

## Discovery Studio 2.0 的模块扩展及应用： 残基水平非键相互作用能量的自动批量计算

高跃东<sup>1,2</sup>, 黄京飞<sup>1,3,\*</sup>

(1. 中国科学院昆明动物研究所 遗传资源与进化国家重点实验室, 云南 昆明 650223 2. 中国科学院研究生院, 北京 100049;  
3. 中国科学院昆明动物研究所 - 香港中文大学生物资源与疾病分子机理联合实验室, 云南 昆明 650223)

**摘要:** 非键相互作用对于生物体系中的分子识别和结合过程起着关键作用。然而, 传统的方法并不能在残基水平自动批量计算非键相互作用。近年来, 已经发展了一些方法和工具进行非键相互作用的计算分析。该文研究发展了一种可以自动计算残基间非键相互作用的方法, 即用 Perl 脚本调用 Discovery Studio 2.0 (DS2.0, Accelrys Inc.) 底层模块中的非键相互作用协议, 实现了直接利用命令行批量计算非键相互作用能量, 而无需通过 DS2.0 的图形界面。该方法扩展了 DS2.0 的计算模块, 并于近期运用到了复合结构的研究分析中。

**关键词:** 非键相互作用; 模块扩展; Discovery Studio 2.0  
中图分类号: Q5-3; Q811.4; TP399 文献标志码: A 文章编号: 0254-5853-(2011)03-0262-05

## An extension strategy of Discovery Studio 2.0 for non-bonded interaction energy automatic calculation at the residue level

GAO Yue-Dong<sup>1,2</sup>, HUANG Jing-Fei<sup>1,3,\*</sup>

(1. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650223, China;  
2. Graduate School of Chinese Academy of Sciences, Beijing 100039, China; 3. Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, the Chinese Academy of Sciences, Chinese University of Hong Kong, Kunming 650223, China)

**Abstract:** Non-bonded interaction forces play crucial roles in molecular recognition and binding in biological systems. However, it is difficult for traditional methods to automatically calculate and batch the non-bonded energy at the residue level. In recent years, many studies have focused on non-bonded interactions and developed tools to calculate and analyze such interactions. In this study, we present a highly automated approach for the calculation of non-bonded energy. Our strategy invoked protocols relevant to non-bonded interactions within Discovery Studio 2.0 (DS2.0, Accelrys Inc.) bottom module using Perl script, and determined the direct command line operation of calculating non-bonded interaction energy batches without accessing the graphical interface of DS. This approach extended the DS2.0 module and was applied to a recent study of complex structure analysis.

**Key words:** Non-bonded energy; Protocol extension; Discovery Studio 2.0

在生物分子体系中, 除了共价键外还存在许多非键相互作用, 涉及生命活动的许多方面。非键相互作用在生物分子结构的稳定性方面起着重要作用, 例如在蛋白与 DNA、蛋白与 RNA 分子复合物体系中, 非键相互作用是维持复合物结构稳定性的主要因素。另外, 它们在生物分子识别、离子载体的选择性、分子组装、分子簇形成等方面也都起着

关键作用(McCullagh et al, 2008)。在生物过程中, 信息的传递不仅需要分子间能够识别并形成复合物体系, 有时也要求复合物能容易解离。非键相互作用不形成共价化学键的特性, 使得其能够在生命过程中发挥关键作用。深入的研究和理解非键相互作用机制是了解复合物分子功能的重要方面(Waters, 2002)。

收稿日期: 2010-08-05; 接受日期: 2011-01-18

基金项目: 国家重点基础研究发展计划项目“973”(2009CB941302); 国家自然科学基金项目(30470939、30623007); 中国科学院基金项目(2007211311091)

\*通讯作者(Corresponding author), Tel: +86 871 5195183, E-mail: huangjf@mail.kiz.ac.cn

非键相互作用可以分为分子内相互作用和分子间相互作用。分子内相互作用决定了分子结构的形成及其构象(conformations)变化;而分子间相互作用则在许多方面起着重要作用,如药物小分子与靶蛋白的识别和结合(Pyrkov et al, 2010)、结构化学(Ashenhurst, 2010)和材料分子研究(Nakashima et al, 2008),以及 RNA 分子体系研究(Spriggs et al, 2008)等。因此,分子间相互作用更为人们所广泛关注。

对于生物分子复合物体系,其结构和功能方面的研究不可避免地需要进行非键相互作用的计算和评估。一般来说,在经验力场中,对非键相互作用描述的重要性等同于对共价键性质的描述(Brooks et al, 1983; Weiner & Kollman, 1981),前者包含静电力(coulombic)和范德华相互作用(Sagui & Darden, 1999);后者包含键长、键角和二面角。在分子动力学模拟过程中,非键相互作用的计算和评估非常耗费计算机时。尽管已有大量的研究涉及对非键相互作用的计算和分析(Sundaram & Prasad, 1982; Van Der Spoel et al, 2005),但传统的计算分析工具无法在残基水平上成对批量自动计算非键相互作用能量。

Discovery Studio 2.0(DS2.0, Accelrys Inc.)计算分析引擎提供了相对完善的分子建模和模拟环境,目前已被广泛应用于医药、生物技术、化学和化工等领域的科学研究中(Hirashima & Huang, 2008; Spassov & Yan, 2008)。在该软件包中,计算非键相互作用需在图形界面下完成,无法实现残基对之间非键相互作用的批量自动计算。本文对 DS2.0 的非键相互作用的计算模块和协议(protocol)进行了扩展。

## 1 方法

### 1.1 程序概述

程序设计的目的是通过扩展开发,使程序模块能在残基水平自动计算非键相互作用。总体策略是通过编辑 DS2.0 原有的底层组件(components),使其具有接收外部程序传递参数信息的接口,同时接合 Perl (Stajich et al, 2002)语言脚本进行相应的流程控制,以达到批量计算的功能。输入文件为 PDB 格式文件以及包括具体参数信息的 ASCII 文本文件,通过命令行调用改进后的非键相互作用计算模块,程序将自动计算指定残基间的非键相互作用。以上操作过程不需要用户打开 DS2.0 程序客户端窗口而实现了后台计算。

### 1.2 程序结构

程序分两个结构单元,即能够接受输入参数的非键相互作用计算模块和 Perl 脚本控制程序。两个结构单元在程序底层通过管道控制模块服务(pipeline pilot server, Accelrys Inc.)进行通讯,以达到协同计算的目的。

在这里,能够接受输入参数的非键计算模块是在 DS2.0 模块“Calculate Interaction Energy protocol”的基础上扩展开发而来的,并被命名为“AutoNBEC”(auto non-bonded energy calculator protocol)。参数传递由管道控制模块 Pipeline Pilot v.7.5 (Accelrys Inc)完成。通过重新编辑相关组件(components),将修改过的计算模块“AutoNBEC”导入到 DS2.0 环境中。AutoNBEC 主要根据用户指定的残基,通过读取 PDB 文件中残基的结构信息,利用相关力场所定义的能量函数进行非键相互作用(静电作用和范德华作用)能量计算。DS2.0 计算引擎可调用多种力场计算,在此,我们调用了经典的 CHARMM 力场进行计算。在 DS2.0 的计算模块中,对非键相互作用的计算截断(cut off)值采用默认值 14 Å。

Perl 脚本利用 DS2.0 的应用程序编程接口(application programming interface, API)调用 DS2.0 的相应计算模块。API 基于面向对象程序设计(object-oriented programming, OPP)策略,封装了相应的计算及操作模块。Perl 脚本的优势在于它能够直接调用 DS2.0 的底层模块。因此,可避免打开 DS2.0 的图形窗口,达到了通过命令行运行批量计算任务的目的。

### 1.3 程序流程

程序的工作流程包括主要的六个阶段:(1)对用户输入文件内容进行解析;(2)对计算环境进行初始化并对计算分子分配 CHARMM 力场参数;(3)产生所有原子的索引;(4)利用哈希(Hash)模型构建需要计算的残基对的数组列表;(5)把需要计算的相关参数传递给 AutoNBEC 模块;(6)解析计算结果并输出结果文件。图 1 显示了 AutoNBEC protocol 的计算流程细节。

## 2 结果

### 2.1 输入文件

用户以 ASCII 文本文件形式指定 PDB 文件中需要计算的链编号信息。文件包括三列内容,以 tab 键分隔,第一列是需要计算的 PDB 代码(PDB

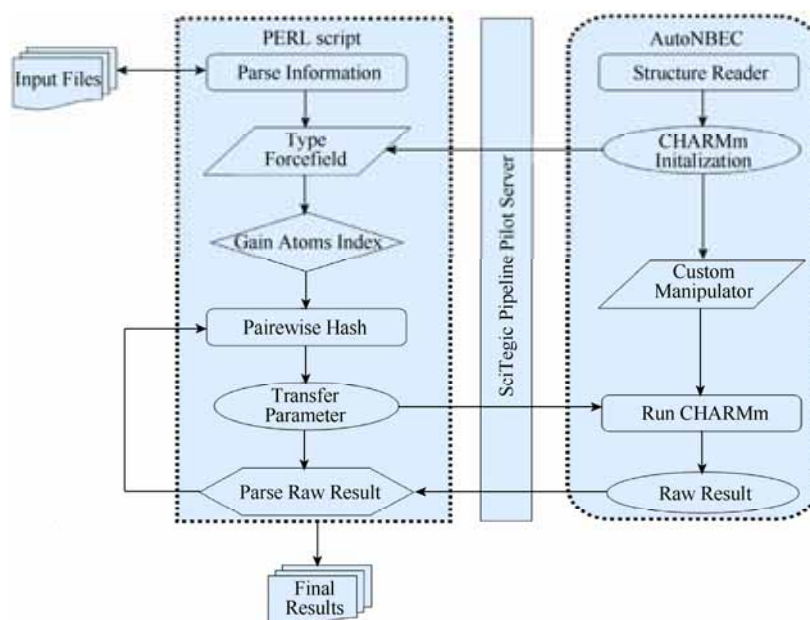


图1 AutoNBEC 非键相互作用计算模块流程

Fig. 1 Schematic workflow of the AutoNBEC module

code), 其余两列是该 PDB 文件中需要计算链的顺序索引。第一条链索引号为 0, 第二条链索引号为 1, 并以此类推。同时, 用户还需要提供计算所需的 PDB 格式文件。图 2(A)显示了输入文件的格式。

## 2.2 程序的执行

AutoNBEC 是对 DS2.0 非键相互作用计算模块的扩展。因此, 必须在 DS2.0 计算引擎环境下运行。首次运行时需要把 AutoNBEC 导入 DS2.0 的计算引擎环境中。剩余步骤为: 在 Windows 环境下打开 CMD 窗口(linux 环境下则打开 terminal 终端), 随

后以命令行方式运行程序的 Perl 脚本。

另外, 读者可向本文通讯作者(Email: huangjf@mail.kiz.ac.cn, Tel: 86-871-5195183)提出计算要求, 并通过 Email 形式获取计算结果。

## 2.3 程序输出及应用实例

计算结果以文本文件形式输出, 数据之间以 tab 键分隔。每一行包括进行计算的残基名称、编号、所属链、计算时间、非键能量总值、静电作用能量、范德华作用能量等结果信息(图 2, C)。结果文件可通过文本编辑软件或图形显示软件进行深

```

A
2F9d  0    1
2ad9  0    1
.....

B Connected to https://localhost:9943 as user ibn.

Task info
Task name:   ibn/AutoNBECalculator
Host:       https://localhost:9943
Protocol Path: C:\Documents and Settings\ibn\My Documents\AutoNBECalculator_2009_05_13_010829_234

Protocol is running.....done
current number is 1
Protocol complete

Task info
Task name:   ibn/AutoNBECalculator
Host:       https://localhost:9943
Protocol Path: C:\Documents and Settings\ibn\My Documents\AutoNBECalculator_2009_05_13_010909_078

Protocol is running.....done
current number is 2
Protocol complete
.....

C
localhost:9943 file C:/GaoF550pan/rnapdb/2f9d/2F9d.pdb have 5 chains
Tottle chains: 0 1 2 3 4 AA chains: 0 1 2 3 BA chains:
chain 0 (AA Chain) residues interaction with chain 1 (AA chain):
LEU13 Chain1 Residue2 ASB84 Chain2 Residue73 [04/08 17:25:03] Potential Energy = -0.16241 kcal/mol Van der Waals Energy = -0.00076 kcal/mol Electrostatic Energy = -0.16165 kcal/mol
ASB18 Chain1 Residue7 ARG98 Chain2 Residue74 [04/08 17:29:55] Potential Energy = -0.53174 kcal/mol Van der Waals Energy = 0.00000 kcal/mol Electrostatic Energy = -0.53174 kcal/mol
ARG19 Chain1 Residue8 LEU26 Chain2 Residue15 [04/08 17:32:29] Potential Energy = -1.54884 kcal/mol Van der Waals Energy = -0.00001 kcal/mol Electrostatic Energy = -1.54883 kcal/mol
ARG19 Chain1 Residue8 TYR28 Chain2 Residue17 [04/08 17:33:21] Potential Energy = -1.49811 kcal/mol Van der Waals Energy = -0.00018 kcal/mol Electrostatic Energy = -1.49807 kcal/mol
ARG19 Chain1 Residue8 PHE80 Chain2 Residue69 [04/08 17:36:17] Potential Energy = -0.27767 kcal/mol Van der Waals Energy = 0.00000 kcal/mol Electrostatic Energy = -0.27767 kcal/mol
ARG19 Chain1 Residue8 ASB84 Chain2 Residue73 [04/08 17:38:02] Potential Energy = -1.96594 kcal/mol Van der Waals Energy = -0.01143 kcal/mol Electrostatic Energy = -1.95411 kcal/mol
ARG19 Chain1 Residue8 TYR56 Chain2 Residue75 [04/08 17:38:55] Potential Energy = -0.63007 kcal/mol Van der Waals Energy = -0.03604 kcal/mol Electrostatic Energy = -0.59393 kcal/mol
ARG19 Chain1 Residue8 LEU87 Chain2 Residue75 [04/08 17:39:22] Potential Energy = -3.05092 kcal/mol Van der Waals Energy = -0.00239 kcal/mol Electrostatic Energy = -3.04853 kcal/mol
ILE30 Chain1 Residue19 GLU55 Chain2 Residue44 [04/08 17:50:19] Potential Energy = -3.79990 kcal/mol Van der Waals Energy = -0.02457 kcal/mol Electrostatic Energy = -3.79533 kcal/mol
THR31 Chain1 Residue20 THR53 Chain2 Residue42 [04/08 17:52:04] Potential Energy = -0.90606 kcal/mol Van der Waals Energy = -0.00017 kcal/mol Electrostatic Energy = -0.90591 kcal/mol
THR31 Chain1 Residue20 GLU55 Chain2 Residue44 [04/08 17:52:58] Potential Energy = -0.15995 kcal/mol Van der Waals Energy = -0.07637 kcal/mol Electrostatic Energy = -0.06356 kcal/mol
THR31 Chain1 Residue20 THR56 Chain2 Residue45 [04/08 17:53:24] Potential Energy = -0.12563 kcal/mol Van der Waals Energy = 0.00000 kcal/mol Electrostatic Energy = -0.12563 kcal/mol
.....

```

图2 AutoNBEC 非键相互作用计算模块输入输出文件格式及运行状态截图

Fig. 2 Input file, output file and running state of the AutoNBEC module

A: 输入文件, 包括 PDB 编号, 链编号; B: 命令行运行状态截图; C: 结果输出文件格式。

A: Input file contains PDB code and chain order index; B: State of command line running; C: Result output file format.

入后续分析。

AutoNBEC可应用于生物大分子复合物的分析中,如蛋白质-DNA、蛋白质-RNA、蛋白质-蛋白质等复合体系中生物分子间的相互作用分析。我们利用该工具对依赖 $\text{NAD}^+$ 的DNA连接酶与底物DNA的

复合物进行残基水平上的非键相互作用分析(Gao & Huang, 未发表资料)。结果显示,连接酶与DNA的非键相互作用在几个区域有明显的强相互作用(图 3),其中I、V、VI三个区域对应于该类酶先前研究报道过的序列高度保守的功能区

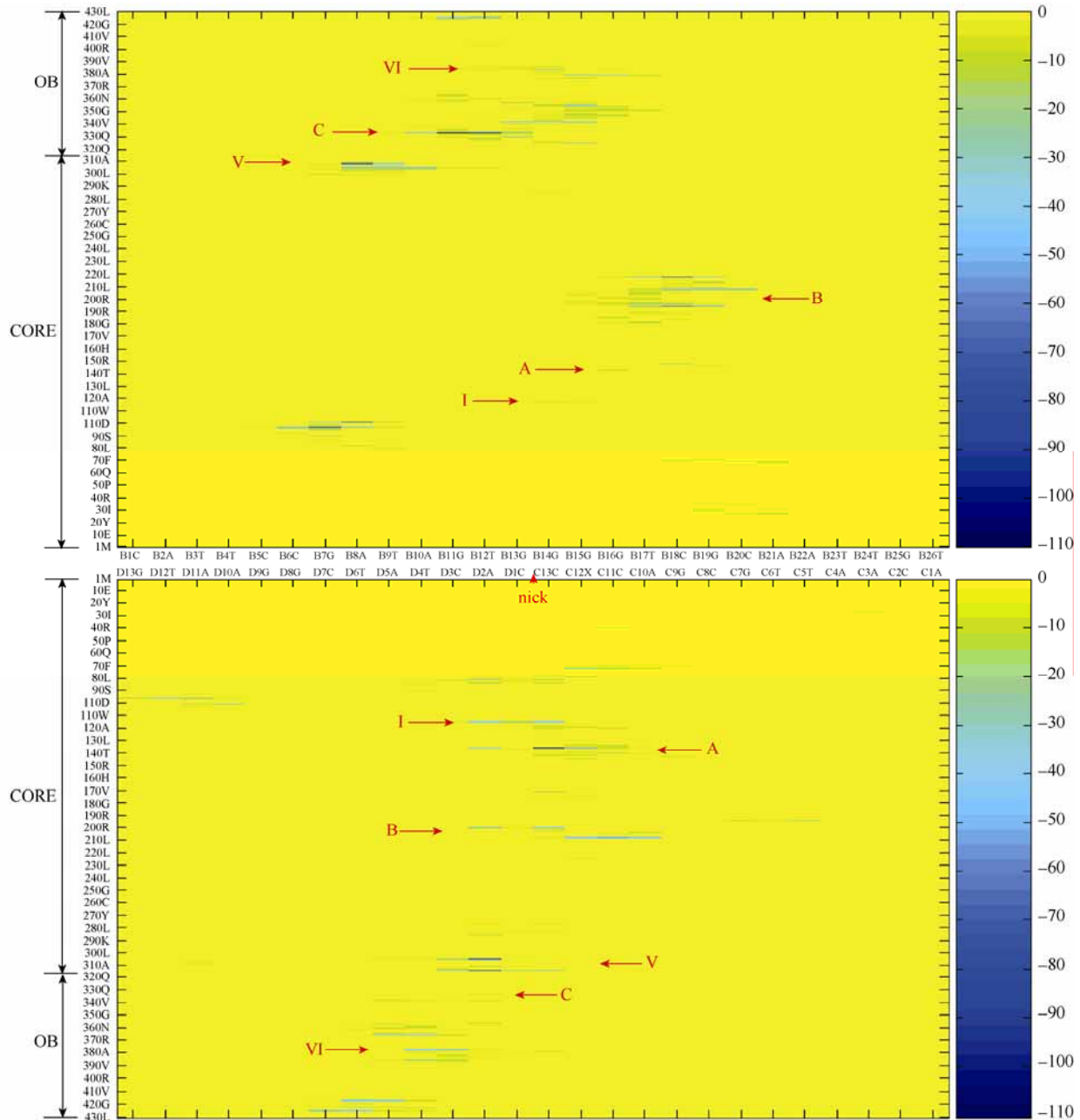


图 3 依赖 $\text{NAD}^+$ 的DNA连接酶(ligase)与DNA复合物(PDB code:2owo)残基水平的非键相互作用 Heatmap 图(adenylation 和 OB 结构域)

Fig. 3 Heatmap representation of pair-wise nonbonded energy values in  $\text{NAD}^+$  dependent DNA ligase complex (adenylation domain and OB domain, PDB code: 2owo)

纵坐标: ligase 的蛋白序列; 横坐标: 与之接合的双链 DNA; 标尺: 残基对间的非键相互作用能量值(单位 kcal/mol)。在显示出很强非键相互作用的区域中, I、V、VI(图中箭头指向的区域)在序列上对应先前研究报道过的高度保守的功能区; A、B、C: 新发现的对应于高度保守序列的区域, 提示功能上的重要性。Vertical axis: protein sequence of ligases; Abscissa: double-stranded DNA bonded to the ligase; The scale is pair-wise nonbonded energy values (kcal/mol). Several regions in  $\text{NAD}^+$  dependent DNA ligase with significant non-bonded interaction have been obtained (indicated by the arrow). Three of these regions have been found in previous studies (I,V,VI) and three of these have not been defined in previous studies (A,B,C).

(Shuman & Lima, 2004)。进一步地结合序列保守性分析及结构比对信息,我们发现其他三个强作用区域(A、B、C)也高度保守,提示可能也具有功能作用。最近研究发现的一些功能残基位于我们预测的这三个强作用区域(Wang et al, 2008, 2009)。

### 3 讨论

传统的非键相互作用的计算工具无法在残基水平上批量自动计算,而批量自动计算非键相互作用能量则大大方便了对生物分子复合体系进行深入的结构与功能分析。本研究基于 DS2.0 的计算引擎,结合 Perl 脚本的编程,扩展了 DS2.0 原有的计算模块功能,实现了残基水平上的非键相互作用能量的批量计算。

在 PDB 结构数据库中,分子复合体系的结构数据远远少于单晶体的蛋白质数据,但是,生物分

子间的识别、结合和相互作用构成了一切生命活动的基础。因此,深入了解分子识别与接合机制对于了解生命过程具有重要意义。本文发展的新方法有助于对现有复合物结构数据的非键相互作用能量进行自动批量计算,便于研究者全面、准确掌握分子复合体中残基水平上的非键相互作用特征,对深入研究生物分子的结构-功能关系以及药物设计和筛选均有帮助意义。

由于 AutoNBEC 是在 DS2.0 非键相互作用模块的基础上开发出来的。因此,必须在 DS2.0 计算引擎环境下才能运行。DS2.0 为需要购买授权许可(license)的商业软件包,这一特点限制了 AutoNBEC 模块的发布和应用。

致谢:感谢 Nico Chen (Neotrident Inc.)在 protocol 编写中的建议与帮助。

### 参考文献:

- Ashenhurst JA. 2010. Intermolecular oxidative cross-coupling of arenes [J]. *Chem Soc Rev*, **39**(2): 540-548.
- Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations [J]. *J Comput Chem*, **4**: 187-217.
- Hirashima A, Huang H. 2008. Homology modeling, agonist binding site identification, and docking in octopamine receptor of *Periplaneta americana* [J]. *Comput Biol Chem*, **32**(3): 185-190.
- McCullagh M, Prytkova T, Tonzani S, Winter ND, Schatz GC. 2008. Modeling self-assembly processes driven by nonbonded interactions in soft materials [J]. *J Phys Chem B*, **112**(34): 10388-10398.
- Nakashima H, Furukawa K, Kashimura Y, Torimitsu K. 2008. Self-assembly of gold nanorods induced by intermolecular interactions of surface-anchored lipids [J]. *Langmuir*, **24**(11): 5654-5658.
- Pyrkov TV, Ozerov IV, Blitskaia ED, Efremov RG. 2010. Molecular docking: role of intermolecular contacts in formation of complexes of proteins with nucleotides and peptides [J]. *Bioorg Khim*, **36**(4): 482-492.
- Sagui C, Darden TA. 1999. Molecular dynamics simulations of biomolecules: long-range electrostatic effects [J]. *Annu Rev Biophys Biomol Struct*, **28**: 155-179.
- Shuman S, Lima CD. 2004. The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases [J]. *Curr Opin Struct Biol*, **14**(6): 757-764.
- Spasov VZ, Yan L. 2008. A fast and accurate computational approach to protein ionization [J]. *Protein Sci*, **17**(11): 1955-1970.
- Spriggs S, Garyu L, Connor R, Summers MF. 2008. Potential intra- and intermolecular interactions involving the unique-5' region of the HIV-1 5'-UTR [J]. *Biochemistry*, **47**(49): 13064-13073.
- Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigan C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. 2002. The Bioperl toolkit: Perl modules for the life sciences [J]. *Genome Res*, **12**(10): 1611-1618.
- Sundaram K, Prasad CV. 1982. A program to calculate non-bonded interaction energy in biomolecular aggregates [J]. *Comput Programs Biomed*, **14**(1): 41-46.
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. 2005. GROMACS: fast, flexible, and free [J]. *J Comput Chem*, **26**(16): 1701-1718.
- Wang LK, Nair PA, Shuman S. 2008. Structure-guided mutational analysis of the OB, HhH, and BRCT domains of *Escherichia coli* DNA ligase [J]. *J Biol Chem*, **283**(34): 23343-23352.
- Wang LK, Zhu H, Shuman S. 2009. Structure-guided mutational analysis of the nucleotidyltransferase domain of *Escherichia coli* DNA ligase (LigA) [J]. *J Biol Chem*, **284**(13): 8486-8494.
- Waters ML. 2002. Aromatic interactions in model systems [J]. *Curr Opin Chem Biol*, **6**(6): 736-741.
- Weiner PK, Kollman PA. 1981. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions [J]. *J Comput Chem*, **2**: 287-303.