5.95	5.41	3.39	1.77	1.891	3.54	2.37	1.3	0.61	3.52	2.41

No.	Abbrev	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	Ala	1.36	0.94	0	5.15	2.64	2.49	1.33	3.51	1.63	1.13	0.39	5.59	1.89	2.59	1.76	22	2	2.2	1.29	1.44	0.68	14.5
2	Arg	2.47	1.5	0.68	9.56	5.54	4.9	2.83	6.71	3.6	2.43	1.32	11.18	5.78	6.57	5.39	48	3	6.07	1.5	1.66	0.41	28
3	Asn	2.24	1.24	0.71	7.44	4.04	3.85	1.98	4.7	2.3	1.62	0.74	8.22	3.29	3.87	3	36	3	3.99	1.39	1.56	0.5	22
4	Asp	2.31	1.27	0.74	7.44	4.04	3.85	1.98	4.57	2.24	1.54	0.71	8.22	3.29	3.87	3	36	3	3.99	1.39	1.56	0.5	22.5
5	Cys	1.14	0.88	0.24	5.86	3.18	2.63	1.78	4.56	2.41	1.49	0.91	6.94	3.01	2.73	2.98	26	2	2.89	1.36	1.51	0.59	16.5
6	Gln	2.24	1.32	0.51	8.15	4.54	4.19	2.34	5.41	2.8	1.95	1.02	9.22	4.1	4.71	3.78	40	3	4.68	1.43	1.6	0.46	24
7	Glu	2.32	1.35	0.53	8.15	4.54	4.19	2.34	5.28	2.74	1.88	0.99	9.22	4.1	4.71	3.78	40	3	4.68	1.43	1.6	0.46	24.5
8	Gly	1.2	0.64	0	4.28	2.27	1.8	0.82	2.64	1.19	0.6	0.17	4.59	1.87	4.22	1.72	16	1	1.79	1.3	1.43	0.75	12.5

9	His	2.4	2.02	1.33	8.27	5.2	4.61	3.32	5.82	3.16	2.21	1.3	8.06	3.38	2.23	2.48	50	6	6.76	1.69	1.85	0.39	25.5
10	Ile	1.24	1.05	0.53	7.44	4.09	3.49	2.59	5.79	3.08	2.25	1.54	8.59	3.58	2.64	3.42	36	3	3.99	1.39	1.56	0.5	20.5
11	Leu	1.28	0.94	0.41	7.44	4.04	3.85	1.98	5.79	3.02	2.57	1.04	8.59	3.58	4.21	3.42	36	3	3.99	1.39	1.56	0.5	20.2
12	Lys	1.49	1.11	0.42	7.98	4.68	3.72	2.4	5.92	3.37	2.23	1.29	9.55	5.34	5.02	5.1	38	2	4.97	1.5	1.64	0.45	23
13	Met	0.66	0.56	0	7.28	4.18	3.36	2.15	6.15	4.05	2.71	1.59	8.94	4.78	4.54	4.75	34	2	4.28	1.46	1.61	0.48	20.5
14	Phe	2.33	1.95	1.27	8.97	5.7	4.96	3.57	6.6	3.72	2.63	1.62	8.9	3.99	2.76	2.96	54	6	7.46	1.71	1.86	0.37	26.5
15	Pro	1.3	0.98	0.72	5.98	3.8	3.29	2.34	4.55	2.77	1.99	1.36	5.72	2.24	1.24	1.6	36	5	4.97	1.71	1.86	0.45	18.5
16	Ser	1.5	1.27	0.39	5.86	3.18	2.63	1.78	3.66	1.77	1.13	0.51	6.55	2.68	2.39	2.51	26	2	2.89	1.36	1.51	0.59	17.5
17	Thr	1.74	1.3	0.56	6.73	3.55	3.35	2.1	4.53	2.22	1.6	0.8	7.55	2.76	2.46	2.6	32	3	3.3	1.33	1.51	0.55	19.5
18	Trp	2.99	2.86	2.28	10.84	7.18	6.5	5.29	8.1	4.72	3.51	2.43	10.04	3.94	1.98	2.64	76	11	10.35	1.84	1.99	0.31	33
19	Tyr	2.77	2.27	1.39	9.84	6.09	5.58	3.98	6.97	3.86	2.82	1.71	9.85	4.13	2.98	3.13	60	7	7.86	1.66	1.83	0.36	29.5
20	Val	1.24	1.1	0.48	6.73	3.55	3.35	2.1	5.09	2.54	2.11	1	7.59	2.79	2.49	2.65	32	3	3.3	1.33	1.51	0.55	18.5
21	Acp	1.28	0.97	0.32	8.69	5.18	4.07	2.65	7.05	4.17	2.79	1.69	10.59	6.31	6.09	6.07	42	2	5.66	1.54	1.67	0.42	24.5
22	Aec	0.76	0.41	-0.04	7.98	4.68	3.72	2.4	6.43	4.1	2.96	2	9.9	5.67	5.35	5.61	38	2	4.97	1.5	1.64	0.45	23
23	Afa	2.27	2.02	1.4	8.27	5.22	4.5	3.44	5.9	3.25	2.23	1.42	7.91	3.27	1.89	2.36	50	6	6.76	1.69	1.85	0.39	24.5
24	Aib	1.49	1.18	0	6.08	2.94	3.52	1.73	4.43	1.97	2.03	0.55	6.59	1.6	2.42	1.5	30	4	2.49	1.25	1.43	0.63	16.5
25	Ail	1.28	0.94	0.41	7.44	4.04	3.85	1.98	5.79	3.02	2.57	1.04	8.59	3.58	4.21	3.42	36	3	3.99	1.39	1.56	0.5	20.5
26	Alg	1.54	1.14	0.51	6.57	3.68	3.01	1.88	4.5	2.27	1.47	0.74	7.33	3.35	2.97	3.07	30	2	3.58	1.41	1.57	0.53	18.5
27	Aba	1.27	1.02	0.3	5.86	3.18	2.63	1.78	4.22	2.17	1.35	0.76	6.59	2.72	2.42	2.55	26	2	2.89	1.36	1.51	0.59	16.5
28	Aph	2.69	2.23	1.37	9.84	6.09	5.58	3.98	7.1	3.92	2.89	1.75	9.85	4.13	2.98	3.13	60	7	7.86	1.66	1.83	0.36	29
29	β-Ala	1.25	0.51	0.45	4.99	2.77	2.18	0.87	3.35	1.69	0.93	0.36	5.59	2.81	4.59	2.62	20	1	2.49	1.39	1.51	0.63	14.5
30	Brp	1.89	1.76	1.14	9.84	6.09	5.58	3.98	8.49	4.61	3.69	2.21	10.37	4.5	3.29	3.59	60	7	7.86	1.66	1.83	0.36	28.25
31	Cha	1.28	0.94	0.34	8.97	5.7	4.96	3.57	7.33	4.68	3.68	2.63	9.68	4.58	3.25	3.69	54	6	7.46	1.71	1.86	0.37	26.5
32	Cit	2.52	1.52	0.7	9.56	5.54	4.9	2.83	6.62	3.55	2.38	1.31	11.18	5.78	6.57	5.39	48	3	6.07	1.5	1.66	0.41	28.5
33	Cla	1.22	0.97	0.27	5.86	3.18	2.63	1.78	4.35	2.26	1.41	0.82	6.88	2.96	2.67	2.91	26	2	2.89	1.36	1.51	0.59	16.83
34	Cle	1.35	1.22	0.54	6.57	4.15	4.02	3.14	5.26	3.17	2.67	1.92	6.71	2.08	1.11	1.55	44	7	5.26	1.61	1.79	0.44	20.5
35	Clp	2.37	2.04	1.28	9.84	6.09	5.58	3.98	7.66	4.2	3.21	1.94	10.18	4.37	3.17	3.42	60	7	7.86	1.66	1.83	0.36	28.83
36	Cya	2.1	0.74	0.43	8.36	4.33	4.99	2.07	5.71	3.87	3.19	1.49	9.37	2.97	4.85	2.78	44	5	4.28	1.35	1.53	0.48	25.5
37	Dab	1.49	1.09	0.49	6.57	3.68	3.01	1.88	4.5	2.37	1.52	0.79	7.55	3.54	3.16	3.34	30	2	3.58	1.41	1.57	0.53	19
38	Dap	1.45	1.21	0.37	5.86	3.18	2.63	1.78	3.79	1.87	1.18	0.57	6.55	2.68	2.39	2.51	26	2	2.89	1.36	1.51	0.59	17
39	Dhp	1.64	1.29	0.99	5.98	3.8	3.29	2.34	4.29	2.43	1.65	1.05	5.47	2.07	1.11	1.41	36	5	4.97	1.71	1.86	0.45	18.5
40	Dha	3.12	2.74	1.64	10.72	6.5	6.09	4.55	7.34	4	2.97	1.81	10.81	4.3	2.85	3.32	66	8	8.27	1.62	1.81	0.35	32.5
41	Fph	2.81	2.29	1.4	9.84	6.09	5.58	3.98	6.9	3.82	2.78	1.69	9.83	4.11	2.96	3.11	60	7	7.86	1.66	1.83	0.36	30
42	Gaa	2.62	2.67	1.28	10.59	5.91	5.21	4.25	6.74	3.55	2.6	1.58	12.43	5.22	3.38	4.99	56	5	6.19	1.42	1.61	0.4	32.5
43	Hag	2.47	1.5	0.68	9.56	5.54	4.9	2.83	6.71	3.6	2.43	1.32	11.18	5.78	6.57	5.39	48	3	6.07	1.5	1.66	0.41	28
44	Hly	1.9	1.55	0.47	8.85	5.07	4.31	2.95	6.23	3.44	2.41	1.4	10.51	5.2	4.86	4.97	44	3	5.38	1.47	1.63	0.43	26
45	Hnv	1.65	1.48	0.61	7.44	4.09	3.49	2.59	5.24	2.76	1.84	1.12	8.55	3.55	2.61	3.37	36	3	3.99	1.39	1.56	0.5	21.5
46	Hog	2.24	1.32	0.56	8.85	5.04	4.55	2.58	6.12	3.3	2.3	1.26	10.22	4.95	5.74	4.6	44	3	5.38	1.47	1.63	0.43	26
47	Hop	2.33	1.99	1.22	9.65	6.2	5.3	3.89	7.31	4.22	2.97	1.9	9.89	4.74	3.36	3.61	58	6	8.15	1.73	1.87	0.35	28.5
48	Hos	1.55	1.12	0.53	6.57	3.68	3.01	1.88	4.37	2.27	1.46	0.76	7.55	3.54	3.16	3.34	30	2	3.58	1.41	1.57	0.53	19.5
49	Hpr	1.73	1.27	0.95	6.85	4.2	3.92	2.68	4.87	2.84	2.19	1.41	6.67	2.43	1.48	1.8	42	6	5.38	1.64	1.82	0.43	21.5
50	Iph	1.58	1.58	1.05	9.84	6.09	5.58	3.98	9.06	4.9	4.02	2.41	10.62	4.69	3.44	3.83	60	7	7.86	1.66	1.83	0.36	27.9
51	Ise	1.7	1.08	0	5.15	2.64	2.49	1.33	2.96	1.31	0.78	0.25	5.55	1.86	2.55	1.72	22	2	2.2	1.29	1.44	0.68	15.5
52	Mle	1.42	1.18	0.35	8.36	4.36	4.74	2.5	6.72	3.38	3.32	1.32	9.59	3.12	3.33	2.99	44	5	4.28	1.35	1.53	0.48	22.5
53	Msm	-0.6	0.13	-0.28	8.15	4.54	4.19	2.34	7.15	5.27	4.8	2.21	9.94	4.71	5.39	4.69	40	3	4.68	1.43	1.6	0.46	22.5
54	1Nala	3.07	2.9	2.32	11.54	7.68	6.86	5.54	8.76	5.13	3.79	2.64	10.8	4.43	2.27	2.99	80	11	11.04	1.85	1.99	0.3	34.5
55	2Nala	3.11	2.9	2.28	11.54	7.66	6.94	5.49	8.76	5.13	3.83	2.59	10.8	4.43	2.47	2.99	80	11	11.04	1.85	1.99	0.3	34.5
56	Nle	1.28	0.97	0.31	7.28	4.18	3.36	2.15	5.63	3.17	2.09	1.19	8.59	4.47	4.21	4.26	34	2	4.28	1.46	1.61	0.48	20.5
57	Nma	1.28	0.98	0.35	5.86	3.18	2.63	1.78	4.43	2.08	1.35	0.8	6.59	2.71	2.42	2.55	26	2	2.89	1.36	1.51	0.59	16.5
58	Nva	1.28	0.97	0.35	6.57	3.68	3.01	1.88	4.92	2.67	1.73	0.92	7.59	3.57	3.2	3.39	30	2	3.58	1.41	1.57	0.53	18.5
59	Obs	2.71	2.41	1.59	10.39	6.7	5.66	4.13	7.72	4.3	2.95	1.73	10.85	5.49	4.03	4.25	62	6	8.84	1.75	1.88	0.34	31.5
60	Obt	3.86	3.57	2.7	14.37	9.65	8.43	6.52	11.03	6.39	4.57	2.94	14.4	6.96	4.54	5.01	96	11	13.81	1.88	2	0.27	43.5
61	Oet	2.91	2.5	1.73	11.26	7.13	6.13	4.51	8.64	4.83	3.23	2.01	11.84	5.6	3.79	4.42	68	7	9.25	1.7	1.85	0.33	33.5
62	Oms	1.61	1.22	0.54	6.57	3.68	3.01	1.88	4.63	2.15	1.4	0.66	7.55	3.54	3.16	3.34	30	2	3.58	1.41	1.57	0.53	19.5

63	Omt	1.68	1.43	0.68	7.44	4.09	3.49	2.59	5.5	2.6	1.8	1.6	8.55	3.55	2.61	3.37	36	3	3.99	1.39	1.56	0.5	21.5
64	Omy	2.76	2.45	1.57	10.55	6.63	5.75	4.39	7.93	4.25	3	1.93	10.85	4.85	3.21	3.76	64	7	8.55	1.68	1.84	0.34	31.5
65	Orn	1.49	1.11	0.4	7.28	4.18	3.36	2.15	5.21	2.87	1.88	1.04	8.55	4.43	4.18	4.21	34	2	4.28	1.46	1.61	0.48	21
66	Pen	0.78	1.06	0.59	7.65	3.85	4.4	2.37	6.35	3.01	3.62	1.31	8.94	2.68	2.9	2.66	40	5	3.58	1.3	1.49	0.53	20.5
67	Pga	1.93	1.38	1.06	6.85	4.2	3.92	2.68	4.76	2.72	2	1.29	6.38	2.24	1.34	1.59	42	6	5.38	1.64	1.82	0.43	21.5
68	Pip	1.3	0.98	0.74	6.69	4.3	3.64	2.59	5.26	3.27	2.34	1.61	6.71	2.94	1.81	2.19	40	5	5.66	1.74	1.88	0.42	20.5
69	Sar	1.26	0.54	0.43	4.99	2.77	2.18	0.87	3.56	1.63	0.92	0.33	5.59	2.81	4.59	2.62	20	1	2.49	1.39	1.51	0.63	14.5
70	Tfa	3.04	1.74	0.86	7.65	3.85	4.4	2.37	4.14	1.91	1.36	0.62	8.38	2.32	2.49	2.16	40	5	3.58	1.3	1.49	0.53	25
71	Thp	3.5	3.07	1.93	11.59	6.91	6.62	5.01	7.71	4.14	3.12	1.93	11.76	4.47	2.84	3.51	72	9	8.67	1.58	1.78	0.34	35.5
72	Vig	1.53	1.24	0.37	5.86	3.18	2.63	1.78	3.79	1.79	1.1	0.54	6.33	2.5	2.2	2.26	26	2	2.89	1.36	1.51	0.59	16.5
73	Aas	2.44	1.33	0.98	9.78	5.39	5.34	2.94	7.25	5.1	4.58	2.9	11.37	4.41	4.75	4.18	52	5	5.66	1.43	1.6	0.42	29
74	Ahd	2.42	1.93	1.01	10.1	6.18	4.78	3.15	7.31	3.93	2.36	1.22	12.47	8.09	7.89	7.76	50	2	7.05	1.59	1.72	0.38	31.5
75	Aho	1.97	1.55	0.7	7.98	4.68	3.72	2.4	5.49	2.85	1.74	0.85	9.51	5.31	4.98	5.05	38	2	4.97	1.5	1.64	0.45	24.5
76	Ahs	2.5	1.37	1.04	9.78	5.39	5.34	2.94	7.12	5.01	4.51	2.82	11.37	4.41	4.75	4.18	52	5	5.66	1.43	1.6	0.42	29.5
77	Ahp	0.83	0.49	0.01	7.98	4.68	3.72	2.4	6.3	4.01	2.89	1.92	9.9	5.67	5.35	5.61	38	2	4.97	1.5	1.64	0.45	23.5
78	Ahd	2.87	2.31	1.32	12.23	7.68	5.84	3.9	9.13	5.01	2.97	1.59	15.43	10.93	10.71	10.54	62	2	9.13	1.66	1.77	0.33	38.5
79	Dad	2.36	1.9	1	10.1	6.18	4.78	3.15	7.44	4.02	2.42	1.25	12.47	8.09	7.89	7.76	50	2	7.05	1.59	1.72	0.38	31
80	Dat	2.81	2.28	1.31	12.23	7.68	5.84	3.9	9.26	5.1	3.03	1.62	15.43	10.93	10.71	10.54	62	2	9.13	1.66	1.77	0.33	38
81	Dfn	2.77	1.28	0.46	9.07	4.83	5.32	2.49	6.18	3.19	2.55	1.21	10.45	3.73	5.96	3.54	48	5	4.97	1.39	1.57	0.45	27.5
82	Dfv	2.77	1.18	0.73	8.36	4.33	4.99	2.07	5.47	2.69	2.22	0.9	9.45	3.03	4.92	2.86	44	5	4.28	1.35	1.53	0.48	25.5
83	Dtc	2.25	1.25	1.04	8.48	5.01	5.26	3.19	6.6	5	4.69	3.19	8.51	2.8	2.21	2.17	54	8	6.36	1.59	1.78	0.4	26.5
84	Hfn	5.26	3.93	1.61	13.36	6.83	8.18	5.54	7.36	3.63	2.93	1.61	15.1	4.17	3.03	3.93	80	11	7.05	1.34	1.55	0.38	45
85	Pfn	4.21	2.89	0.69	11.57	6.08	6.89	4.39	6.81	3.51	2.69	1.5	13.24	4.18	3.86	3.95	66	8	6.36	1.38	1.58	0.4	38
86	Pfv	4.21	2.79	0.87	10.86	5.58	6.56	4	6.11	3.01	2.35	1.21	12.24	3.54	3.26	3.34	62	8	5.66	1.35	1.55	0.42	36
87	Tca	0.58	0.32	0.16	6.69	4.3	3.64	2.59	5.78	4	3.07	2.28	7.05	3.2	2.01	2.5	40	5	5.66	1.74	1.88	0.42	20.5
88	Sec	0.73	0.44	0.06	5.86	3.18	2.63	1.78	5.54	3.1	1.89	1.34	6.59	2.71	2.42	2.55	26	2	2.89	1.36	1.51	0.59	16.5
89	Pyl	3.07	2.83	2.04	13.54	8.52	7.41	5.62	10.48	6.01	4.34	2.78	14.69	7.15	4.81	5.84	82	8	11.04	1.69	1.85	0.3	40.5
90	Ath	3.83	3.8	3.31	14.11	9.66	8.77	7.42	10.91	6.54	4.95	3.62	12.79	5.04	2.32	3.22	106	16	14.62	1.94	2.08	0.26	42.5
91	Bal	2.19	1.97	1.41	10.84	7.18	6.5	5.29	8.83	5.5	4.32	3.33	10.42	4.18	2.13	2.9	76	11	10.35	1.84	1.99	0.31	32.5
92	Bip	3.42	3.22	2.49	12.96	8.66	7.64	6.05	9.91	5.79	4.21	2.83	12.49	5.58	3.26	3.87	88	11	12.43	1.86	1.99	0.28	38.5
93	Dip	3.31	3.09	2.66	12.96	8.7	7.48	6.05	9.86	5.78	4.16	2.96	12.49	5.58	2.84	3.87	88	11	12.43	1.86	1.99	0.28	38.5
94	Tbt	3.34	3.81	3.18	20.95	12.03	14.16	8.99	18.37	9.72	10.82	5.18	21.81	6.84	4.51	5.52	116	17	13.64	1.64	1.86	0.27	47
95	Tpc	6.02	5.6	5.42	25	15.82	16.36	13.01	20.54	12.27	10.63	7.71	24.34	8.46	4.37	6.05	142	24	18.38	1.92	2.09	0.23	54.5
96	Asu	2.32	1.36	0.59	10.27	6.04	5.25	3.08	7.4	4.24	2.94	1.72	12.22	6.71	7.65	6.3	52	3	6.76	1.53	1.68	0.39	30.5
97	Hcy	1.11	0.87	0.25	6.57	3.68	3.01	1.88	5.27	2.91	1.9	1.02	7.94	3.88	3.52	3.85	30	2	3.58	1.41	1.57	0.53	18.5
98	Sta	1.76	1.04	0.92	9.72	5.43	5.21	2.66	7.52	4.09	3.46	1.62	11.55	5.24	5.84	5.04	50	4	5.78	1.44	1.62	0.42	27.5
99	Thi	1.34	1.13	0.72	7.56	4.72	4.15	3.19	5.97	3.61	2.81	2.07	7.53	3.01	1.67	2.26	46	6	6.07	1.67	1.84	0.41	22.5
100	γ -Abu	1.25	0.54	0.36	5.7	3.27	2.54	1.14	4.05	2.19	1.29	0.59	6.59	3.77	5.87	3.55	24	1	3.18	1.45	1.58	0.56	16.5
101	Aca	1.25	0.54	0.38	7.11	4.27	3.24	1.64	5.47	3.19	1.99	1.09	8.59	5.72	7.77	5.46	32	1	4.56	1.54	1.66	0.47	20.5
102	Ach	1.35	1.22	0.54	7.61	4.65	4.37	3.39	5.97	3.67	3.02	2.17	7.69	2.68	1.56	2.06	48	7	5.95	1.64	1.81	0.41	22.5
103	Afb	3.88	3.01	2.59	13.05	7.89	7.46	5.38	8.87	4.78	3.61	2.4	13.4	5.11	3.16	4.02	82	10	9.94	1.61	1.79	0.32	40
104	Aoq	3.3	3.14	2.55	11.54	7.68	6.86	5.54	8.51	4.87	3.56	2.4	10.97	4.54	2.35	3.12	80	11	11.04	1.85	1.99	0.3	36
105	Bpa	3.95	3.85	2.9	14.54	9.58	8.5	6.87	10.82	6.25	4.54	3.02	14.12	6.23	3.61	4.4	98	12	13.52	18.82	1.97	0.27	43.5
106	Mas	2.22	1.72	0.85	8.31	4.46	4.21	2.93	5.44	2.68	1.99	1.21	9.22	3.4	2.58	3.14	42	4	4.39	1.36	1.55	0.48	24.5
107	Ceg	1.21	0.93	0.31	6.56	3.68	3.01	1.88	5.06	2.76	1.8	0.96	7.88	3.83	3.46	3.77	30	2	3.58	1.41	1.57	0.53	18.83
108	Cha	1.69	1.36	0.61	10.55	6.63	5.75	4.39	8.61	5.14	4.06	3.03	11.62	5.44	3.68	4.51	64	7	8.55	1.68	1.84	0.34	31.5
109	Dty	3.3	3.29	2.49	13.05	7.95	7.05	6	9.34	5.03	3.75	2.71	13.32	5.05	2.5	3.96	82	10	9.94	1.61	1.79	0.32	37.5
110	Chg	1.24	1	0.51	8.27	5.22	4.5	3.44	6.62	4.2	3.26	2.44	8.68	3.83	2.3	3.02	50	6	6.76	1.69	1.85	0.39	24.5
111	Cpa	2.37	2.04	1.28	9.84	6.09	5.58	3.98	7.66	4.2	3.21	1.94	10.18	4.37	3.17	3.42	60	7	7.86	1.66	1.83	0.36	28.83
112	Deg	1.35	1.32	0.51	7.49	4.06	3.52	3.01	5.85	3.09	2.17	1.69	8.59	2.94	1.81	2.81	38	4	3.87	1.37	1.54	0.51	20.5
113	Dmt	2.95	2.37	1.84	11.59	6.91	6.65	4.79	8.82	4.69	3.71	2.42	11.84	4.53	2.88	3.57	72	9	8.67	1.58	1.78	0.34	33.5
114	Dvg	1.88	1.41	0.58	7.28	4.18	3.36	2.15	4.95	2.46	1.48	0.74	8.07	4	3.74	3.59	34	2	4.28	1.46	1.61	0.48	20.5
115	Gav	2.4	1.32	1.03	9.56	5.57	4.78	2.65	6.71	3.61	2.38	1.32	11.18	5.78	5.52	5.39	48	3	6.07	1.5	1.66	0.41	28
116	Hat	2.9	2.65	1.94	11.05	7.01	6.99	5.61	8.23	4.84	4.09	2.96	10.29	3.41	1.8	2.34	80	13	9.94	1.75	1.94	0.32	33.5

117	Hai	2.89	2.63	1.99	10.34	6.51	6.66	5.25	7.53	4.34	3.77	2.63	9.33	2.88	1.46	1.92	76	13	9.25	1.73	1.94	0.33	31.5
118	Hpp	2.77	2.34	2.1	10.84	7.19	6.52	5.15	8.18	4.93	3.76	2.81	10.29	4.1	2.08	2.81	76	11	10.35	1.84	1.99	0.31	33.5
119	Ing	2.37	2.25	1.81	10.13	6.7	6.05	5.12	7.81	4.76	3.69	2.87	9.24	3.44	1.55	2.27	72	11	9.65	1.83	1.99	0.32	30.5
120	Mhp	2.76	2.49	1.52	11.26	7.13	6.09	4.71	8.64	4.75	3.33	2.22	11.84	5.6	3.79	4.42	68	7	9.25	1.7	1.85	0.33	33.5
121	Oct	1.28	0.97	0.31	10.1	6.18	4.78	3.15	8.46	5.17	3.5	2.19	12.59	8.2	8.01	7.94	50	2	7.05	1.59	1.72	0.38	28.5
122	Oic	1.34	1.06	0.81	8.55	5.77	5.28	4.22	7.12	4.76	3.94	3.16	8.2	3.15	1.53	2.15	62	10	8.55	1.9	2.04	0.34	26.5
123	Pal	2.45	2.07	1.36	8.97	5.7	4.96	3.57	6.47	3.57	2.51	1.5	8.83	3.94	2.72	2.9	54	6	7.46	1.71	1.86	0.37	27
124	Tic	2.38	2.14	1.89	9.26	6.27	5.62	4.52	7.16	4.3	3.24	2.38	8.41	3.28	1.63	2.12	66	10	9.25	1.9	2.04	0.33	28.5
125	Thz	0.65	0.39	0.14	5.98	3.8	3.29	2.34	5.07	3.5	2.64	1.95	6.07	2.48	1.42	1.88	36	5	4.97	1.71	1.86	0.45	18.5
126	Tle	1.22	1.17	0.63	7.65	3.85	4.4	2.37	6.01	2.84	3.18	1.2	8.59	2.45	2.64	2.34	40	5	3.58	1.3	1.49	0.53	20.5
127	Dpg	3.35	3.42	2.91	12.3	8.15	7.19	6.21	9.21	5.29	3.85	2.8	11.51	4.49	1.99	3.04	86	12	11.61	1.84	1.98	0.29	36.5
128	Dbz	3.44	3.2	2.52	13.72	9.1	8.23	6.36	10.62	6.2	4.79	3.15	13.46	5.78	3.51	4.1	94	12	13	1.85	1.98	0.28	40.5
129	β-Phe	2.39	1.68	1.63	8.97	5.7	5	3.3	6.6	3.72	2.62	1.62	8.9	3.99	2.76	2.96	54	6	7.46	1.71	1.86	0.37	26.5
130	α-Abu	1.28	1.02	0.3	5.86	3.18	2.63	1.78	4.22	2.17	1.35	0.76	6.59	2.71	2.42	2.55	26	2	2.89	1.36	1.51	0.59	16.5
131	Mpr	1.28	0.99	1	6.85	4.22	3.82	2.78	5.42	3.18	2.54	1.8	6.71	2.45	1.31	1.83	42	6	5.38	1.64	1.82	0.43	20.5
132	Hva	1.93	1.34	0.7	7.65	3.85	4.4	2.37	5.46	2.56	2.47	1.03	8.55	2.43	2.61	2.31	40	5	3.58	1.3	1.49	0.53	21.5
133	Dcp	3.13	2.8	1.91	11.59	6.91	6.61	5.05	8.4	4.48	3.47	2.25	12.09	4.69	3	3.78	72	9	8.67	1.58	1.78	0.34	34.83
134	Car	3.03	2.09	1.15	10.43	5.95	5.41	3.39	6.91	3.54	2.37	1.3	11.85	5.49	5.2	5	54	4	6.47	1.47	1.65	0.39	31

No.	Abbrev	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
1	Ala	13.48	9.67	0.38	108	61	5.4	8.94	6.94	9.33	-4.59	1.52	47	29	2.99	2.63	2.57	10.55	16.08	15.51	20	0.92
2	Arg	44.38	41.17	0.38	904	735	12.8	79.84	16.6	26.92	-18.67	4.05	739	247	3.2	2.67	3.22	19.29	62.85	43.02	38	2.12
3	Asn	27.36	21.33	0.35	348	248	9.03	32.13	11.92	17.97	-9.82	2.43	206	96	3.38	2.88	2.99	19.12	59.05	28.53	33	1.44
4	Asp	27.36	21.33	0.35	348	248	9.03	32.13	11.92	17.97	-9.82	2.43	206	96	3.38	2.96	2.85	20.97	70.17	28.53	33	1.44
5	Cys	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.67	2.76	12.26	25.02	19.65	21	1.38
6	Gln	32.84	27.2	0.36	494	373	10.32	44.1	13.48	20.87	-12.45	2.99	330	136	3.3	2.78	3.1	20.33	62.31	33.22	34	1.76
7	Glu	32.84	27.2	0.36	494	373	10.32	44.1	13.48	20.87	-12.45	2.99	330	136	3.3	2.83	2.99	22.29	71.57	33.22	34	1.76
8	Gly	9.76	7.2	0.44	68	36	4.3	5.2	5.22	6.67	-3.49	1.54	28	18	2.54	2.63	2.24	8.82	8.14	11.61	13	0.97
9	His	37.23	30	0.32	678	590	11.22	25.55	16.54	25.6	0	2.73	397	233	2.18	2.67	3.26	15.98	29.69	38.05	13	1.77
10	Ile	26.96	20.44	0.33	332	232	8.89	31.18	12.07	18.23	-9.41	2.1	188	92	3.58	2.58	3.18	13.61	40.2	28.53	33	1.44
11	Leu	27.36	21.33	0.35	348	248	9.03	32.13	11.92	17.97	-9.82	2.43	206	96	3.38	2.58	3.11	13.27	37.65	26.53	33	1.44
12	Lys	33.31	28.6	0.38	520	401	10.53	32.56	13.23	20.45	-13.09	3.24	368	143	3.13	2.6	3.14	15.76	45.18	33.22	28	1.92
13	Met	27.88	22.67	0.38	370	272	9.27	23.94	11.67	17.55	-10.44	2.73	235	102	3.16	2.58	3.07	13.07	35.24	28.53	25	1.84
14	Phe	42.57	35.33	0.32	864	766	12.31	30.77	18.32	29.17	0	2.82	542	320	2.2	2.63	3.38	14.49	32.35	39.02	13	1.7
15	Pro	21.8	15.5	0.33	270	217	7.68	7.68	11.42	16	0	1.63	114	106	2.24	2.39	3.14	11.69	14.71	24	6	1.25
16	Ser	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.85	2.65	14.62	33.94	19.65	21	1.38
17	Thr	22.14	16.25	0.34	236	155	7.71	22.74	10.39	15.17	-7.54	1.87	122	65	3.46	2.85	2.86	16.64	50.32	24	30	1.3
18	Trp	58.16	49.2	0.28	1576	1507	15.03	43.23	25.64	42.95	16.7	2.83	1018	772	1.88	2.64	3.63	16.83	34.65	58.6	13	1.51
19	Tyr	48.11	41.23	0.31	1077	955	13.36	50.41	20.25	32.99	0	2.9	735	388	2.22	2.69	3.4	20.39	42.63	44.11	20	1.66
20	Val	22.14	16.25	0.34	236	155	7.71	22.74	10.39	15.17	-7.54	1.87	122	65	3.46	2.6	3.02	12.81	32.53	22	30	1.3
21	Acp	38.95	35.27	0.39	708	566	11.76	43.37	14.79	23.44	-16.06	3.74	552	194	3.11	2.55	3.24	14.15	49.78	38.05	29	2.19
22	Aec	33.31	28.6	0.38	520	401	10.53	32.56	13.23	20.45	-13.09	3.24	368	143	3.13	2.62	3.12	15.63	44.55	33.22	28	1.92
23	Afa	36.61	28.36	0.3	641	553	11.03	22.82	16.8	26.12	0	2.15	353	252	2.32	2.67	3.38	14.45	28.65	34.05	12	1.46
24	Aib	17.25	12	0.33	156	90	6.37	15.1	8.92	12.5	-5.62	1.44	69	42	3.54	2.65	2.9	12.2	27.91	17.65	29	0.86
25	Ail	27.36	21.33	0.35	348	248	9.03	32.13	11.92	17.97	-9.82	2.43	206	96	3.38	2.58	3.11	13.27	37.65	26.53	33	1.44
26	Alg	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.65	2.92	12.91	30.28	24	24	1.5
27	Aba	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.6	2.81	11.76	22.89	19.65	21	1.38
28	Aph	48.11	41.23	0.31	1077	955	13.34	50.41	20.25	32.99	0	2.9	735	388	2.22	2.66	3.42	17.59	39.05	44.11	20	1.66
29	β-Ala	14.07	10.67	0.43	118	73	5.7	7.21	6.71	9	-5.07	2.05	58	32	2.63	2.5	2.52	10.4	12.25	15.51	14	1.46
30	Brp	48.11	41.23	0.31	1077	955	13.34	50.41	20.25	32.99	0	2.9	735	388	2.22	2.64	3.44	15.84	37.5	44.11	20	1.66
31	Cha	42.57	35.33	0.32	864	766	12.31	30.77	18.32	29.17	0	2.82	542	320	2.2	2.53	3.45	14.97	34.66	39.02	13	1.7
32	Cit	44.38	41.17	0.38	904	735	12.8	79.84	16.6	26.92	-18.67	4.05	739	247	3.2	2.7	3.2	21.29	67.73	43.02	38	2.12
33	Cla	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.73	2.72	13.02	27.32	19.65	21	1.38

34	Cle	25.86	18	0.29	344	272	8.56	12.61	13.61	19.83	0	1.46	148	133	2.51	2.51	3.4	14.9	27.25	24.53	11	1.23
35	Clp	48.11	41.23	0.31	1077	955	13.34	50.41	20.25	32.99	0	2.9	735	388	2.22	2.66	3.42	17.75	39.22	44.11	20	1.66
36	Cya	31.93	24.8	0.32	450	325	9.97	60.86	13.99	21.7	-11.35	2.5	272	124	3.68	5.19	2.96	27.34	200.11	31.22	44	1.37
37	Dab	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.66	2.91	14.23	32.43	24	24	1.5
38	Dap	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.72	2.73	12.92	27.01	19.65	21	1.38
39	Dhp	21.8	15.5	0.33	270	217	7.68	7.68	11.42	16	0	1.63	114	106	2.24	2.47	3.09	11.27	13.7	24	6	1.25
40	Dha	53.3	45.86	0.3	1276	1128	14.21	90.8	22.34	37.31	0	2.9	893	469	2.33	2.77	3.42	26.69	61.51	53.3	29	1.62
41	Fph	48.11	41.23	0.31	1077	955	13.34	50.41	20.25	32.99	0	2.9	735	388	2.22	2.72	4.44	23.2	47.3	44.11	20	1.66
42	Gaa	47.39	39.08	0.29	916	716	13.06	227.26	19.48	32.79	-17.66	2.83	640	254	4.15	3.17	3.22	35.4	311.16	48.11	61	1.73
43	Hag	44.38	41.17	0.38	904	735	12.8	79.84	16.6	26.92	-18.67	4.05	739	247	3.2	2.67	3.22	19.29	62.85	43.02	38	2.12
44	Hly	38.15	32.73	0.35	654	510	11.45	58.31	15.15	24.1	-14.9	3.27	469	180	3.37	2.71	3.16	21.34	74.39	38.05	37	1.86
45	Hnv	26.96	20.44	0.33	332	232	8.89	31.18	12.07	18.23	-9.41	2.1	188	92	3.58	2.83	3.02	18.05	63	28.53	33	1.44
46	Hog	38.52	33.82	0.37	678	534	11.58	59.67	15.04	23.86	-15.4	3.53	504	186	3.24	2.71	3.19	21.21	67.89	38.05	37	1.86
47	Hop	48.7	43.08	0.33	1135	1027	13.54	39.86	19.88	32.32	0	3.45	799	400	2.1	2.6	3.4	14.59	36.02	44.11	14	1.91
48	Hos	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.74	2.85	16.08	37.87	24	24	1.5
49	Hpr	26.49	19.33	0.32	366	296	8.76	12.41	13.28	19.32	0	1.71	173	142	2.3	2.55	3.17	17.02	26.48	28.53	11	1.23
50	Iph	48.11	41.23	0.31	1077	955	13.36	50.41	20.25	32.99	0	2.9	735	388	2.22	2.63	3.44	14.95	37.03	44.11	20	1.67
51	Ise	13.48	9.67	0.38	108	61	5.4	8.94	6.94	9.33	-4.59	1.52	47	29	2.99	3.07	2.32	12.28	25.89	15.51	20	0.92
52	Mle	31.57	24	0.31	434	309	9.87	58.86	14.13	21.97	-10.98	2.22	254	120	3.87	2.6	3.43	14.92	56.2	31.22	44	1.37
53	Msm	32.84	27.2	0.36	494	373	10.32	44.1	13.48	20.87	-12.45	2.99	330	136	3.3	2.56	3.19	13.98	44.93	31.22	34	1.76
54	1Nala	63.61	54.25	0.26	1840	1767	15.92	46.51	27.63	47.43	18.17	2.78	1213	970	1.92	2.63	3.72	16.11	37.69	62	13	1.45
55	2Nala	64.46	57.25	0.28	1948	1887	16.16	52.77	27.36	46.73	18.63	2.96	1377	958	1.81	2.63	3.64	15.79	37.11	64	13	1.45
56	Nle	27.88	22.67	0.38	370	272	9.27	23.84	11.67	17.56	-10.44	2.73	235	102	3.16	2.57	3.08	13.15	35.56	28.53	25	1.84
57	Nma	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.48	2.87	10.1	17.6	19.65	21	1.38
58	Nva	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.58	2.96	12.56	29.05	24	24	1.5
59	Obs	54.97	51.57	0.34	1458	1340	14.74	50.05	21.45	35.56	0	4.05	1137	493	2.02	2.71	3.34	17.41	42.91	46.55	15	2.11
60	Obt	91.44	98.8	0.31	4128	4097	20.59	102.76	33.8	61.33	34.26	4.56	4052	1396	1.59	2.66	3.68	20.16	52.02	78.44	13	1.28
61	Oet	60.03	56	0.32	1659	1493	15.54	80.19	23.67	40.23	0	3.25	1342	564	2.19	2.65	3.52	18.67	46.66	54.6	26	2.56
62	Oms	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.7	2.88	13.91	31.17	24	24	1.5
63	Omt	26.96	20.44	0.33	332	232	8.89	31.18	12.07	18.23	-9.41	2.1	188	92	3.58	2.7	3.1	15.37	44.05	28.53	33	1.44
64	Omy	53.94	48.14	0.32	1340	1196	14.43	64.31	21.99	36.64	0	3.04	997	469	2.22	2.65	3.48	17.87	41.82	49.3	23	1.73
65	Orn	27.88	22.67	0.38	370	272	9.27	23.94	11.67	17.55	-10.44	2.73	235	102	3.16	2.62	3.04	15.12	38.52	28.53	25	1.84
66	Pen	26.38	19.11	0.31	312	208	8.64	41.12	12.49	18.83	-8.8	1.87	164	86	3.88	2.69	3.17	14.99	50.61	26.53	41	1.22
67	Pga	26.49	19.33	0.32	366	296	8.76	12.41	13.28	19.32	0	1.71	173	142	2.3	2.61	3.38	17.66	25.8	28.53	11	1.23
68	Pip	26.62	19.56	0.32	375	315	8.86	9.17	13.14	19.15	0	1.66	176	160	2.23	2.38	3.29	12.37	17.9	28.53	6	1.19
69	Sar	14.07	10.67	0.43	118	73	5.7	7.21	6.71	9	-5.07	2.05	58	32	2.63	2.49	2.54	8.78	9.68	15.51	14	1.46
70	Tfa	26.38	19.11	0.31	312	208	8.64	41.12	12.49	18.83	-8.8	1.87	164	86	3.88	6.09	3.19	27.85	454.16	23.77	41	1.22
71	Thp	58.36	49.87	0.28	1473	1297	15.01	175.51	24.54	41.99	0	2.79	1035	560	2.48	2.87	3.45	33.41	88.52	58.6	40	1.57
72	Vig	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.71	2.73	11.59	23.07	19.65	21	1.38
73	Aas	42.46	35.17	0.31	764	591	12.2	111.77	17.46	26.59	-15.95	2.86	529	211	3.87	4.32	3.92	30.62	234.39	41.02	48	1.79
74	Ahd	50.78	50.77	0.4	1214	1020	14.13	72.19	17.93	29.64	-22.94	4.69	1118	330	3.07	2.75	3.2	23.55	79.34	48.11	33	2.47
75	Aho	33.31	28.6	0.38	520	401	10.53	32.56	13.23	20.45	-13.09	3.24	368	143	3.13	2.77	3.03	19.29	53.28	33.22	28	1.92
76	Ahs	42.46	35.17	0.31	764	591	12.2	111.78	17.46	28.59	-15.95	2.86	529	211	3.87	4.45	3.84	32.67	264.76	41.02	48	1.79
77	Ahp	33.31	28.6	0.38	520	401	10.53	32.56	13.23	20.45	-13.09	3.24	368	143	3.13	2.66	3.09	17.67	48.98	33.22	28	1.92
78	Ahd	69.63	79.25	0.4	2358	2071	17.5	135.4	22.67	39.47	-35.54	6.07	2657	634	3.04	2.75	3.29	27.68	110.09	64	40	2.75
79	Dad	50.78	50.77	0.4	1214	1020	14.13	72.19	17.93	29.64	-22.94	4.69	1118	330	3.07	2.73	3.21	21.68	74.37	48.11	33	2.47
80	Dat	69.63	79.25	0.4	2358	2071	17.5	135.4	22.67	39.47	-35.54	6.07	2657	634	3.04	2.74	3.3	25.77	104.51	64	40	2.75
81	Dfn	37.68	31.27	0.34	626	478	11.26	85.75	15.54	24.68	-14.24	3.11	428	172	3.54	4.08	4.1	28.88	150.41	36.05	45	1.69
82	Dfv	31.93	24.8	0.32	450	325	9.97	60.86	13.99	21.7	-11.35	2.5	272	124	3.68	4.25	3.99	27.62	158.59	31.22	44	1.37
83	Dtc	36.04	26.91	0.28	605	505	10.77	26.31	17.25	26.78	0	1.69	317	256	2.46	4.04	3.86	27.16	90.08	36.05	18	1.16
84	Hfn	62.88	51.88	0.25	1500	1195	15.6	2386.74	26.56	47.36	-23.26	2.98	1097	415	4.92	7.73	4.5	67.11	15920.68	55.25	112	1.55
85	Pfn	53.22	45.57	0.29	1160	925	14.12	532.77	21.63	36.89	-20.53	3.4	872	319	4.13	6.91	4.18	49.83	3046.44	46.55	76	1.66
86	Pfv	46.89	37.54	0.28	882	676	12.87	349.39	20.08	33.65	-16.96	2.76	593	244	4.34	7.01	4.07	47.81	3295.34	41.35	73	1.55
87	Tca	26.62	19.56	0.32	375	315	8.86	9.17	13.14	19.15	0	1.66	176	160	2.23	2.4	3.28	12.28	17.68	28.53	6	1.19

88	Sec	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.67	2.76	12.26	25.02	19.65	21	1.38
89	Pyl	79.59	84.67	0.33	3026	2827	18.81	242.25	28.68	51.06	0	5.52	2975	922	2.04	2.65	4.66	26.68	75.16	75.06	34	2.6
90	Ath	85.81	74.4	0.22	3232	3258	19.23	58.13	37.5	68.67	45.93	2.42	2176	2152	1.76	2.64	4.05	18.18	43.48	74.44	13	2.12
91	Bal	58.16	49.2	0.28	1576	1507	15.03	43.23	25.64	42.96	16.7	2.83	1018	772	1.88	2.62	3.64	15.8	34.2	58.6	13	1.51
92	Bip	77.55	75.89	0.3	2870	2815	18.38	75.51	30.7	54.21	25.08	3.52	2414	1043	1.7	2.64	3.67	15.99	43.44	67.06	13	1.36
93	Dip	74.48	63.89	0.24	2414	2339	17.63	43	31.62	56.87	22.54	2.04	1610	935	2.05	2.66	4.16	18.28	52.68	55.06	12	1.13
94	Tbt	98.47	89.82	0.22	3975	3728	20.96	2192.9	40.64	77.71	31.5	2.48	3128	1886	2.18	2.63	4.16	22.23	71.16	88.11	76	1.54
95	Tpc	120.18	124.72	0.25	6956	7429	24.19	199.05	48.63	93.17	112.02	4.05	6264	3252	1.33	4.86	5.58	31.39	79.17	112.1	17	0.61
96	Asu	50.41	49.23	0.38	1176	980	13.98	105.62	18.18	30.07	-22.45	4.55	1047	320	3.16	2.67	3.25	24.56	91.01	48.12	41	2.19
97	Hcy	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.63	2.93	13.5	30.81	24	24	1.5
98	Sta	42.97	36.67	0.33	798	627	12.36	113.49	17.14	28.04	-16.63	3.11	576	220	3.68	2.68	3.23	20.64	87.19	41.02	48	1.79
99	Thi	31.44	23.6	0.31	490	411	9.92	18.47	15.02	22.63	0	2.08	249	178	2.32	2.65	3.28	14.02	24.79	33.22	12	1.52
100	γ -Abu	18.74	14.86	0.43	190	130	7.06	9.75	8.21	11.48	-6.96	2.54	108	52	2.68	2.44	2.7	11.41	16.29	19.65	17	1.56
101	Aca	28.9	25.33	0.43	416	320	9.64	16.52	11.26	16.84	-11.66	3.45	297	114	2.75	2.38	2.93	12.62	25.59	28.53	21	1.98
102	Ach	30.84	22.2	0.29	464	384	9.72	16.09	15.4	23.23	0	1.49	220	195	2.48	2.51	3.55	15.65	32	29.22	11	1.19
103	Afb	69.94	62.59	0.27	2075	1847	16.96	263.56	28.51	50.74	0	2.75	1616	700	2.55	2.96	4.02	36.96	123.52	63.49	48	1.69
104	Aoq	63.61	54.25	0.26	1840	1767	15.92	46.51	27.63	47.43	18.17	2.78	1213	970	1.92	2.68	3.68	19.35	40.35	64	13	1.45
105	Bpa	89.92	91.6	0.29	3806	3733	20.22	168.62	34.61	63.24	31.21	3.9	3464	1324	1.73	2.7	5.17	26.16	64.04	78.44	20	1.33
106	Mas	31.66	24.2	0.31	436	313	9.89	60	13.99	21.8	-11.07	2.23	257	121	3.81	2.96	3.1	23.22	100.65	33.22	44	1.37
107	Ceg	22.71	17.5	0.37	254	175	7.98	17.32	10.11	14.73	-8.12	2.23	143	70	3.17	2.66	2.91	14.33	32.69	24	24	1.5
108	Cha	53.94	48.14	0.32	1340	1196	14.43	64.31	21.99	36.64	0	3.04	997	469	2.22	2.56	3.54	18.87	46.87	49.3	23	1.73
109	Dty	68.72	58.12	0.25	1931	1711	16.65	289.73	29	51.96	0	2.51	1363	662	2.75	2.86	4.2	26.84	104.21	65.49	48	1.69
110	Chg	36.61	28.36	0.3	641	553	11.03	22.83	16.8	26.12	0	2.15	353	252	2.32	2.52	3.47	15.15	32.01	34.05	12	1.46
111	Cpa	48.11	41.23	0.31	1077	955	13.34	50.41	20.25	32.99	0	2.9	735	388	2.22	2.66	3.42	17.75	39.22	44.11	20	1.66
112	Deg	26.34	19.11	0.3	310	208	8.67	28.27	12.4	18.75	-8.8	1.46	163	86	3.92	2.58	3.36	14.61	47.78	24.53	35	1.35
113	Dmt	58.11	49.07	0.27	1449	1273	14.95	174.58	24.62	42.21	0	2.73	999	566	2.52	2.68	3.62	22.2	55.45	52.6	40	1.57
114	Dvg	27.88	22.67	0.38	370	272	9.27	23.94	11.67	17.55	-10.44	2.73	235	102	3.16	2.7	3	13.24	33.36	28.53	25	1.84
115	Gav	42.93	36.67	0.32	796	627	12.38	75.47	16.95	27.82	-16.63	2.6	568	220	3.7	2.68	3.49	20.17	65.9	43.02	42	1.89
116	Hat	57.2	46.13	0.26	1476	1394	14.75	53.13	26.45	44.39	13.55	1.78	903	810	2.01	2.62	3.98	24.11	50.3	58.6	18	1.03
117	Hai	51.39	40	0.26	1203	1120	13.69	44.74	24.57	40.29	12.14	1.74	682	636	2.05	2.66	3.89	23.51	46.97	53.3	18	1.06
118	Hpp	57.86	48.27	0.27	1545	1479	14.96	29.16	25.75	43.23	15.85	2.44	978	669	1.92	2.48	3.94	21.06	38.07	54.6	11	0.99
119	Ing	51.68	40.86	0.26	1233	1164	13.81	33.09	24.12	39.65	13.28	2.16	705	650	2	2.6	3.76	16.63	32.91	53.3	12	1.29
120	Mhp	60.35	57.07	0.33	1699	1541	15.61	82.54	23.56	39.95	0	3.67	1390	572	2.13	2.62	3.48	17.97	45.6	54.6	24	1.94
121	Oct	50.78	50.77	0.4	1214	1020	14.13	72.19	17.93	29.64	-22.94	4.69	1118	330	3.07	2.55	3.34	14.75	65.73	48.11	33	2.47
122	Oic	41.12	31.33	0.28	838	792	11.87	13.45	20.14	31.37	9.3	1.75	431	465	1.94	2.34	3.68	14.35	23.79	43.02	6	1.03
123	Pal	42.57	35.33	0.32	864	766	12.31	30.77	18.32	29.17	0	2.82	542	320	2.2	2.65	3.36	16.05	33.16	43.02	13	1.7
124	Tic	46.59	36.62	0.28	1048	1004	12.93	15.17	22	35.22	10.59	1.76	580	606	1.91	2.44	3.7	14.01	23.99	48.11	6	0.99
125	Thz	21.8	15.5	0.33	270	217	7.68	7.68	11.42	16	0	1.63	114	106	2.24	2.41	3.13	11.57	14.42	24	6	1.25
126	Tle	26.38	19.11	0.31	312	208	8.64	8.64	12.49	18.83	-8.8	1.87	164	86	3.88	2.61	3.22	13.82	45.4	23.77	41	1.22
127	Dpg	68.03	55.77	0.23	2008	1938	16.59	31.86	30.3	53.43	19.13	2.16	1251	810	2.11	2.75	4.5	19.99	57.13	49.49	11	0.92
128	Dbz	81.7	75.16	0.26	3010	29.68	18.89	43.21	33.2	60.33	27.3	3.26	2254	1098	1.91	2.64	4.51	20.08	64.37	58.71	11	0.87
129	β -Phe	42.38	34.83	0.31	850	750	12.26	27.7	18.38	29.3	0	2.67	526	317	2.24	2.54	3.3	14.77	29.1	39.02	13	1.7
130	α -Abu	17.87	13.14	0.37	168	106	6.67	12.45	8.54	12	-6.16	1.78	83	46	3.14	2.6	2.81	11.76	22.89	19.65	21	1.38
131	Mpr	26.18	18.67	0.3	354	284	8.66	12.61	13.38	19.5	0	1.59	159	145	2.4	2.39	3.35	12.55	19.7	28.53	11	1.23
132	Hva	26.38	19.11	0.31	312	208	8.64	41.12	12.49	18.83	-8.8	1.87	164	86	3.88	2.86	3.06	18.58	72.85	26.53	41	1.22
133	Dcp	58.53	50.4	0.28	1489	1313	15.05	174.55	24.5	41.88	0	2.86	1061	556	2.44	2.82	3.46	30.42	78.29	52.6	40	1.57
134	Car	49.11	44.62	0.34	1058	860	13.6	162.77	18.72	31.21	-20.16	3.79	851	290	3.55	3.02	3.91	26.25	124.17	48.11	49	2.05

附表 4 所选 134 个氨基酸分子拓扑与化学结构信息 6 个显著主成分得分矢量(VTSA)

No.	Abbr	Name	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆
1	Ala	alanine	-9.11	-1.63	0.63	1.04	2.26	-0.204
2	Arg	arginine	0.23	3.89	-1.16	-0.39	-0.06	1.051
3	Asn	asparagine	-4.62	0.66	1.16	-0.22	0.93	0.292
4	Asp	aspartic acid	-4.65	0.75	1.39	-0.40	1.05	0.309
5	Cys	cysteine	-7.35	-0.86	-0.33	0.80	0.98	0.645
6	Gln	glutamine	-3.00	1.72	0.28	-0.39	0.33	0.984
7	Glu	glutamic acid	-3.03	1.82	0.51	-0.58	0.43	0.998
8	Gly	glycine	-10.61	-1.21	-0.12	0.75	3.25	-0.227
9	His	histidine	-1.01	-1.31	0.01	-1.81	-0.21	1.544
10	i	isoleucine	-4.25	-0.28	-0.15	1.40	-0.21	-0.204
11	Leu	leucine	-4.38	0.28	-0.49	1.45	0.02	-0.151
12	Lys	lysine	-2.59	2.34	-1.69	0.41	-0.21	0.059
13	Met	methionine	-4.08	0.98	-2.34	1.64	-0.79	-0.132
14	Phe	phenylalanine	0.49	-0.94	-0.63	-1.27	-0.44	0.679
15	Pro	proline	-5.11	-3.54	-0.53	-0.36	-0.29	0.872
16	Ser	serine	-7.44	-0.65	0.68	-0.17	1.58	1.032
17	Thr	threonine	-5.97	-0.62	1.11	0.31	0.95	0.910
18	Trp	tryptophan	5.73	-2.67	-0.07	-1.96	-0.54	1.363
19	Tyr	tyrosine	2.08	-0.47	0.07	-1.67	-0.35	1.305
20	Val	valine	-5.87	-0.94	0.28	1.10	0.48	0.721
21	Acp	α -aminocaprylic acid	-0.89	3.26	-3.07	0.91	-0.92	0.286
22	Aec	(S)-2-aminoethyl-L-cysteine-HCl	-2.46	2.10	-2.56	1.55	-1.01	-0.919
23	Afa	aminophenylacetate	-1.05	-2.07	0.22	-1.58	-0.27	1.353
24	Aib	α -aminoisobutyric acid	-7.66	-1.90	1.46	1.66	1.42	-0.447
25	Ail	alloisoleucine	-4.38	0.28	-0.49	1.45	0.02	-0.254
26	Alg	L-allylglycine	-5.93	0.06	-0.07	0.12	0.94	0.108
27	Aba	α -aminobutyric acid	-7.34	-0.93	-0.07	0.53	1.15	0.750
28	Aph	p-aminophenylalanine	2.08	-0.57	-0.21	-1.46	-0.43	1.249
29	β -Ala	β -alanine	-8.79	-0.23	-0.98	0.11	2.06	0.807
30	Brp	p-bromophenylalanine	2.14	-0.93	-1.38	-0.18	-1.23	0.554
31	Cha	cyclohexylalanine	1.04	-1.16	-2.26	0.91	-1.86	-0.627
32	Cit	citrulline	0.22	3.95	-0.95	-0.55	-0.00	0.803
33	Cla	β -chloroalanine	-7.41	-0.81	-0.06	0.53	1.14	0.515
34	Cle	cycloleucine	-3.53	-3.86	0.16	0.30	-0.96	0.371
35	Clp	p-chlorophenylalanine	2.07	-0.66	-0.63	-1.00	-0.74	0.880
36	Cya	cysteic acid	-2.95	0.84	2.57	1.41	-0.04	-1.770
37	Dab	2,4-diaminobutyric acid	-5.83	0.21	-0.16	0.28	0.84	-0.061
38	Dap	2,3-diaminopropionic acid	-7.41	-0.75	0.42	0.03	1.47	0.858
39	Dhp	3,4-dehydroproline	-5.29	-3.45	-0.00	-1.06	0.18	1.056
40	Dha	3,4-dihydroxyphenylalanine	3.60	-0.14	1.04	-1.79	-0.24	1.263
41	Fph	p-fluorophenylalanine	2.33	-0.47	0.52	-1.93	-0.52	0.738
42	Gaa	D-glucoseaminic acid	1.69	2.87	2.77	-0.12	-0.45	0.366
43	Hag	homoarginine	0.23	3.89	-1.16	-0.39	-0.06	0.745
44	Hly	δ -hydroxylysine · HCl	-1.23	2.55	-0.48	0.24	-0.31	-0.040
45	Hnv	DL- β -hydroxynorvaline	-4.36	0.09	0.74	0.58	0.30	-0.213
46	Hog	homoglutamine	-1.36	2.80	-0.43	-0.08	0.02	-0.051
47	Hop	homophenylalanine	2.17	-0.08	-1.46	-1.28	-0.65	0.927
48	Hos	homoserine	-5.84	0.30	0.09	0.08	0.95	-0.132
49	Hpr	hydroxyproline	-3.67	-2.92	0.28	-0.51	-0.24	-0.126
50	Iph	p-iodophenylalanine	2.21	-1.02	-1.89	0.37	-1.56	0.152
51	Ise	isoserine	-9.24	-1.37	1.37	0.42	2.70	-0.463
52	Mle	α -methylleucine	-2.88	-0.26	0.44	2.19	-0.71	-0.615

53	Msm	DL-methionine-s-methylsulfoniumchloride	-2.46	0.79	-2.97	3.66	-2.12	-1.264
54	1Nala	β -(1-naphthyl)alanine	7.16	-2.66	-0.23	-1.71	-0.51	1.026
55	2Nala	β -(2-naphthyl)alanine	7.35	-2.26	-0.51	-1.70	-0.25	0.483
56	Nle	norleucine(or 2-aminohexanoic acid)	-4.16	1.15	-1.56	0.63	-0.11	0.296
57	Nma	N-methylalanine	-7.35	-1.01	-0.23	0.54	1.14	0.604
58	Nva	norvaline(or 2-aminopentanoic acid)	-5.77	0.03	-0.66	0.79	0.54	-0.397
59	Obs	O-benzylserine	3.80	1.23	-1.52	-1.82	-0.15	0.568
60	Obt	O-benzyltyrosine	14.75	0.43	-2.06	-1.86	2.44	0.094
61	Oet	O-ethyltyrosine	5.45	1.43	-1.41	-1.46	-0.76	0.874
62	Oms	O-methylserine	-5.89	0.25	-0.02	0.03	1.01	-0.091
63	Omt	O-methylthreonine	-4.32	-0.05	0.53	0.61	0.27	-0.303
64	Omy	O-methyltyrosine	3.72	0.14	-0.60	-1.39	-0.37	0.480
65	Orn	ornithine	-4.21	1.33	-1.06	0.11	0.18	0.406
66	Pen	penicillamine	-4.42	-1.12	0.66	2.27	-0.38	0.012
67	Pga	pyroglutamic acid	-3.74	-2.95	0.68	-0.93	-0.06	-0.236
68	Pip	pipecolic acid	-3.62	-3.09	-0.90	0.02	-0.66	-0.454
69	Sar	sarcosine	-8.81	-0.29	-1.10	0.14	2.06	0.564
70	Tfa	3,3,3-trifluoroalanine	-4.56	0.24	5.16	-0.76	1.50	-0.454
71	Thp	6-hydroxydopa	5.12	0.20	2.17	-1.79	-0.20	1.152
72	Vig	L-vinylglycine	-7.51	-0.90	0.50	-0.12	1.55	0.675
73	Aas	(-)-(2R)-2-amino-3-(2-aminoethylsulfonyl)propanoic acid dihydrochloride	0.79	2.08	1.87	1.04	-1.03	-1.744
74	Ahd	(2S)-2-amino-9-hydroxy-4,7-dioxanonanoic acid	2.22	6.21	-2.41	-1.11	0.06	0.555
75	Aho	(2S)-2-amino-6-hydroxy-4-oxahexanoic acid	-2.73	2.71	-0.74	-0.64	0.42	0.005
76	Ahs	(-)-(2R)-2-amino-3-(2-hydroxyethylsulfonyl)propanoic acid	0.78	2.20	2.17	0.82	-0.91	-1.745
77	Ahp	(-)-(2R)-2-amino-3-(2-hydroxyethylsulfanyl)propanoic acid	-2.48	2.20	-2.30	1.32	-0.90	-1.292
78	Ahd	(2S)-2-amino-12-hydroxy-4,7,10-trioxadodecanoic acid	7.68	9.71	-3.91	-0.93	0.68	-0.126
79	Dad	(2S)-2,9-diamino-4,7-dioxanonanoic acid	2.22	6.13	-2.65	-0.90	-0.01	0.514
80	Dat	(2S)-2,12-diamino-4,7,10-trioxadodecanoic acid	7.70	9.63	-4.15	-0.72	0.61	-0.150
81	Dfn	(S)-5,5-difluoronorleucine	-1.26	2.19	1.98	0.15	-0.12	-0.588
82	Dfv	(S)-4,4-difluoronorvaline	-2.86	1.08	2.94	0.28	0.38	-1.327
83	Dtc	(3R)-1-1-dioxo-[1] and [4]thiaziane-3-carboxylic acid	0.64	-2.37	1.56	0.15	-1.19	-2.881
84	Hfn	(S)-4,4,5,5,6,6,6-heptafluoronorleucine	7.11	6.32	15.84	1.43	-2.07	0.320
85	Pfn	(S)-5,5,6,6,6-pentafluoronorleucine	3.43	4.14	7.69	-0.34	-0.27	-1.408
86	Pfv	(S)-4,4,5,5,5-pentafluoronorvaline	1.69	3.15	8.24	-0.81	-0.00	-0.838
87	Tca	(3R)-1,4-thiazinane-3-carboxylic acid	-3.50	-3.41	-1.74	1.17	-1.49	-1.517
88	Sec	selenocysteine	-7.31	-1.45	-1.02	1.76	0.32	-0.525
89	Pyl	pyrrollysine	11.25	3.40	-1.91	-0.72	-0.13	-0.536
90	Ath	β -(9-anthracenyl)alanine	14.19	-3.65	-0.21	-1.74	0.06	0.498
91	Bal	β -(3-benzothienyl)alanine	5.89	-3.15	-1.21	-0.47	-1.60	-0.287
92	Bip	β -(4,4'-biphenyl)alanine	10.77	-1.14	-1.30	-1.77	0.76	0.101
93	Dip	β , β -diphenylalanine	10.02	-2.62	-0.06	-1.47	-0.18	0.518
94	Tbt	β -[3-(2,5,7-tri-tert-butyl-indolyl)]alanine	24.18	-1.01	0.76	13.40	-2.14	5.977
95	Tpc	β -{3-[2-(2,2,5,7,8-pentamethyl-chroman-6-sulfonyl)-indolyl]}alanine	36.90	-4.22	-0.72	5.46	8.53	-4.978
96	Asu	Aminosuberic acid	2.03	5.00	-1.72	0.1	-0.40	-0.586
97	Hcy	homocysteine	-5.77	0.11	-0.91	1.07	0.36	-0.960
98	Sta	statine	0.24	2.51	-0.67	1.60	-0.97	-0.714
99	Thi	β -(2-thienyl)alanine	-2.35	-2.75	-0.52	0.04	-1.04	-0.991
100	γ -Abu	L- γ -aminobutyric acid	-7.20	0.89	-1.63	0.36	1.40	-0.647
101	Aca	ϵ -aminocaproic acid	-3.95	3.02	-3.14	0.46	0.22	-0.718
102	Ach	1-aminocyclohexane-1-carboxylic acid	-2.04	-3.43	-0.22	0.70	-1.28	-1.076
103	Afb	β -amino- β -phenyl-p-nitro-L-butyric acid	8.47	0.86	2.47	-1.57	-0.17	0.096
104	Aoq	α -amino- β -[4-(1,2-dihydro-2-oxo-quinoliny)]propionic acid	7.16	-2.34	0.28	-2.26	-0.21	0.888
105	Bpa	4'-benzoylphenylalanine	15.14	-0.69	-0.58	-3.65	3.19	-1.042
106	Mas	β -methyl aspartic acid	-3.09	0.41	2.11	0.15	0.34	-0.518

107	Ceg	2-chloroethylglycine	-5.83	0.16	-0.65	0.79	0.50	-0.957
108	Cha	β -cyclohexyl(p-methoxyl)-L-alanine	4.29	-0.07	-2.20	0.78	-1.79	-1.065
109	Dty	α,β -divinyltyrosine	8.27	-0.13	1.89	-0.83	-0.80	0.690
110	Chg	2-L-cyclohexylglycine	-0.51	-2.32	-1.39	0.56	-1.68	-0.451
111	Cpa	4-chlorophenylalanine	2.07	-0.66	-0.63	-1.00	-0.74	0.312
112	Deg	α,α -diethyl glycine	-4.28	-1.30	0.72	1.42	-0.38	-0.224
113	Dmt	2',6'-dimethyltyrosine	5.26	-0.56	0.59	-0.32	-1.04	0.458
114	Dvg	divinyl glycine	-4.50	1.18	-0.44	-0.70	0.69	0.427
115	Gav	2-guanidine-5-amino-L - n-valeric acid	0.02	2.83	-0.10	0.08	-0.41	-0.488
116	Hat	2-amino-6-hydroxytetralin-2-carboxylic acid	6.03	-3.75	0.94	-1.03	-1.10	0.042
117	Hai	2-amino-5-hydroxyindan-2-carboxylic acid	4.40	-3.99	1.43	-1.16	-1.02	-0.412
118	Hpp	3-(4'-hydroxyphenyl)proline	5.88	-3.31	-0.01	-1.38	-0.88	0.095
119	Ing	1-indanylglycine	4.34	-3.82	-0.1	-0.80	-1.40	-0.581
120	Mhp	p-methoxyhomophenylalanine	5.48	1.01	-1.43	-1.29	-0.39	0.399
121	Oct	n-octylglycine	2.52	5.39	-4.60	1.36	-1.35	-0.806
122	Oic	octahydroindole-2-carboxylic acid	1.66	-4.70	-1.30	0.80	-2.21	-1.986
123	Pal	β -pyridylalanine	0.45	-0.86	-0.33	-1.59	-0.28	-0.017
124	Tic	1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid	2.69	-4.12	-0.21	-1.28	-0.90	-0.300
125	Thz	L-4-thiazolidine carboxylic acid	-5.01	-3.85	-1.33	0.72	-1.07	-0.820
126	Tle	L-tert-butylglycine	-4.43	-1.15	0.89	1.88	-0.13	-0.201
127	Dpg	diphenylglycine	8.69	-3.44	0.94	-2.00	-0.38	0.339
128	Dbz	dibenzylglycine	11.79	-2.00	-0.54	-1.27	-0.16	-0.048
129	β -Phe	β -phenylalanine	0.49	-1.07	-0.62	-1.27	-0.49	-0.140
130	α -Abu	α -aminobutyric acid	-7.34	-0.93	-0.07	0.52	1.15	-0.018
131	Mpr	3-methylproline	-3.56	-3.67	-0.36	0.17	-0.85	-0.166
132	Hva	3-hydroxyvaline	-4.52	-0.77	1.81	1.05	0.34	-0.064
133	Dep	3,5-dihydroxy-4-chloro-phenylalanine	5.17	-0.03	1.42	-1.16	-0.58	0.371
134	Car	β -carbonylarginine	1.41	3.64	1.32	-1.50	0.09	-0.058

附表 5 血管紧张素转化酶(ACE)抑制剂: 58 个二肽样本集一级序列及生物活性

No.	Peptide	pIC ₅₀ (obsd)	pIC ₅₀ (cald_1)	pIC ₅₀ (cald_2)	No.	Peptide	pIC ₅₀ (obsd)	pIC ₅₀ (cald_1)	pIC ₅₀ (cald_2)
1	VW	5.80	5.26	5.10	30	KG	2.49	2.63	2.62
2	IW	5.70	5.46	5.29	31	FG ^Δ	2.43	2.34	2.35
3	IY	5.43	4.91	4.62	32	GS ^Δ	2.42	2.30	2.61
4	AW	5.00	5.19	4.76	33	GV ^Δ	2.34	2.59	2.48
5	RW	4.80	5.07	5.38	34	MG	2.32	2.99	2.71
6	VY ^Δ	4.66	4.21	4.43	35	GK	2.27	2.24	2.33
7	GW	4.52	4.63	4.56	36	GE	2.27	2.30	2.90
8	VF ^Δ	4.28	4.10	4.23	37	GT ^Δ	2.24	2.49	2.68
9	AY ^Δ	4.06	3.84	4.09	38	WG	2.23	2.22	2.32
10	IP	3.89	4.16	4.17	39	HG ^Δ	2.20	2.19	2.22
11	RP	3.74	3.78	4.26	40	GQ ^Δ	2.15	2.33	2.45
12	AF ^Δ	3.72	3.73	3.88	41	GG	2.14	1.99	1.86
13	GY ^Δ	3.68	3.58	3.89	42	QG ^Δ	2.13	2.36	2.47
14	AP	3.64	3.59	3.63	43	SG ^Δ	2.07	2.15	2.10
15	RF	3.64	3.92	4.51	44	LG ^Δ	2.06	2.78	2.59
16	VP ^Δ	3.38	3.96	3.97	45	GD	2.04	2.34	2.79
17	GP	3.35	3.33	3.44	46	TG ^Δ	2.00	2.37	2.29
18	GF ^Δ	3.20	3.47	3.69	47	EG ^Δ	2.00	2.30	2.45
19	IF ^Δ	3.03	4.30	3.42	48	DG	1.85	2.21	2.31
20	VG ^Δ	2.96	2.62	2.60	49	PG	1.77	2.49	2.15
21	IG	2.92	2.82	2.59	50	LA ^Δ	3.51	3.10	3.48
22	GI ^Δ	2.92	2.65	2.66	51	KA	3.42	2.94	2.92
23	GM	2.85	2.36	2.09	52	RA	3.34	2.75	2.98
24	GA ^Δ	2.70	2.30	2.16	53	YA	3.34	2.55	2.68
25	YG	2.70	2.24	2.38	54	AA ^Δ	3.21	2.56	2.95
26	GL ^Δ	2.60	2.49	2.28	55	FR	3.04	2.60	3.04
27	AG ^Δ	2.60	2.25	2.06	56	HL	2.49	2.69	2.63
28	GH ^Δ	2.51	2.35	2.85	57	DA ^Δ	2.42	2.52	2.60
29	GR	2.49	2.25	2.55	58	EA ^Δ	2.00	2.61	2.34

附表 6 人工合成的弹性蛋白酶模拟底物序列和活性: ($\lg 1/K_m$, $\lg k_{cat}$, $\lg k_{cat}/K_m$)

No.	Code	$\lg(1/K_m)$	$\lg k_{cat}$	$\lg(k_{cat}/K_m)$	No.	Code	$\lg(1/K_m)$	$\lg k_{cat}$	$\lg(k_{cat}/K_m)$
1	GlyAla	-0.354	0.049	2.695	45	IlePro	0.071	1.274	4.346
2	GlyVal	-0.231	-0.105	2.703	46	IlePhe	-0.188	0.158	2.971
3	GlyLeu	0.125	-0.796	2.328	47	IleAbu	-0.086	1.068	3.982
4	GlyIle	0.036	-0.337	2.704	48	IleNva	-0.228	1.061	3.833
5	GlyPro	-0.322	0.068	2.746	49	IleNle	0.585	0.407	3.992
6	GlyPhe	-0.260	-0.854	1.881	50	PheGly	-0.812	0.037	2.225
7	GlyAbu	0.059	-1.121	2.832	51	PheAla	-0.425	0.890	3.465
8	GlyNva	-0.220	0.041	2.882	52	PheVal	-0.207	0.511	3.303
9	GlyNle	0.149	-0.155	2.994	53	PheLeu	-0.344	0.310	2.965
10	AlaGly	-0.686	0.107	2.422	54	Phelle	-0.255	0.490	3.236
11	AlaAla	-0.225	0.936	3.711	55	PhePro	-0.107	1.000	3.893
12	AlaVal	-0.049	0.979	3.930	56	PhePhe	0.013	-0.409	2.605
13	AlaLeu	-0.193	0.515	3.322	57	PheAbu	0.053	0.556	3.610
14	AlaIle	-0.083	0.951	3.869	58	PheNva	-0.033	0.612	3.579
15	AlaPro	-0.207	1.436	4.230	59	PheNle	0.108	0.553	3.660
16	AlaPhe	-0.072	0.375	3.303	60	AbuGly	-0.906	0.565	2.658
17	AlaAbu	-0.114	1.033	3.920	61	AbuAla	-0.155	1.086	3.931
18	AlaNva	-0.061	1.045	3.985	62	AbuVal	0.215	0.619	3.834
19	AlaNle	0.638	0.622	4.265	63	AbuLeu	-0.037	0.471	3.435
20	ValGly	-0.493	0.467	2.974	64	Abulle	0.222	0.653	3.876
21	ValAla	-0.215	1.185	3.970	65	Abupro	-0.362	1.581	4.220
22	ValVal	0.041	0.876	3.920	66	AbuPhe	0.252	-0.076	3.179
23	ValLeu	-0.243	0.561	3.318	67	AbuAbu	0.060	1.045	4.107
24	ValIle	-0.170	0.880	3.710	68	AbuNva	-0.244	0.787	4.029
25	ValPro	0.033	1.476	4.509	69	AbuNle	0.538	0.589	4.124
26	ValPhe	0.018	0.210	3.225	70	NvaGly	-0.377	0.199	2.822
27	ValAbu	0.276	0.866	4.140	71	NvaAla	-0.362	1.060	3.699
28	ValNva	0.060	0.965	4.025	72	NvaVal	0.051	0.814	3.863
29	ValNle	0.228	0.864	4.090	73	NvaLeu	-0.246	0.471	3.226
30	LeuGly	-0.559	0.188	2.628	74	NvaIle	-0.228	0.860	3.631
31	LeuAla	-0.326	1.107	3.781	75	NvaPro	-0.143	1.380	4.238
32	LeuVal	-0.185	0.780	3.595	76	NvaPhe	0.222	-0.102	3.124
33	LeuLeu	-0.340	0.412	3.072	77	NvaAbu	0.131	0.925	4.061
34	Leulle	-0.452	0.991	3.539	78	NvaNva	0.316	0.741	4.057
35	Leupro	0.142	0.996	4.140	79	NvaNle	0.569	0.604	4.057
36	Leuphe	-0.152	0.033	2.881	80	NleGly	-0.851	0.581	2.730
37	LeuAbu	-0.158	1.053	3.895	81	NleAla	-0.328	1.143	3.814
38	LeuNva	-0.045	0.925	3.880	82	NleVal	-0.188	0.840	3.653
39	LeuNle	0.071	0.956	4.041	83	NleLeu	-0.371	0.303	2.947
40	IleGly	-0.708	0.607	2.900	84	NleIle	-0.068	0.730	3.662
41	IleAla	-0.371	1.210	3.838	85	NlePro	-0.196	1.380	4.185
42	IleVal	-0.252	0.927	3.674	86	NlePhe	0.244	-0.086	3.158
43	IleLeu	-0.312	0.358	3.045	87	NleAbu	0.215	0.812	4.025
44	IleIle	-0.152	0.738	3.585	88	NleNva	-0.262	1.176	3.914