

## 固有无序蛋白与蛋白质相互作用位点残基特征分析\*

董川<sup>1)</sup> 曹赞霞<sup>1, 2)</sup> 赵立岭<sup>1, 2)</sup> 索振鹏<sup>1)</sup> 王吉华<sup>1, 2)\*\*</sup>

<sup>1)</sup>德州学院物理与电子信息学院, 德州 253023; <sup>2)</sup>山东省功能大分子生物物理重点实验室(德州学院), 德州 253023)

**摘要** 本文对固有无序蛋白(IDPs)与其他蛋白质相互作用位点残基特征进行了研究。首先在数据库中选出满足条件的 109 条 IDPs 蛋白质链及与其他配体蛋白形成的 299 个 IDPs-蛋白质复合物, 然后提取复合物中作为相互作用位点的 IDPs-蛋白质残基。这 109 条 IDPs 链中共含有 50 031 个氨基酸残基, 其中处于作用位点的残基有 4 822 个。通过分析发现, 20 种氨基酸在形成 IDPs-蛋白质相互作用位点残基时具有不同的倾向性, 根据形成作用位点残基的倾向性, 20 种氨基酸可分成三大类: 倾向型氨基酸(ILE、LEU、ARG、PHE、TYR、MET、TRP)、中间型氨基酸(GLN、GLU、THR、LYS、VAL、ASP、HIS)、非倾向型氨基酸(PRO、SER、GLY、ALA、ASN、CYS)。研究结果还进一步表明, 不同氨基酸在有序区域与无序区域形成 IDPs-蛋白质作用位点残基的倾向性不同。其中, 氨基酸 TRP、LEU、ILE、CYS 在有序和无序区域形成作用位点残基的差异性尤为明显, 而氨基酸 GLU、PHE、HIS、ALA 则基本没有多大差别。对 IDPs-蛋白质相互作用位点残基理化特征进行分析发现: 疏水性强、侧链净电荷量较少、极性较小、溶剂可及性表面积较大、侧链体积较大、极化率较大的氨基酸比较倾向于形成作用位点残基。主成分分析结果显示, 残基的极化率、侧链体积和溶剂可及表面积对作用位点残基影响最大。

**关键词** 固有无序蛋白(IDPs), 作用位点残基特征, 有序区, 无序区, 氨基酸的理化性质

**学科分类号** Q615, Q7

**DOI:** 10.3724/SP.J.1206.2013.00214

蛋白质的变性理论在 20 世纪二三十年代被提出, 随后 Linderstrom-Lang 提出了蛋白质的一级结构、二级结构、三级结构学说, 在此基础上提出了蛋白质的序列决定结构、结构决定功能的假说。随后, 这一假说得到了实验的证实。然而, 20 世纪 90 年代发现的固有无序蛋白质 (intrinsically disordered protein, IDPs) 或称为天然无折叠蛋白质 (natively unfolded protein, NUP) 却打破了“序列-结构-功能”的金科玉律<sup>[1-2]</sup>。

IDPs 是一类与以往的传统蛋白质不同类型的蛋白质。这类蛋白质在天然孤立状态下, 完全没有稳定的三维结构<sup>[3-6]</sup>, 称为完全无序蛋白; 或者部分区域没有稳定的结构, 称为部分无序蛋白, 相应的区域称为无序区 (intrinsically disordered regions), 有稳定结构的区域称为有序区 (ordered regions)。由于 IDPs 无序区域的氨基酸残基有较低的疏水性以及较大的静电排斥作用, 这导致了无序区呈现一种松散的状态。当 IDPs 与配体结合后, 由于净电荷的降低, IDPs 的无序区会倾向于形成折叠状

态<sup>[7-8]</sup>。该类蛋白在自然界中极为丰富, 并且履行着很重要的生物学功能, 通常是蛋白质相互作用网络的关键蛋白。IDPs 在分子识别、细胞周期调控<sup>[9-10]</sup>、细胞信号转导<sup>[11]</sup>、细胞内调节等多种生理活动中起到关键作用, 细胞中多种生理活动的进行是通过蛋白质与其配体相互作用来实现的<sup>[12]</sup>, IDPs-配体的相互作用亦介导了重要的生物学功能。此外, IDPs 还与一些重大疾病的发生有密切关系<sup>[13]</sup>, 如与癌症相关的 P53 蛋白、与帕金森疾病相关的  $\alpha$ -synuclein 和 tau 蛋白等均是固有无序蛋白。近年来有文献指出<sup>[14-15]</sup>, IDPs 的无序区域在与生物分子的相互作用中也起到了重要的作用。IDPs 的无序区域可以参与一对多或者多对一的相互作用

\* 国家自然科学基金(31000324, 61271378, 30970561)和山东省自然科学基金(ZR2011FL011, ZR2012CL09) 资助项目。

\*\* 通讯联系人。

Tel: 0534-8985933, E-mail: jhw25336@126.com

收稿日期: 2013-05-17, 接受日期: 2013-08-14

用, 这是 IDPs 功能多样性的重要原因<sup>[6]</sup>. 因此, 研究固有无序蛋白与其他生物大分子的相互作用, 尤其是与其他蛋白质的相互作用具有重要意义.

随着实验技术的改进及提高, 被发现的 IDPs 的数量正在迅速增加. 由美国印第安那大学生物信息中心创建并维护的无序蛋白数据库(<http://www.disprot.org>)中, 所收录的无序蛋白及无序区域的数目已经由 2004 年第 1 版的 154 个和 190 条增加到 2012 年最新版的 654 个和 1 513 条. 此外, IDPs 所具有的未知生物学功能也正在进一步被揭示.

对 IDPs 的研究已经成为生物学领域的热点问题之一. 目前, 国内外针对 IDPs 开展了大量的研究工作并取得了一系列的成果. 本课题组<sup>[7]</sup>曾对 IDPs 有序与无序接点处的残基特征进行了分析, 发现 IDPs 接点两侧氨基酸的疏水性有较大差别. Wojtas 等<sup>[8]</sup>研究了抗衡离子(counter ion)与 IDPs 的作用, 结果表明抗衡离子会通过减小 IDPs 的静电排斥力进而改变 IDPs 的构象. 有的研究小组<sup>[9]</sup>针对 IDPs 展开了分子动力学模拟研究, 并发展出针对 IDPs 的全原子及原子团的模拟方法. Fuxreiter 小组<sup>[20]</sup>研究了 24 条 IDPs 在与配体结合状态时骨架转角和二级结构的分布, 以及配体对 IDPs 折叠的影响, 发现形成 coil 结构的氨基酸占较大比例, 即使 IDPs 与配体结合时, 它们的结构也并不是完全有序的. 除此之外, 有很多学者基于现有的蛋白质-配体结构数据, 开发了一系列的软件用于预测蛋白质与其他配体相互作用时的作用位点残基<sup>[21-23]</sup>. 但是, 一方面, 这些预测软件的预测精度有待进一步提高, 另一方面, 尚未见针对 IDPs 与其他成分相互作用位点残基预测的报道. 随着实验技术及计算方法的发展, 结合位点的残基特征将被进一步挖掘, 这些都有利于进一步提高预测精度, 与此同时, 随着 IDPs 与其他蛋白质相互作用研究的深入, 可发展新的针对 IDPs 与其他成分相互作用位点残基的预测软件.

本文基于现有的数据库对固有无序蛋白(IDPs)与其他蛋白质相互作用位点残基特征进行了研究. 首先提取出所有相应复合物中的作用位点残基, 接着对 IDPs-蛋白质复合物的作用位点残基进行了分析.

## 1 数据集的构建与研究方法

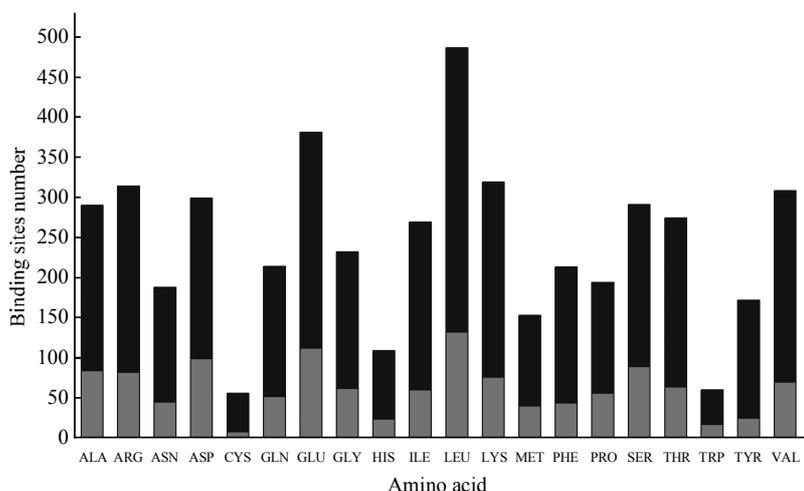
本研究中固有无序蛋白取自固有无序蛋白数据

库(Disprot, <http://www.disprot.org/>), 蛋白质复合物结构来源于蛋白质结构数据库(PDB, <http://www.rcsb.org/pdb/home/home.do>). PDB 数据库中无序蛋白-蛋白质复合物的选取满足以下条件: a. 所研究的 IDPs 链在复合物中没有突变; b. 对采用 X 射线晶体衍射技术得到的结构, 要求分辨率在 5Å 以内; c. 对采用核磁共振技术得到的结构, 选取 PDB 文件中第一个结构; d. 不考虑非上述 b、c 中的测定技术得到的复合物结构. 根据这 4 个条件, 筛选出 126 条 IDPs 链. 为了避免对相似性较高的 IDPs 链重复研究, 使数据集更加可靠, 我们基于 PISCES 在线软件(<http://dunbrack.fccc.edu/PISCES.php>), 对筛选出的 126 条固有无序蛋白质链进行了去冗余处理. 选择 IDPs 链的总长度大于 40 个氨基酸的蛋白质链, 并且将相似性高于 25% 的 IDPs 链全部去除. 最后我们筛选出数据集中符合条件的 109 条 IDPs 链及 299 个 IDPs-蛋白质复合物. 在以往的研究中<sup>[24-25]</sup>, 蛋白质-蛋白质相互作用位点残基是根据目标蛋白与其他配体蛋白之间的距离来定义的, 当目标蛋白的氨基酸残基与其他配体蛋白氨基酸残基之间的距离小于某个阈值时, 就认为该氨基酸是目标蛋白-蛋白质的作用位点残基. 本文亦根据残基间距离来定义 IDPs-蛋白质的作用位点残基, 当无序蛋白氨基酸残基的任一原子与其相互作用蛋白质的任一原子间距小于 5Å 时, 认为这个氨基酸残基就是 IDPs 与其他蛋白质相互作用时的作用位点残基. IDPs 无序区域与有序区域的划分, 依据 disprot 数据库中的标注.

## 2 结果与讨论

### 2.1 IDPs-蛋白质作用位点残基数据集的特征

经过去冗余处理后得到的 109 条 IDPs 中, 共包含 50 031 个氨基酸残基, 其中属于 IDPs 无序区域的残基有 11 285 个(占 22.6%), 属于有序区域的残基有 38 746 个(占 77.4%). 根据作用位点残基的定义, 这些氨基酸残基形成 IDPs-蛋白质作用位点残基 4 822 个, 其中, 落在 IDPs 无序区域的有 1 241 个, 落在有序区域的有 3 581 个. 每种残基形成 IDPs-蛋白质作用位点残基的情况如图 1 所示, 图中用不同灰度表示作用位点残基位于有序区域或无序区域.



**Fig. 1 Distribution of 20 kinds of amino acids forming IDPs-Protein interaction sites**

Grey represents binding sites located in disordered regions and black represents binding sites located in ordered regions. ■ : Binding sites in ordered region; ■: Binding sites in disordered region.

由图 1 可以看出，20 种氨基酸残基在形成作用位点时，它们在 IDPs 的有序区域出现的频数由小到大依次是：TRP、CYS、HIS、MET、PRO、ASN、TYR、GLN、PHE、GLY、ASP、SER、ALA、ILE、THR、ARG、VAL、LYS、GLU、LEU，在无序区域出现的频数由小到大依次是：CYS、TRP、HIS、TYR、MET、PHE、ASN、GLN、PRO、ILE、GLY、THR、VAL、LYS、ARG、ALA、SER、ASP、GLU、LEU。从总体上看 LEU、GLU、LYS、ARG、VAL 形成作用位点残基时具有较大的频数，而 CYS、TRP、HIS、MET、TYR 形成作用位点残基时的频数较小。从图 1 可知，IDPs-蛋白质的作用位点残基在有序区域出现的频数比较大，并不意味着它们在无序区出现的频数也很大。例如 PRO 与 PHE 这两种氨基酸，虽然 PHE 在 IDPs 作用位点残基的有序区域处出现的频数较 PRO 大，但是 PHE 在无序区出现的频数要小于 PRO。这提示：同一种氨基酸在形成 IDPs-蛋白质作用位点残基时，在有序区域与无序区之间可能具有不同的倾向性，这可能是由 IDPs 无序区与有序区理化性质的差异引起的。

由图 1 还可看出，对于同一种氨基酸而言，落在 IDPs 有序区的频数总是大于落在 IDPs 无序区的频数，但这并不意味着 IDPs 更倾向于通过有序区与其他蛋白质发生相互作用。首先，在所研究体系

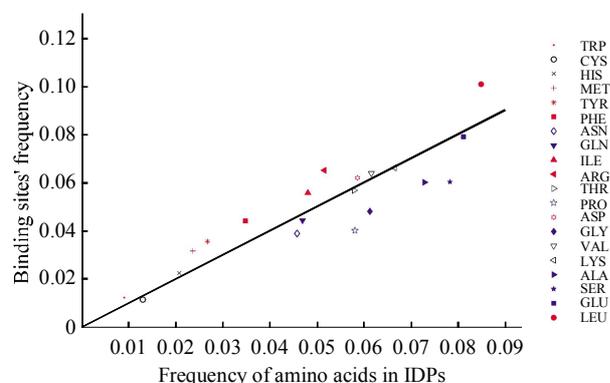
中，IDPs 有序区域的长度大于无序区域的长度，相应氨基酸的含量分别为 77.4%和 22.6%；其次，研究体系中 IDPs 仅通过有序区域或仅通过无序区域与其他蛋白质相互作用的复合物所占比重分别为 50.2%和 20.4%，同时通过有序区和无序区域与其他蛋白质相互作用的复合物占 29.4%。进一步分析发现，那些仅通过有序区域与其他蛋白质分子作用体系的作用位点残基却倾向于落在 IDPs 的有序与无序区域的接点附近，这说明 IDPs 的无序区在 IDPs-蛋白质相互作用中起了重要的作用。

### 2.2 IDPs-蛋白质作用位点残基倾向性

仅根据作用位点位于有序和无序区域的数目来判断氨基酸形成 IDPs-蛋白质相互作用位点残基的倾向性，有一定局限性，这是因为作用位点残基出现的频率会受到相应氨基酸在 IDPs 中出现频率的影响。有些学者<sup>[20]</sup>曾用 BP(binding propensities)来衡量蛋白质-核酸复合物界面氨基酸与核苷酸的偏好性。为了分析不同氨基酸形成作用位点残基的偏好，我们计算了所研究数据集中的每种氨基酸在 IDPs 中出现的几率和在 IDPs-蛋白质复合物作用位点残基集合中出现的几率，如图 2 所示，图中的直线斜率为 1。

图 2 中每一个点代表一种氨基酸残基，横坐标表示氨基酸残基在 IDPs 中出现的几率，纵坐标表示相应氨基酸在作用位点残基集合中出现的几率。

由图 2 可以看出, 20 种氨基酸残基被一条斜率为 1 的直线分成三部分: 位于直线附近、位于直线下方、位于直线上部。



**Fig. 2 The distribution of each amino acid in residue ensemble of IDPs and IDPs-proteins interface sites**

x-Axis represents amino acids' probability in 109 IDPs chains and y-Axis represents amino acids' probability in IDPs-protein interface. Red and blue points denote that amino acids deviate from the straight-line obviously. The slope of straight-line is 1.

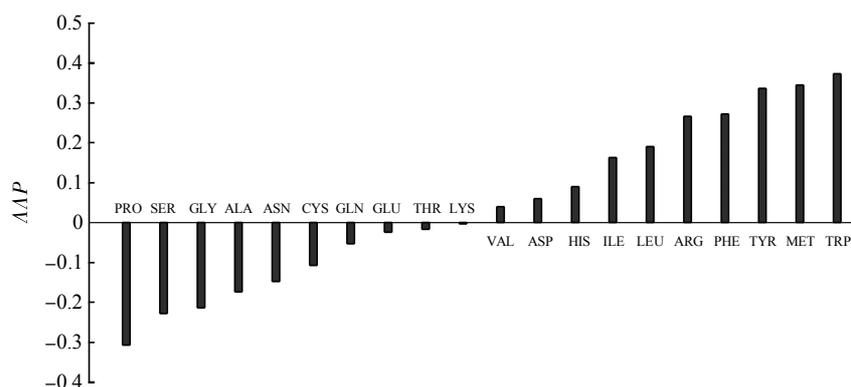
对于落在直线上方的氨基酸而言, 其在 IDPs-蛋白质作用位点集合中出现的几率大于在 IDPs 数

据集中出现的几率. 也就是说, 该部分氨基酸在 IDPs 中出现的几率虽然小, 但是它们在形成 IDPs-蛋白质的作用位点时却可能会有一个较大的几率; 同样的道理, 对于落在直线下方的氨基酸残基, 在 IDPs 中出现的几率较大, 但是它们在 IDPs-蛋白质的作用位点集合中却有一个相对较小的几率. 20 种氨基酸的这些特性表明不同的氨基酸残基在形成 IDPs-蛋白质的作用位点时表现为不同的倾向性. 为了衡量哪些氨基酸倾向于形成 IDPs-蛋白质的作用位点残基, 我们引入了 *AAP* (amino acids propensity) 参数. *AAP* 定义为:

$$AAP(X) = \frac{P_{BS}(X)}{P_{PR}(X)} - 1$$

其中,  $P_{PR}(X)$  表示氨基酸  $X$  在 109 条 IDPs 链中出现的几率,  $P_{BS}(X)$  表示氨基酸  $X$  在作用位点集合中出现的几率.

根据 *AAP* 的大小, 可以判断氨基酸形成 IDPs-蛋白质作用位点残基的偏好性.  $AAP > 0$ , 对应于图 2 中位于单位斜率直线上方的氨基酸, 该部分氨基酸具有形成 IDPs-蛋白质作用位点残基的倾向性; 相反,  $AAP < 0$ , 对应于图 2 中位于单位斜率直线下方的氨基酸, 该部分氨基酸不具有形成作用位点残基的倾向性. 我们计算了 20 种氨基酸的 *AAP* 值, 如图 3 所示.



**Fig. 3 The AAP values of twenty kinds of amino acids in IDPs**

根据 *AAP* 值的大小, 我们可以将 20 种氨基酸分成 3 类:  $AAP > 0.1$  的氨基酸, 定义为倾向型氨基酸;  $AAP < -0.1$  的氨基酸, 称之为非倾向型氨基酸;  $-0.1 < AAP < 0.1$  的氨基酸称之为中间型氨基酸. 在图 3 中可以看出, 倾向型氨基酸主要有: ILE、LEU、ARG、PHE、TYR、MET、TRP; 非

倾向型氨基酸主要有: PRO、SER、GLY、ALA、ASN、CYS; 中间型氨基酸有: GLN、GLU、THR、LYS、VAL、ASP、HIS. Fuxreiter 等<sup>[20]</sup>研究发现, 无序蛋白的某些区域倾向于形成  $\alpha$ -helix, 这些局部的二级结构常参与蛋白质-蛋白质的相互作用. 而 PRO 是  $\alpha$ -helix 的最强终止者, 所以说

PRO 不易出现在蛋白质 - 蛋白质相互作用的界面处. 这与图 3 中 PRO 具有最小的 AAP 值最不容易形成 IDPs- 蛋白质的作用位点残基相一致.

### 2.3 氨基酸在 IDPs 的不同区域特性

同一种氨基酸在形成 IDPs- 蛋白质作用位点残基时, 在 IDPs 的有序与无序区域可能具有不同倾向性. 为了分析它们的差异性, 我们将 109 条 IDPs 蛋白质链残基集合分成两部分: 有序区域残基子集  $S_{OPR}$  和无序区域残基子集  $S_{DPR}$ , IDPs- 蛋白质作用位点残基集合也分为两部分: 落在有序区域的残基子集  $S_{OBS}$  和落在无序区域的残基子集  $S_{DBS}$ .

由此计算出每种氨基酸在不同区域的 AAP 值, 即无序区域残基的 AAP 值(disAAP)和有序区域残基的 AAP(orderAAP). 计算公式如下所示:

$$\text{orderAAP}(X) = \frac{P_{OBS}(X)}{P_{OPR}(X)} - 1$$

$$\text{disAAP}(X) = \frac{P_{DBS}(X)}{P_{DPR}(X)} - 1$$

式中  $P_{OBS}(X)$  和  $P_{OPR}(X)$  分别是氨基酸  $X$  在子集  $S_{OBS}$  和  $S_{OPR}$  中出现的几率,  $P_{DBS}(X)$  和  $P_{DPR}(X)$  分别是氨基酸  $X$  在子集  $S_{DBS}$  和  $S_{DPR}$  中出现的几率. 各氨基酸在不同区域的 AAP 值如图 4a 所示.

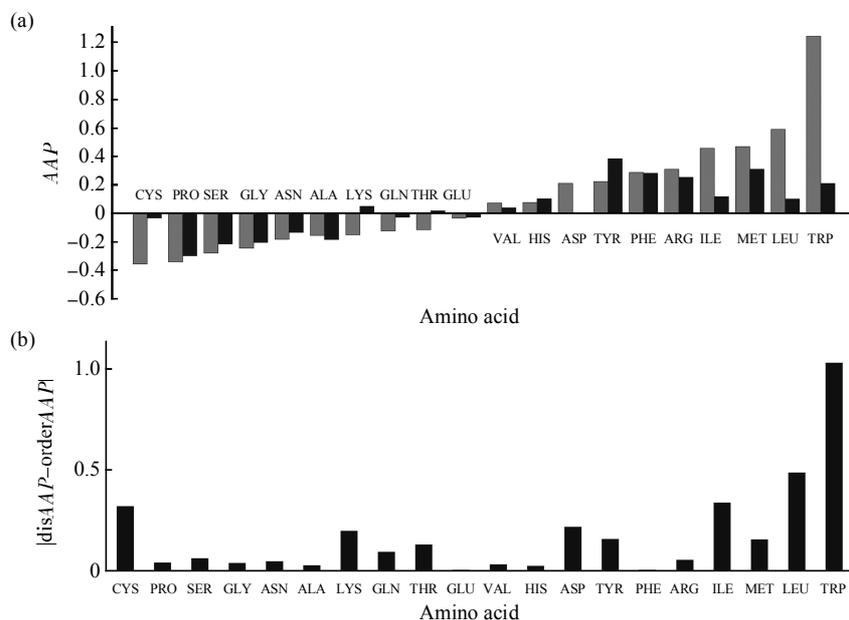


Fig. 4 Compared with the 20 kinds of amino acids' AAP values in IDPs' different regions

(a) AAP values of amino acids in different regions of IDPs. Grey indicates disordered regions and black indicates ordered regions. ■: disAAP; ■: orderAAP. (b) The absolute value |disAAP-orderAAP| of AAP in IDPs different regions.

图 4a 反映了氨基酸在 IDPs 有序与无序区域 AAP 值的差异性. 从中可以看出, 对于 IDPs 的有序区域而言 20 种氨基酸的 AAP 值由小到大依次是: PRO、SER、GLY、ALA、ASN、CYS、GLN、GLU、ASP、THR、VAL、LYS、LEU、HIS、ILE、TRP、ARG、PHE、MET、TYR, 其中 HIS、ILE、TRP、ARG、PHE、MET、TYR 在 IDPs 的有序区比较容易形成作用位点残基, 而 PRO、SER、GLY、ALA、ASN 不倾向于形成作用位点残基; 无序区域的 AAP 值由小到大依次是

GLN、THR、GLU、VAL、HIS、ASP、TYR、PHE、ARG、ILE、MET、LEU、TRP. 对于无序区域而言 ASP、TYR、PHE、ARG、ILE、MET、LEU、TRP 容易形成作用位点残基, CYS、PRO、SER、GLY、ASN、ALA、LYS、GLN、THR 则不倾向于形成作用位点残基; 无序区域的 CYS、GLN、PRO、SER 比在有序区域更不利于形成作用位点残基, ASP、LEU、ILE、TRP 则比有序区域更利于形成 IDPs- 蛋白质的作用位点残基, 而有序区的 TYR、HIS、THR、LYS 比无序区的相应氨基酸更加有利于形成 IDPs- 蛋白质的作用位点残基.

为了进一步衡量同一种氨基酸在有序区域与无序区域形成 IDPs- 蛋白质复合物作用位点倾向性的差别, 我们计算了同一种氨基酸在 IDPs 有序与无序区域的  $AAP$  之差, 即  $|\text{disAAP} - \text{orderAAP}|$ , 如图 4b 所示. 这个差值越大说明该氨基酸在 IDPs- 蛋白质作用位点的不同区域的倾向性差别越明显, 反之则越小. 由图 4b 可以看出, 同一种氨基酸在 IDPs- 蛋白质不同区域形成作用位点残基的倾向性不同, 其中 TRP、LEU、ILE、CYS 的差别比较明显; GLU、PHE、HIS、ALA 基本没有多大差别. 不同氨基酸形成作用位点残基的这种差异性可能是由不同氨基酸侧链的理化性质以及它所处的序列环境决定的.

有综述文章指出<sup>[27]</sup>, IDPs 的无序区域序列与有序区域序列在平均电荷与平均疏水性等多种理化性质方面存在差异, 这种差异导致有序与无序区域氨基酸分布的不对称性, 从而使 IDPs 有序与无序区域之间存在序列环境上的差异, 这可能是影响相应

区域氨基酸形成 IDPs- 蛋白质作用位点倾向性的原因之一.

#### 2.4 氨基酸的理化性质对 IDPs-蛋白质作用位点残基的影响

有些文献中提到<sup>[28-29]</sup>氨基酸残基的序列保守性、疏水程度、所处的二级结构、溶剂可及性表面积及侧链构象熵对作用位点残基的影响较大, 除此之外, 我们猜测氨基酸的侧链体积和极性等理化性质对作用位点的残基倾向性可能有影响. 基于以上原因, 我们选取了氨基酸的 6 种主要理化性质, 来研究其对 IDPs- 蛋白质作用位点残基倾向性的影响, 这 6 种理化性质是: 疏水性(hydrophobicity, H)、极性(polarity, PI)、极化率(polarizability, P2)、侧链体积(side chain's volume, VSC)、侧链的净电荷指数(net charge index of side chains, NCISC)和溶剂可及表面积(solvent accessible surface area, SASA). 每种氨基酸所对应的 6 种理化参数如表 1 所示<sup>[30]</sup>.

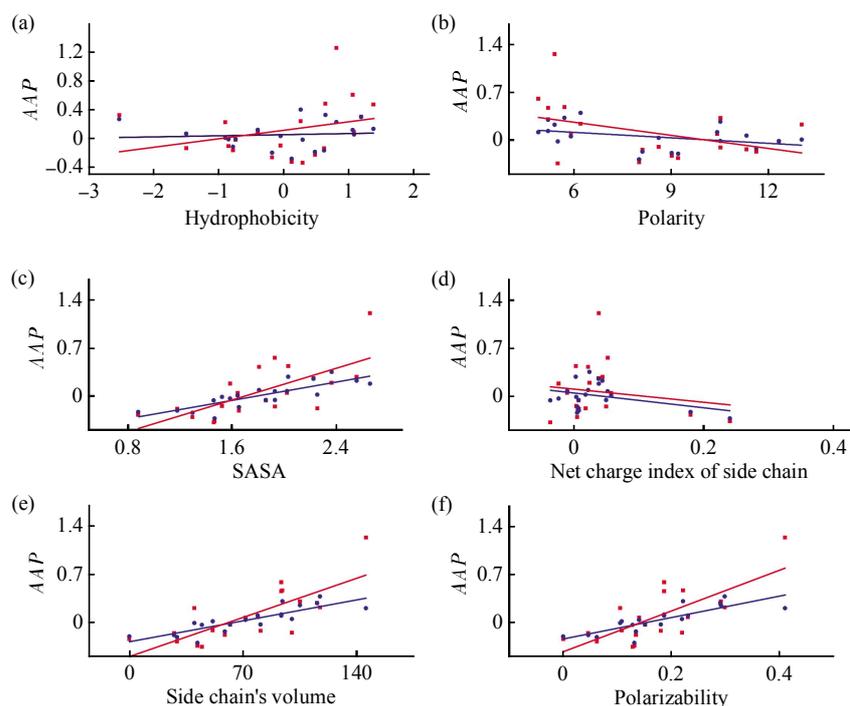
**Table 1 Six major physical and chemical properties of the 20 kinds of amino acids<sup>[30]</sup>**

Amino acid	H	VSC	PI	P2	SASA	NCISC
ALA	0.62	27.5	8.1	0.046	1.181	0.007187
ARG	-2.53	105	10.5	0.291	2.56	0.043587
ASN	-0.78	58.7	11.6	0.134	1.655	0.005392
ASP	-0.9	40	13	0.105	1.587	-0.02382
CYS	0.29	44.6	5.5	0.128	1.461	-0.03661
GLN	-0.85	80.7	10.5	0.18	1.932	0.049211
GLU	-0.74	62	12.3	0.151	1.862	0.006802
GLY	0.48	0	9	0	0.881	0.179052
HIS	-0.4	79	10.4	0.23	2.025	-0.01069
ILE	1.38	93.5	5.2	0.186	1.81	0.021631
LEU	1.06	93.5	4.9	0.186	1.931	0.051672
LYS	-1.5	100	11.3	0.219	2.258	0.017708
MET	0.64	94.1	5.7	0.221	2.034	0.002683
PHE	1.19	115.5	5.2	0.29	2.228	0.037552
PRO	0.12	41.9	8	0.131	1.468	0.239531
SER	-0.18	29.3	9.2	0.062	1.298	0.004627
THR	-0.05	51.3	8.6	0.108	1.525	0.003352
TRP	0.81	145.5	5.4	0.409	2.663	0.037977
TYR	0.26	117.3	6.2	0.298	2.368	0.023599
VAL	1.08	71.5	5.9	0.14	1.645	0.057004

H: Hydrophobicity; VSC: Side chain's volume; PI: Polarity; P2: Polarizability; SASA: Solvent accessible surface area; NCISC: Net charge index of side chain.

在上述的研究中定义了  $AAP$ , 在这里我们仍然使用上述计算出的  $AAP$  值. 这里的  $AAP$  值主要包括: IDPs 无序区域氨基酸的  $AAP$  值、IDPs 有序

区域氨基酸的  $AAP$  值. 氨基酸的  $AAP$  值与各种参数的关系如图 5 所示.



**Fig. 5 The effects of physical and chemical properties on IDP-Protein interface residues**

The  $x$ -axis is (a) Hydrophobicity. (b) Polarity. (c) Solvent accessible surface area. (d) Net charge index of side chain. (e) Side chain's volume. (f) Polarizability. The  $y$ -axis is the AAP value of amino acid. Red circle: amino acids in disordered region; Blue triangle: amino acids in ordered region. The AAP values are fitted through line. ■: Amino acids in disordered region; ●: Amino acids in ordered region; —: Linear Fit of disAAP; —: Linear Fit of orderAAP.

由氨基酸疏水性与 AAP 的关系(图 5a)发现,无论是有序区域的氨基酸还是无序区域的氨基酸,具有较大 AAP 的氨基酸多数落在疏水性指数较大的区域,当然也有个别较大 AAP 的氨基酸落在疏水性指数较小的区域,多数具有中等疏水性的氨基酸的 AAP 值比较小,这说明极端的疏水性环境有利于氨基酸形成 IDPs-蛋白质的作用位点。图 5 显示,溶剂可及性表面(图 5c)和侧链体积(图 5e)、氨基酸的极化率(图 5f)有一定相关性。大部分倾向型氨基酸(具有较大的 AAP 值)一般具有较大溶剂可及性表面、较大侧链体积和较大的极化率;而非倾向型氨基酸(具有较小的 AAP 值)一般具有较大的极性、较小溶剂可及性表面和较小侧链体积。大部分的倾向型氨基酸分布在侧链净电荷指数小的部分(图 5d),这可能是由于侧链的带电量较少,有利于 IDPs 的折叠,进而形成 IDPs-蛋白质界面处比较规则的结构。

对比图 5 的 6 幅图可以看出,对处于相同理化

性质下的 IDPs 不同区域的同一种氨基酸来说,它们形成 IDPs-蛋白质作用位点时的 AAP 值随理化性质具有类似的分布,唯一不同的是不同区域的相同氨基酸对主要理化性质的敏感程度有区别。

以上的研究表明:疏水性指数较大、侧链净电荷量指数较小、极性较小、溶剂可及性表面积较大、侧链体积较大、极化率较大的氨基酸比较倾向于形成 IDPs-蛋白质的作用位点残基。

通过计算标准化后的理化参数构建的协方差矩阵,利用 MATLAB 得到 3 个最大的主成分(principal components, PC)对应协方差矩阵的特征值分别为 3.086、1.875 和 0.887,分别记为 PC1、PC2 和 PC3。PC1、PC2 和 PC3 可以反映出原来 6 种主要理化性质大约 97%的信息,6 种理化参数对 3 个最大主成分的载荷如表 2 所示。从表 2 可以看出,残基的侧链体积、溶剂可及性表面积及极化率对 PC1 影响较大,疏水性以及极性对 PC2 影响较大,侧链的净电荷量对 PC3 的影响尤为明显。

**Table 2 The loading of the different parameters on the top three principal components**

Parameter	PC1	PC2	PC3
H	0.0233	0.6964	-0.2069
NCISC	-0.1585	0.2487	0.9531
VSC	0.5637	0.0069	0.0552
SASA	0.5465	-0.1816	0.1221
PI	-0.2172	-0.6470	0.1013
P2	0.5574	-0.0394	0.1436

### 3 结论及展望

我们对固有无序蛋白与其他蛋白质相互作用时的作用位点残基进行了研究, 发现不同的氨基酸形成 IDPs- 蛋白质作用位点残基的倾向性有一定的差别. 为了比较氨基酸形成表面残基的倾向性, 定义并引入了相异于 BP 的参数 AAP. 根据 IDPs- 蛋白质作用位点残基的 AAP 值, 20 种氨基酸可分成 3 类: 倾向型氨基酸主要有: ILE、LEU、ARG、PHE、TYR、MET、TRP; 非倾向型氨基酸主要有: PRO、SER、GLY、ALA、ASN、CYS; 中间型氨基酸有: GLN、GLU、THR、LYS、VAL、ASP、HIS.

考虑到 IDPs 无序区域与有序区域的差异, 该研究分别计算了有序区域与无序区域的 AAP 值, 计算结果表明: HIS、ILE、TRP、ARG、PHE、MET、TYR 在 IDPs 的有序区比较容易形成作用位点, 而 ASP、TYR、PHE、ARG、ILE、MET、LEU、TRP 在无序区容易形成 IDPs- 蛋白质的作用位点残基. 进一步考察 IDPs 不同区域的作用位点残基, 发现同一种氨基酸在不同区域形成 IDPs- 蛋白质作用位点残基的倾向性不同. 作用位点残基在 IDPs 的不同区域差异性比较明显的是 TRP、LEU、ILE、CYS, 然而, GLU、PHE、HIS、ALA 基本没有多大差别. 除此之外, 研究也进一步表明 IDPs- 蛋白质的作用位点残基一般具有以下特征: 疏水能力较强、侧链净电荷量较少、极性较小、溶剂可及性表面较大、侧链体积较大、极化率较大.

进一步对 6 种理化性质进行主成分分析发现, 残基的侧链体积、溶剂可及性表面积及极化率对第一主成分影响较大且贡献相似, 疏水性以及极性对第二主成分影响较大, 第三主成分主要由侧链的净电荷量决定.

该研究将揭示 IDPs- 蛋白质作用位点残基的更多信息, 对于进行合理的药物设计, 分析蛋白质分子相互作用的关键残基具有重要的意义, 除此之外, 我们的研究也将为 IDPs- 蛋白质作用位点残基的预测提供更多的输入信息, 进一步提高预测的精度, 为发展针对 IDPs- 蛋白质作用位点残基的预测软件奠定一些基础性工作.

### 参 考 文 献

- [1] 黄永棋, 刘志荣. 天然无序蛋白质: 序列 - 结构 - 功能的新关系. 物理化学学报, 2010, **26**(8): 2061-2072  
Huang Y Q, Liu Z R. Acta Phys Chim Sin, 2010, **26**(8): 2061-2072
- [2] 王克夷. 天然无折叠蛋白质. 生命的化学, 2006, **26**(3): 199-202  
Wang K Y. Chemistry of Life, 2006, **26**(3): 199-202
- [3] Dunker A K, Obradovic Z, Romero P, *et al.* Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform, 2000(11): 161-171
- [4] Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci, 2002, **27**(10): 527-533
- [5] Jurneczko E, Cruickshank F, Porrini M, *et al.* Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry. Biochem Soc Trans, 2012, **40**(5): 1021-1026
- [6] Uversky V N, Dunker A K. Intrinsically Disordered Protein Analysis. New York: Springer New York, 2012, 429-437
- [7] Kovacs D, Szabo B, Pancsa R, *et al.* Intrinsically disordered proteins undergo and assist folding transitions in the proteome. Arch Biochem Biophys, 2013, **531**(1-2): 80-89
- [8] Sotomayor-Perez A C, Karst J C, Ladant D, *et al.* Mean net charge of intrinsically disordered proteins: experimental determination of protein valence by electrophoretic mobility measurements. Methods Mol Biol, 2012, **896**: 331-349
- [9] Yoon M K, Mitrea D M, Ou L, *et al.* Cell cycle regulation by the intrinsically disordered proteins p21 and p27. Biochem Soc Trans, 2012, **40**(5): 981-988
- [10] Williamson M P, Potts J R. Intrinsically disordered proteins: administration not executive. Biochem Soc Trans, 2012, **40**(5): 945-949
- [11] Dogan J, Schmidt T, Mu X, *et al.* Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. J Biol Chem, 2012, **287**(41): 34316-343124
- [12] Sigalov A B. Protein intrinsic disorder and oligomericity in cell signaling. Mol Biosyst, 2010, **6**(3): 451-461
- [13] Staneva I, Huang Y, Liu Z, *et al.* Binding of two intrinsically disordered peptides to a multi-specific protein: a combined Monte Carlo and molecular dynamics study. PLoS Comput Biol, 2012, **8**(9): e1002682
- [14] Mészáros B, Simon I, Dosztányi Z. The expanding view of protein-protein interactions: complexes involving intrinsically disordered proteins. Phys Biol, 2011, **8**(3): 035003
- [15] Uversky V N. Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta, 2013, **1837**(5):932-951

- [16] Uversky V N. Intrinsic disorder-based protein interactions and their modulators. *Curr Pharm Des*, 2013, **19**(23): 4191-4213
- [17] 曹赞霞, 张红梅, 赵岩, 等. 固有无序蛋白质无序与有序接点处的氨基酸序列分析. *生物物理学报*, 2011, **27**(9): 801-811  
Cao Z X, Zhang H M, Zhao Y, *et al.* *Acta Biophys Sin*, 2011, **27**(9): 801-811
- [18] Wojtas M, Kaplon T M, Dobryszycski P, *et al.* The effect of counterions on the conformation of intrinsically disordered proteins studied by size-exclusion chromatography. *Methods Mol Biol*, 2012, **896**: 319-330
- [19] Smith W W, Schreck C F, Hashem N, *et al.* Molecular simulations of the fluctuating conformational dynamics of intrinsically disordered proteins. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2012, **86**(4 Pt 1): 041910
- [20] Fuxreiter M, Simon I, Friedrich P, *et al.* Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol*, 2004, **338**(5): 1015-1026
- [21] Tong J, Jiang P, Lu Z. RISP: a web-based server for prediction of RNA-binding sites in proteins. *Comput Meth Prog Bio*, 2008, **90**(2): 148-153
- [22] Wang L, Brown S J. BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res*, 2006, **34**(Suppl 2): 243-248
- [23] Jones S, Thornton J M. Prediction of protein-protein interaction sites using patch analysis. *J Mol Biol*, 1997, **272**(1): 133-143
- [24] Ezkurdia I, Bartoli L, Fariselli P, *et al.* Progress and challenges in predicting protein-protein interaction sites. *Brief Bioinform*, 2009, **10**(3): 233-246.
- [25] Mihel J, Šikic M, Tomic S, *et al.* PSAIA-protein structure and interaction analyzer. *BMC Struct Biol*, 2008, **8**(1)
- [26] 吴建盛, 胡栋, 伍洪涛, 等. 蛋白质-核酸复合物界面氨基酸与核苷酸偏好性分析. *生物物理学报*, 2010, **26**(3): 234-244  
Wu J S, Hu D, Wu H T, *et al.* *Acta Biophys Sin*, 2010, **26**(3): 234-244
- [27] Uversky V N. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol*, 2011, **43**(8): 1090-1103
- [28] 罗慧萍. 蛋白质-蛋白质相互作用界面和热点预测的方法研究[D]. 武汉: 武汉科技大学, 2011  
Luo H P. Approach to Prediction of Protein-protein Interaction Interface and Hot Spot[D]. Wuhan: Wuhan University of Science and Technology, 2011
- [29] Zhou H X, Qin S. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics*, 2007, **23**(17): 2203-2209
- [30] 夏俊峰. 蛋白质相互作用及其结合面热点残基的预测方法研究[D]. 安徽: 中国科学技术大学, 2010  
Xia J F. Computational Prediction of Protein-protein Interactions and Hot Spot Residues in Protein Interfaces[D]. Anhui: University of Science and Technology of China, 2010

## An Investigation on Characteristic of Residues Involved in Intrinsically Disordered Protein-Protein Interaction\*

DONG Chuan<sup>1)</sup>, CAO Zan-Xia<sup>1,2)</sup>, ZHAO Li-Ling<sup>1,2)</sup>, SUO Zhen-Peng<sup>1)</sup>, WANG Ji-Hua<sup>1,2)\*\*</sup>

<sup>1)</sup> *The College of Physics and Electronic Information of Dezhou University, Dezhou 253023, China;*

<sup>2)</sup> *Key Laboratory of Functional Macromolecular Biophysics of Shandong (Dezhou University), Dezhou 253023, China)*

**Abstract** In this study, we investigated the characters of residues in IDPs that involved in the interactions with other proteins. Firstly, 109 IDPs' chains and 299 IDPs-protein compounds which meet the requirement in database were selected; secondly, 4 822 interface amino acid residues which involving in interaction were extracted from the totally 50 031 amino acid residues in the 109 IDPs chains. The results indicated that the 20 amino acids have different propensities when forming IDPs-protein's interfaces. Therefore, we divided the 20 amino acids into three parts based on their propensities: propensity amino acids (ILE, LEU, ARG, PHE, TYR, MET, TRP), middle amino acids (GLN, GLU, THR, LYS, VAL, ASP, HIS) and non-propensity amino acids (PRO, SER, GLY, ALA, ASN, CYS). Moreover, the results show that the 20 amino acids have different propensities in IDPs' different regions (ordered or disordered regions). For example, TRP, LEU, ILE and CYS are more frequently observed compared with ordered and disordered regions. However, there are no obvious difference for GLU, PHE, HIS and ALA residues between ordered regions and disordered regions. Besides, the amino acids which have larger hydrophobicity, polarizability, side chain's volume, solvent accessible surface area, smaller polarity and net charge index of side chain tend to be IDPs-Protein's interface. The results obtained by principal components analysis showed that the polarizability, side chain's volume and solvent accessible surface area of residues had more affect on the IDPs-Protein's interaction residues.

**Key words** intrinsically disordered protein (IDPs), residue characteristics of action sites, ordered regions, disordered regions, physical-chemical properties of amino acid

**DOI:** 10.3724/SP.J.1206.2013.00214

---

\*This work was supported by grants from The National Natural Science Foundation of China (31000324, 61271378, 30970561) and Natural Science Foundation of Shandong Province, China (ZR2011FL011, ZR2012CL09).

\*\*Corresponding author.

Tel: 86-534-8985933, E-mail: jhw25336@126.com

Received: May 17, 2013 Accepted: August 14, 2013