Methods 65 (2014) 310–319

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Computational modeling of protein–RNA complex structures

Irina Tuszynska [a], Dorota Matelska [a], Marcin Magnus [a], Grzegorz Chojnowski [a], Joanna M. Kasprzak [b], Lukasz P. Kozlowski [a], Stanislaw Dunin-Horkawicz [a], Janusz M. Bujnicki [a,b,*]

[a] Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, PL-02-109 Warsaw, Poland
[b] Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, PL-61-614 Poznan, Poland

ARTICLE INFO

Article history:
Available online 29 September 2013

Keywords:
RNA
Protein
RNP
Protein–RNA binding
Macromolecular complex
Computational docking
Structural bioinformatics

ABSTRACT

Protein–RNA interactions play fundamental roles in many biological processes, such as regulation of gene expression, RNA splicing, and protein synthesis. The understanding of these processes improves as new structures of protein–RNA complexes are solved and the molecular details of interactions analyzed. However, experimental determination of protein–RNA complex structures by high-resolution methods is tedious and difficult. Therefore, studies on protein–RNA recognition and complex formation present major technical challenges for macromolecular structural biology. Alternatively, protein–RNA interactions can be predicted by computational methods. Although less accurate than experimental measurements, theoretical models of macromolecular structures can be sufficiently accurate to prompt functional hypotheses and guide e.g. identification of important amino acid or nucleotide residues. In this article we present an overview of strategies and methods for computational modeling of protein–RNA complexes, including software developed in our laboratory, and illustrate it with practical examples of structural predictions.

1. Introduction

Proteins and RNAs are the two main workhorses in the cell. Quite often they function together in many cellular processes, such as the maintenance of chromosome ends, transcription, RNA transport, RNA processing, regulation of gene expression, and protein synthesis [1,2]. It has been found that most of the human genome is transcribed into RNA, of which only a small fraction is translated into proteins [3], and that non-coding RNA transcription is also pervasive in bacteria [4]. The importance of protein–RNA assemblies is accentuated by the fact that most of RNAs studied so far require proteins to exert their function, and form ribonucleoprotein (RNP) complexes at least at some stage of their life cycle. Protein–RNA interactions are therefore essential to basic functioning of the cellular metabolism and to the survival of all organisms. Defects in protein–RNA interactions are implicated in a number of diseases, ranging from neurological disorders to cancer [5,6].

There are many protein domain families involved in RNA recognition, which are found in RNA-binding proteins (RBPs) [7]. In addition to domains with a well-defined fold, RNA can be recognized and bound by structurally disordered regions that change the conformation upon RNA binding, as exemplified by ribosomal proteins [8]. Obtaining three dimensional (3D) structures of protein–RNA complexes aids the efforts in the determination of physico-chemical principles of protein–RNA interactions. It reveals the details of interactions in RNPs and provides information about the specificity of mutual recognition of different proteins and RNAs. This knowledge is essential for the understanding of biological roles of protein–RNA interactions. However, the experimental determination of RNA–protein complexes (RNPs) structures is a slow and laborious process [9,10]. High-resolution structures of protein–RNA complexes were relatively rare until the end of the 20th century. As of June 2013, 1529 macromolecular complexes involving both protein and RNA components (but excluding RNA/DNA hybrids) were available in the Protein Data Bank (PDB), including 1323 solved by X-ray crystallography, 76 by Nuclear Magnetic Resonance (NMR) spectroscopy, and 130 by other methods. These structures contained 17138 protein chains interacting with RNAs, but many proteins were highly similar to each other. After removing redundant protein chains with sequence identity >90% or >40%, only 929 or 594 RNA-bound proteins with experimentally determined structures remained, respectively.

Despite the fact that the structures of many protein–RNA complexes have been experimentally determined, for many RNPs, e.g. the spliceosome, RNA-Induced Silencing Complex (RISC), complexes containing Clustered Regularly Interspaced Short Palin-

* Corresponding author at: Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, PL-02-109 Warsaw, Poland. Fax: +48 22 597 0715.

E-mail addresses: irena@genesilico.pl (I. Tuszynska), dmatelska@genesilico.pl (D. Matelska), magnus@genesilico.pl (M. Magnus), gchojnowski@genesilico.pl (G. Chojnowski), jkasp@amu.edu.pl (J.M. Kasprzak), lukaskoz@genesilico.pl (L.P. Kozlowski), sdh@genesilico.pl (S. Dunin-Horkawicz), iamb@genesilico.pl (J.M. Bujnicki).

1046-2023/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2013.09.014

dromic Repeats (CRISPR) and Cleavage/Polyadenylation Specificity Factor (CPSF), complete high-resolution structures (proteins complexed with RNA) are not yet available. For many such complexes, partial experimental data are available such as high-resolution structures of all or some isolated components in unbound conformations. For some complexes, low-resolution structures are available, obtained with e.g. cryo-electron microscopy (cryo-EM) or small angle X-ray or neutron scattering (SAXS/SANS) techniques.

Given the scarcity of experimentally determined structures of RNPs, computational techniques can complement existing data to help elucidation of protein–RNA interactions. There exists a wealth of low-resolution experimental data that identify the interacting components and often tie them to particular functional states. As an example, chemical cross-linking combined with mass spectrometry is often used to determine interacting subunits within a complex and sometimes allows the elucidation of individual interacting residues [11]. These data can be exploited by bioinformatics methods for structure prediction. However, while the methodology for prediction and modeling of proteins and protein–protein complexes is very well established (reviews: [12–14]), there are much fewer methods for predicting and modeling structures of RNA molecules and protein–RNA complexes [15,16].

In this article we review computational approaches for modeling of protein–RNA complex structures, with a focus on computer programs that are publicly available (e.g. as web servers or downloadable programs for local installation) and are easy to use by non-experts. We devote particular attention to combination of various complementary methods, and to the use of experimental information in the process of modeling. The selection of computational tools described and recommended in this article is based on various external benchmarks and on the authors' subjective experience. In particular, whenever appropriate, tools developed in the authors' laboratory are described. Hence, we must emphasize that the recommended workflows can be realized by the use of various different tools, also those that were not mentioned in this article due to space constraints.

2. Structure prediction for the individual components of a complex

Ideally, the prediction of protein–RNA complex structure should be based on the knowledge of atomic structures of the components, as determined by X-ray crystallography or NMR. However, in many cases experimentally determined structures of components of the complex are not available. For many applications they can be substituted by structures modeled computationally. To this end, a large number of computer programs have been developed, which now allow for reasonably accurate and practically useful predictions of macromolecular 3D structures. It must be emphasized that the protein 3D structure modeling field is more mature; for nearly two decades, the state-of-the art has been systematically assessed by the Critical Assessment of protein Structure Prediction (CASP) experiment organized by John Moult and coworkers (http://predictioncenter.org/). However, recent developments in the structural bioinformatics of RNA suggests that essentially the same principles are applicable for modeling of RNAs [16]. The most typical modeling strategies are reviewed below.

2.1. Template-based modeling

The most reliable strategy of macromolecular structure prediction is template-based modeling [17], also called comparative modeling or homology modeling. It relies on an observation that evolutionarily related (homologous) macromolecules often retain the same three-dimensional fold (i.e. the 3D arrangement and connectivity of secondary structure elements) despite the accumulation of divergent mutations [18]. Consequently, this strategy requires the identification of a molecule, for which the structure is known, and which can be used as a "template". Further, each element of the target sequence must be aligned to a structurally equivalent element in the template sequence/structure. In fact, the generation of a practically useful alignment is the most important step of template-based modeling. In the protein structure prediction field, the identification of modeling templates and the generation of target-template alignments has been termed "fold-recognition" and it is usually carried out as a separate step, before the actual 3D model building [19].

High sequence similarity is not a prerequisite for template-based modeling. In fact, it is possible to create accurate homology models even if the sequence identity between the target and the template is zero [20]. However, on the average, molecules with higher sequence similarity tend to exhibit more similar structures [18]. For highly similar sequences it is generally easier to generate a correct alignment, therefore the use of templates with higher sequence similarity is recommended. Apart from sequence divergence, structures may change because of environmental factors, e.g. the binding of other molecules [21]. Hence, the outcome of template-based modeling strongly depends on the choice of a template, whose biological state should correspond as closely as possible to the desired biological state of the target molecule to be modeled. For template-based modeling of components of protein–RNA complexes, it is recommended to use elements of related complexes as templates, as this usually helps to preserve the complementarity of the interface. So for instance for modeling of a complex between an amino acid-tRNA synthetase (aaRS) and its target tRNA, the best template is likely to be a structure of another, related aaRS–tRNA complex, rather than structures of related aaRSs and tRNAs solved in isolation.

There exist numerous computational methods for protein fold-recognition that facilitate the task of template structure identification and generation of target-template alignments. Our group has developed a web server (GeneSilico metaserver, http://genesilico.pl/meta2/) that runs numerous third-party methods for protein structure prediction, including more than 10-fold recognition methods, generates a user-friendly graphical output and calculates consensus between different predictions [22].

For RNA there exist much fewer automated computational methods that can be used to identify a modeling template and generate a target-template alignment (see Table 1 for examples). For RNA templates with high sequence identity, simple pairwise sequence search such as with BLAST [23] may be sufficient. Such utility is available e.g. via the ModeRNA modeling server [24], which uses a PARALIGN method [25]. However, the detection of remotely related structural templates for RNA modeling is difficult. Many methods developed so far serve the task of detection of homology to RNA families, e.g. those from the Rfam database, http://rfam.sanger.ac.uk/ regardless of the availability of a 3D structural information (for a critical comparison of algorithms see [26]). Among the methods for the detection of homology and sequence alignment with RNA families that may contain potential templates for modeling, we recommend Infernal (http://infernal.janelia.org/) [27] and CMcompare [28]. Both of these methods rely on covariance models, which describe probabilistically the conservation of both sequence and secondary structure (i.e. patters of Watson–Crick base pairs) in families of RNA molecules. In addition to searching for global matches of the target sequence to RNAs with known structure, we recommend also to search for local structural motifs, with tools such as RMDetect [29] or JAR3D [30].

As mentioned above, the actual modeling of protein and RNA 3D structures is a relatively easy task, provided that a target structure is identified and a reasonable target-template alignment is

Table 1
Examples of software tools for sequence searches and alignment.

Method name	Molecule type	Description	Refs.	Web link	Input
BLAST	Protein, RNA	Fast local sequence similarity search, used to find homologs of protein or RNA sequences	[23]	http://blast.ncbi.nlm.nih.gov/	Sequence
PARALIGN	Protein, RNA	An alternative to BLAST – similar to Smith–Waterman in sensitivity and to BLAST in speed; can be used for finding homologs that may pass undetected by other tools	[25]	http://www.paralign.org/	Sequence
GeneSilico metaserver	Protein	A gateway to over 100 bioinformatics tools for prediction of different features of proteins (domains, protein disorder, secondary structure, transmembrane proteins, fold recognition, protein–RNA interaction, etc.)	[22]	http://genesilico.pl/meta2/	Sequence, alignment
Infernal	RNA	Usage of RNA covariance models (CMs). Provides sensitive sequence database searching for RNA homologs and generates structure-based RNA sequence alignment	[27]	http://infernal.janelia.org/	Sequence, secondary structure
CMcompare	RNA	Comparison of RNA CMs. Can be used to find similarities to models present in the Rfam database	[28]	http://www.tbi.univie.ac.at/software/cmcompare/	Alignment, secondary structure
RMDetect	RNA	Detection of several structural motifs in RNA sequences	[29]	http://rmdetect.sourceforge.net/	Sequence, alignment
JAR3D	RNA	Detection of recurrent structural motifs in RNA sequences	[30]	http://rna.bgsu.edu/main/webapps/jar3d/	Sequence, alignment

available. There exist numerous tools for template-based modeling of proteins; among the most popular programs with user-friendly implementations there are MODELLER [31] and SWISS-MODEL [32] that are based on different approaches: minimization of spatial restraints and fragment copying, respectively. Essentially the same two approaches have been implemented in programs for template-based modeling of RNA: MacroMoleculeBuilder (MMB; previously RNABuilder) [33] and ModeRNA [34] (Table 2).

2.2. Template-free modeling and hybrid modeling

In the absence of a structural template, it is possible to generate a model of a protein or RNA 3D structure using alternative methods. There exist a host of methods that simulate the folding of macromolecular structures starting from sequence information alone, e.g. ROSETTA [35] and I-TASSER [36] for proteins, and FARNA [37], RNAComposer [38] and MC-Fold|MC-Sym [39] for RNA. They sample the conformational space and assess the conformations obtained to identify candidate structures that exhibit low "score" that is typically a combination of physical energy terms and statistical criteria.

However, template-free structure prediction is difficult. In practice, models of protein or RNA molecules longer than 80 residues that are folded completely "*de novo*" are very rarely accurate enough to make them useful for the next step, i.e. prediction of interactions. Template-free modeling is therefore often performed with additional data used as constraints and restraints. First, some parts of the model of a large (>80 residues) macromolecule can be built by a template-based method and constrained as rigid bodies, and free modeling is then applied only to sequence fragments, for which a template is not available or where the target-template alignment is uncertain. Second, information about secondary structure and information about long-range contacts (inferred experimentally or predicted computationally) can be encoded in the form of distance and orientation restraints. Third, information from experiments such as cryo-EM or SAXS can be used in the form of low-resolution restraints on the shape of molecules under consideration. MMB [40] can be cited as an example of a modeling method that is capable of handling both proteins and RNAs, with the aid of experimental data.

There exist methods for interactive (user-guided) modeling of macromolecular structures based on assembly of fragments derived from various structures that are predicted to be similar to different parts of the target. Computational tools and the graphics front-end facilitate the choice, the manipulation, and the visualization of fragments, and often provide specialized algorithms for local optimization of geometry. The approach that allows the expert user to rearrange and recombine multiple template structures has been particularly popular in the RNA modeling field, with methods such as S2S/Assemble [41,42] or RNA2D3D [43]. General-purpose graphical methods that facilitate the modeling of protein and nucleic acid molecules and their complexes include: Chimera [44], VMD [45], PyMOL [46], and SwissPDBViewer [47].

2.3. Model quality evaluation and local refinement

All structural models contain errors. They can range from globally wrong structures (e.g. due to the use of a wrong template, or due to improper folding by the template-free modeling method), to register shifts in sequence (usually due to errors in the alignment), to local inaccuracies such as wrong conformation of loops, steric clashes or problems with stereochemical parameters (e.g. wrong bond lengths). The accuracy of models is always limited by the errors. However, the impact of errors on the practical utility of models depends on the type of the question asked. Importantly, for some users high accuracy of some parts (e.g. regions of protein–RNA interaction) may be subjectively more important than other regions. The interpretation of a model should be made only at the level of its accuracy or lower, i.e. one should not make conclusions about atomic-level interactions for a model that is supposed to be accurate only at a level of individual residues. Hence, accuracy of each structural model has to be evaluated before the model

Table 2
Examples of software tools for protein and/or RNA structure prediction and modeling. Representation (in typical analyses): A-use of full-atom representation of macromolecular 3D structures, CG-use of models with a coarse-grained (reduced) representation.

Method name	Molecule type	Description	Refs.	Web link	Representation	Input
MODELLER	Protein	Implements homology modeling of protein 3D structures by satisfaction of spatial restraints. Can be used as a standalone program or online as a webserver	[31]	http://salilab.org/modeller/	A	Alignment, restraints, pdb file
SWISS-MODEL	Protein	Implements a rigid fragment assembly approach for homology modeling of protein 3D structures. Can be used via fully-automated server (SWISS-MODEL Workspace) or via the DeepView structure viewer	[32]	http://swissmodel.expasy.org/	A	Alignment, pdb file
ROSETTA	Protein	Conducts fragment assembly of protein 3D structures by a Monte Carlo approach, guided by an energy function that mixes knowledge-based and physical elements. Can be installed locally or used via a webserver (ROBETTA)	[35]	https://www.rosettacommons.org/	A, CG	Sequence
I-TASSER	Protein	Conducts hybrid modeling using a combination of fold-recognition and Monte Carlo-based assembly of fragments and folding simulation	[36]	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	A, CG	Sequence
QA-RecombineIt	Protein	Builds a hybrid model using fragments of models generated by other protein structure modeling methods. To accomplish this task, the program uses model quality assessment (QA) and a fragment recombination algorithm (RecombineIt)	[51]	http://genesilico.pl/qarecombineit/	A	Sequence, pdb file
ModeRNA	RNA	Builds comparative models of RNA 3D structures based on known structures as templates, using a fragment replacement approach. Aavailable as a web server and as a standalone program. Allows for modeling of modified nucleotides	[24]	http://genesilico.pl/moderna/	A	Alignment, pdb file
FARNA/FARFAR	RNA	Counterpart of ROSETTA for *de novo* modeling of RNA. Conducts fragment assembly of RNA structures	[37]	https://www.rosettacommons.org/manuals/rosetta3_user_guide/index.html	A, CG	Sequence
RNA Composer	RNA	Predicts large RNA 3D structures by fragment assembly	[38]	http://rnacomposer.ibch.poznan.pl/Home	A	Sequence, secondary structure
MC-Fold | MC-Sym	RNA	Predicts RNA 2D and 3D structures by enumeration of alternative conformations	[39]	http://www.major.iric.ca/MC-Fold/	A	Sequence, secondary structure
S2S/Assemble	RNA	An interface for largely manual modeling of RNA structures with the aid of heterogeneous data, such as multiple sequence alignments, secondary and tertiary structures	[41,42]	http://bioinformatics.org/S2S/	A	Alignment, secondary structure, pdb file
RNA2D3D	RNA	Generates a first-pass, low-resolution 3D structures from secondary structure that can be used to explore alternative 3D conformations	[43]	http://www-lmmb.ncifcrf.gov/users/bshapiro/rna2d3d/rna2d3d.html	A	Sequence, secondary structure
SimRNA	RNA	Conducts folding simulations of RNA using a Monte Carlo approach and a knowledge-based energy function	[16]	Program is available upon request from the authors of this article	CG	Sequence, secondary structure
Macro-Molecule-Builder	RNA, protein	Constructs 3D structures of individual macromolecules and complexes by applying user-specified restraints and physical constraints	[33]	https://simtk.org/home/rnatoolbox	A	Sequence, pdb file
HyperChem	Protein, RNA	A molecular modeling environment allowing 3D visualization and animation with quantum chemical calculations, molecular mechanics, dynamics etc.	HyperCube, Inc.	http://www.hyper.com/	A	Pdb file
FILTREST3D	Protein, RNA	A server for scoring and ranking of models according to their agreement with user-defined restraints (distances, secondary structure etc.)	[73]	http://filtrest3d.genesilico.pl/filtrest3d/index.html	A, CG	Restraints, pdb file

Table 3
Examples of software tools for quality evaluation and local refinement of macromolecular models. Representation (in typical analyses): A-use of full-atom representation of macromolecular 3D structures, CG-use of models with a coarse-grained (reduced) representation.

Method name	Molecule type	Description	Refs.	Web link	Representation	Input
Pcons	Protein	Identifies potentially best models in large datasets of alternative by combination of clustering and application of knowledge-based scoring functions	[48]	http://pcons.net/	A, CG	Pdb file
ModFold	Protein	Identifies potentially best models in large datasets of alternative by combination of clustering and application of knowledge-based scoring functions. If a single model is provided, alternative models are generated automatically. Provides global and local (per-residue) quality of 3D protein models	[49]	http://www.reading.ac.uk/bioinf/ModFOLD/index.html	A, CG	Sequence, pdb file
Meta-MQAP	Protein	Generates consensus prediction of model quality based on results of eight primary MQAPs. Provides global and local (per-residue) quality of 3D protein models	[50]	https://genesilico.pl/toolkit/unimod?method=MetaMQAPII	A	Pdb file
RASP	RNA	Provides global and local (per-residue) quality of 3D RNA models using a set of all-atom and coarse-grained knowledge-based potentials. Also available as a web server (WebRASP)	[52,53]	http://melolab.org/webrasp/home.php	A, CG	Pdb file

is used for any purpose. Ideally, a prospective user of a given model should be informed if a model is likely to be globally accurate (i.e. if its components are placed more or less where they are in reality), and what are the relative inaccuracies of its particular regions.

The absolute evaluation of accuracy requires a comparison of the model to the experimentally determined structure to be made, but in most "real life" situations, models are built for such macromolecules, for which no experimental structure is available, hence predictive methods must be used. This problem has been addressed initially in the protein structure prediction field, and numerous so-called Model Quality Assessment Programs (MQAPs) were developed (Table 3). Based on the results of recent CASP experiments, we recommend the use of methods such as Pcons [48], ModFold [49], and our own method QA-RecombineIt, which not only predicts the quality of individual protein models [50], but also recombines best parts into an optimized hybrid model [51]. More recently, MQAPs for RNA models have been also developed. From our own experience, among the publicly available methods we tested, the RASP force field [52] exhibits very good performance in discrimination of highly to moderately accurate RNA structure models from ones that are completely misfolded. RASP has been recently made available as a web server [53]. Further, for discrimination of protein–RNA complex models specialized scoring functions have been developed, which will be discussed later, in the context of protein–RNA docking.

2.4. Predicting disordered structures

Methods discussed above are based on the assumption that a well-defined three-dimensional structure exists that can be determined experimentally or obtained by a predictive approach. However, many macromolecules are known to have either multiple conformational states (such as allosteric proteins and riboswitches), or possess intrinsically disordered/unstructured regions (IURs). Many RNA-binding proteins possess IURs that may to some extent become ordered only upon RNA binding [54]. The identification of IURs may be therefore used to eliminate disordered sequences from the initial process of model building for the individual components (especially with template-based methods). However, IURs may need to be included in some form at the stage of protein–RNA complex building.

Recently, a number of methods for predicting structural disorder in protein sequences have been developed. To leverage on these developments, our group has developed a web server Metadisorder [55] that predicts IURs based on input from a number of individual methods, with a use of machine learning techniques. To our best knowledge, equivalent methods for IUR prediction in RNA sequences are not yet available. It is however possible to infer certain types of IURs in RNA sequences by studying the results of secondary structure prediction methods and identifying long regions without predicted base-pairing, or in which different methods predict different structures. The RNA secondary structure prediction meta-server developed in our group (http://genesilico.pl/rnametaserver/) can aid the users in accomplishing this task.

3. Structure prediction of protein–RNA complexes

3.1. Prediction of RNA-binding residues in proteins

The question whether a given protein binds RNA or not remains a challenge for computational methods, while it can be now relatively easily answered with experimental methods. However, once a given protein is found to bind RNA, computational methods can be very useful in predicting amino acid residues that are most likely to form RNA-binding sites. In particular, the availability of a protein structure (determined experimentally or predicted computationally) can greatly facilitate the prediction of the RNA-binding site, which is typically formed by surface-exposed residues that are close to each other in space, but not necessarily in sequence.

A number of methods have been developed to predict RNA-binding residues from protein sequence alone or from protein structure (Table 4). Based on recent surveys [56,57] we recommend to use different approaches, depending on the data available. If a structure is available for the target protein, the user should check whether the target exhibits similarity to other RNA-binding proteins with already known structures in complex with RNA. If

Table 4
Examples of software tools for prediction of RNA-binding residues in protein sequences.

Method name	Description	Refs.	Web link	Input(protein only)
RNA BindR Plus	Uses a naive Bayes classifier trained on structures of protein–RNA complexes found in the PDB	[58]	http://einstein.cs.iastate.edu/RNABindRPlus/	Sequence
KYG	Uses RNA-binding propensities of individual residues, doublets of spatially close residues, sequence profiles, and combinations thereof	[59]	http://cib.cf.ocha.ac.jp/KYG/	Alignment, pdb file
OPRA	Calculates a score derived from propensities of residues at known protein–RNA interfaces weighed by their accessible surface	[60]	Program is available upon request from the authors	Pdb file
DRNA	Conducts a structural alignment to identify sites similar to those already known in other protein–RNA complex structures, followed by binding assessment with a DFIRE statistical energy function	[61]	http://sparks.informatics.iupui.edu/yueyang/server/dRNA-DB/	Pdb file

this is the case, the RNA-binding site in the target is best predicted "by homology" to the template(s). This can be done manually, by superposition of structures, or with the use of programs such as RNABindRPlus [[58], new version available at: http://einstein.cs.iastate.edu/RNABindRPlus], KYG [59], OPRA [60], and DRNA [61]. For a target protein with known structure that has no close homolog, we recommend to use KYG and to compare its result with predictors based on sequence alone (e.g. run via a meta-server developed by our group and available at http://iimcb.genesilico.pl/meta2/).

3.2. Modeling of protein–RNA complexes by macromolecular docking

Docking methods aim at predicting three-dimensional structures of macromolecular complexes, starting from the atomic coordinates of their components. The larger molecule is usually referred to as the receptor, while the smaller molecule is usually called the ligand. The docking can be divided into two steps that are analogous to template-free modeling of macromolecular structure: (i) sampling: the search of the conformational space of possible orientations and conformations of the components and thereby generation of sample models (called poses or decoys) and (ii) scoring: assessment of these models by a scoring function to distinguish near-native structures from non-native ones [62]. Some methods combine both steps, while others specialize only in the assessment of models/poses, leaving their generation to the user.

Structures of binding partners, which are solved individually or with another partner, often undergo conformational changes during association, in a process known as induced fit. Although during recent years innovative solutions have been proposed to taking macromolecular flexibility into account [63], modeling conformational changes involving backbone and loop rearrangements remains the biggest challenge in the modeling of macromolecular complexes [64]. Some methods model conformational changes of docking components explicitly, which makes such analyses computationally demanding, while others introduce a certain level of 'fuzziness' (review: [65]). There are also hybrid approaches that initially carry out a global search for approximate solutions, followed by the refinement phase, where small conformational changes are explored.

Thus far, a large number of protein–protein docking methods have been developed and assessed in the course of the Critical Assessment of PRediction of Interactions (CAPRI) experiment, analogous to CASP (review: [66]). Compared to protein–protein docking, protein–RNA docking has received relatively little attention from developers of computational methods. For instance in the last CAPRI experiment [64], only four groups participated in the competition of scoring protein–RNA alternative models. Most methods for protein–RNA docking have been actually developed by modifying protein–protein docking methods in order to accept nucleic acid molecules as receptors and/or ligands. Only recently a docking method 3dRPC was developed with a specific purpose of protein-RNA docking, which takes special features of RNA surfaces into account. Computer programs for macromolecular docking that accept protein and RNA coordinates as an input to generate RNP complex decoys include HADDOCK [68], GRAMM [69], HEX [70], PatchDock [71], and FTDock [72] (Table 5). Among these tools, HADDOCK is most versatile in the ability to use user-defined restraints (e.g. from experimental data) to drive the docking, and in the handling of flexibility both on the protein and the nucleic acid side; this method is also available as a web server [73]. During the last CAPRI experiment, HADDOCK produced relatively best models for protein–RNA complexes [64]. However, in our experience, HADDOCK requires parametrization of RNA molecules to be docked, which can be quite cumbersome even if a user intends to run only the first stage of the docking process, i.e. the generation of poses. On the other end of the spectrum there is GRAMM, which is very easy to install and use, but is unable to directly utilize experimental information.

It must be emphasized that purely theoretical protein–RNA docking, while possible, performs rather poorly. For modeling of protein–RNA complexes it is generally useful to combine the use of docking methods for the generation of physically reasonable alternative poses with additional ranking by external methods. These may include: (i) the use of restraints derived from experimental data or from computational predictions of residues involved in binding (see above), e.g. as implemented in the FILTREST3D method [74] and (ii) the use of scoring functions specialized in the discrimination of native-like structures. For example, our group has developed statistical potentials QUASI-RNP and DARS-RNP that are deliberately coarse-grained to take into account moderate conformational changes [75]. We have also developed a web server RNPdock (http://iimcb.genesilico.pl/RNPdock/) that takes as an input a complex comprising a protein molecule and an RNA molecule and conducts optimization of their mutual position according to the DARS-RNP score.

3.3. Modeling of large macromolecular complexes guided by experimental data

Most of docking methods have been developed having in mind the modeling of binary complexes, i.e. comprising one protein

Table 5
Examples of software tools for modeling of protein–RNA complexes by macromolecular docking and for scoring of protein–RNA structural models. Representation (in typical analyses): A-use of full-atom representation of macromolecular 3D structures, CG-use of models with a coarse-grained (reduced) representation.

Method	Description	Ref.	Web link	Representation	Input
HADDOCK	A very versatile method for docking of various molecules including proteins, nucleic acids and small molecules. Relies on user-defined restraints derived from various sources. The program is available both as a standalone program and a web server	[73]	http://www.nmr.chem.uu.nl/haddock/	A	Pdb files, restraints
GRAMM	A low-resolution rigid body docking program that accepts protein as well as nucleic acids and small molecules. Does not have a special scoring function for protein–RNA complexes	[69]	http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/	A, CG	Pdb files
HEX	A fast algorithm that enables protein–protein and protein–nucleic acid docking. Allows for restriction of docking to pre-defined binding sites. Does not have a special scoring function for protein–RNA complexes	[70]	http://hex.loria.fr/	A, CG	Pdb files, restraints
PatchDock	A geometry-based molecular docking algorithm available both as a standalone program and a web server. Allows for restriction of docking to pre-defined binding sites. Does not have a special scoring function for protein–RNA complexes	[71]	http://bioinfo3d.cs.tau.ac.il/PatchDock/	A	Pdb files, restraints
FTDock	A rigid-body docking method. It does not accept modified residues and does not have a special scoring function for protein–RNA complexes	[72]	http://www.sbg.bio.ic.ac.uk/docking/ftdock.html	A	Pdb files
DARS-RNP & QUASI-RNP	Knowledge-based, statistical and quasi-chemical potentials for scoring of protein–RNA decoys obtained with other methods	[75]	http://iimcb.genesilico.pl/RNP/	A, CG	Pdb files
RNPDock	A server that takes as an input a preliminary model of protein–RNA complex and conducts rigid body optimization using the DARS-RNP score	Unpublished	http://iimcb.genesilico.pl/RNPdock/	A, CG	Pdb files
3dRPC	A protocol specifically developed for protein-RNA docking, which includes both a docking procedure and a scoring function.	[67]	http://biophy.hust.edu.cn/download.html	A, CG	pdb files
Integrative Modeling Platform (IMP)	A versatile software toolkit for structural modeling of biomolecules ranging in size and complexity from small peptides to large macromolecular assemblies. Relies on integration of restraints from diverse experiments	[76]	http://salilab.org/imp/	A, CG	Pdb files, restraints
PyRy3D	A simple software tool for structural modeling of macromolecular complexes based on user-defined restraints. Enables the use of rigid bodies, intrinsically disordered regions and regions of uncertain structure	Unpublished	http://iimcb.genesilico.pl/pyry3d/	A, CG	Pdb files, restraints

component and one RNA component. However, numerous important RNP complexes such as those mentioned earlier in this article, comprise multiple components. For multi-component complexes, structure prediction by purely computational docking becomes extremely difficult, as the number of mutual placements and orientations of subunits that should be considered grows to very large numbers. To address this problem, computational methods have been developed that can use additional information from a variety of sources, to restrict the search space and make the modeling computationally manageable. Among the aforementioned tools, HADDOCK web server is capable of data-driven simultaneous docking of up to 6 components [73]. Methods such as Integrative Modeling Platform (IMP; http://salilab.org/imp/) [76] or PyRy3D (developed in the authors' group: http://iimcb.genesilico.pl/pyry3d/) are capable of building macromolecular models comprising dozens of components. The modeling is guided primarily by restraints that can be derived from biochemical and biophysical experiments for the determination of molecular shape (e.g. by cryo-EM or SAXS), determination of interactions between different subunits (e.g. by chemical cross-linking and mass spectrometry, FRET or EPR spectroscopy, etc.), determination of binding sites by various types of foot-printing and many other types of analyses. With these tools it is possible to integrate, into a common framework, structural information collected at multiple levels of the biological hierarchy: from single atoms, to residues, to protein and nucleic acid subunits, to the higher-order assemblies. Usually a result of such a modeling is not just a single model, but a set of models that are maximally consistent with the input data. Hence, in case of insufficient data, the user is expected to obtain multiple alternative models, which should be further analyzed to plan additional experiments that would be capable of discrimination between various alternatives.

Among the example methods mentioned, IMP is definitely the most versatile. It is actually not a single program, but a software package that delivers a big variety of tools and scripts to proceed from stages of data gathering, selection of representation for models, sampling and optimization schemes, to scoring and analyzing of the models. However, it is also very complex, and for a typical modeling exercise it requires essentially a dedicated program of instructions to be written, which can be a significant barrier for users. PyRy3D lies on the other end of the spectrum – it is relatively simplistic, but was developed with user-friendliness in mind, and can be run entirely via a graphics interface. A highlight of PyRy3D is that it facilitates the modeling of complexes comprising components, for which some of the components are intrinsically disordered or for which the 3D structure cannot be reliably predicted. This feature is particularly useful for the modeling of single-stranded RNA regions and for regions of protein sequence that fold only upon interaction with the target RNA.

4. Practical Examples

4.1. Modeling of a protein–RNA complex by combination of template-based modeling and protein–RNA docking

We used template-based modeling for both RNA structure and RNA–protein complex prediction in the course of the RNApuzzles experiment [77], which is a counterpart of the CASP experiment for RNA modeling. While most of the modeling tasks were focused on modeling of just an RNA molecule, one task involved 3D structure prediction for a complex between the YbxF protein from *Bacillus subtilis* and a SAM-I riboswitch.

First, the structures of the YbxF protein and the SAM-I riboswitch RNA were predicted by template-based modeling. We carried out the protein fold-recognition procedure via the GeneSilico metaserver (http://genesilico.pl/meta2/), and obtained alignment of the YbxF sequence to the L7Ae protein with known structure (PDB code: 2fc3). The model of YbxF 3D structure was built using MODELLER [78] and its accuracy was predicted to be acceptable

by the MetaMQAP method [50]. Likewise, the structure of the SAM-I riboswitch RNA was predicted by first running a search with the ModeRNA server [24] and detecting a match to a homologous RNA from *Thermoanaerobacter. tengcongensis* (PDB code: 3iqp), followed by 3D structure modeling with ModeRNA [34].

Second, having structural models of both components of complex, we analyzed the known structures of protein–RNA complexes that contain proteins with a 3D fold similar to that of YbxF, in order to predict the RNA-binding site of YbxF. Among proteins with a similar fold (d.73 according to the SCOP database [79]) we identified 62 structures that presented a protein–RNA complex and their superposition revealed a clear structural conservation of the RNA binding site (data not shown). Following the structural superposition of RNA molecules from the above-mentioned complexes, we identified conserved RNA sequences in the area of the interaction site (residues 16-GUG-18 and 33-AUGA-36). Central residues U17 and U34 within the two RNA sequence motifs were then selected as potential interaction partners of the corresponding protein residues K21 and A76.

Third, we used the GRAMM method to generate 50,000 alternative models of YbxF–SAM-I riboswitch complex and we ranked these models with FILTREST3D [74]. We retained structures, in which the RNA was bound only in the predicted binding site defined by two distance restraints mentioned above, in which either a distance between NZ-P atoms of Lys21-U17 pair or a distance between CB-P atoms of Ala76-U34 pair was smaller than 13 Å, and in which the sum of deviations from the 13 Å threshold was smaller than 5 Å667 models that fulfilled this criterion were evaluated using the DARS-RNP potential and clustered according to geometrical similarity. The biggest cluster contained solutions, in which the orientation of protein and RNA molecules was similar to that in homologs with known structure, which supported the docking. We selected five representative models as our predictions; we optimized their local geometry using HyperChem (HyperCube, Inc.) and submitted them for evaluation to the organizers of the RNA Puzzles experiment.

The structure of the YbxF–SAM-I riboswitch complex has been determined by crystallography and is now available in the PDB database (PDB code: 3v7e). Thus far, the organizers of the RNA Puzzles experiment have reported only the accuracy of the RNA moiety. In order to assess the accuracy of the protein–RNA complex, we superimposed the binding site on the side of the RNA in the experimental model and in the blind theoretical models (see Fig. 1 for a comparison of our model number 1 to the experimentally determined structure). The results indicate that the complex structure as well as the mutual binding mode of the protein and the RNA were predicted with relatively high accuracy; with RNA molecules superimposed, the protein model shows RMSD of 2.81 Å to its experimentally determined counterpart. Also the models predicted 77% of protein–RNA contacts present in the experimentally determined structure. The biggest differences between the protein model and the protein native structure are within one of the α helices (Pro39–Gly53) (Fig. 1A). The difference in local structure between the protein model and the experimental structure leads to a decrease in the number of the native-like contacts, but it does not prevent an overall correct prediction of orientations and contacts between the protein and RNA molecules (Fig. 1B).

4.2. Modeling of a protein–RNA complex by combination of template-based and template-free modeling

The S6:S18 ribosomal protein complex is known to bind to the 16S rRNA and the molecular details of this interaction are well studied [80]. In our recent study [81], we discovered that this protein complex also interacts with a structural motif present in the 5′

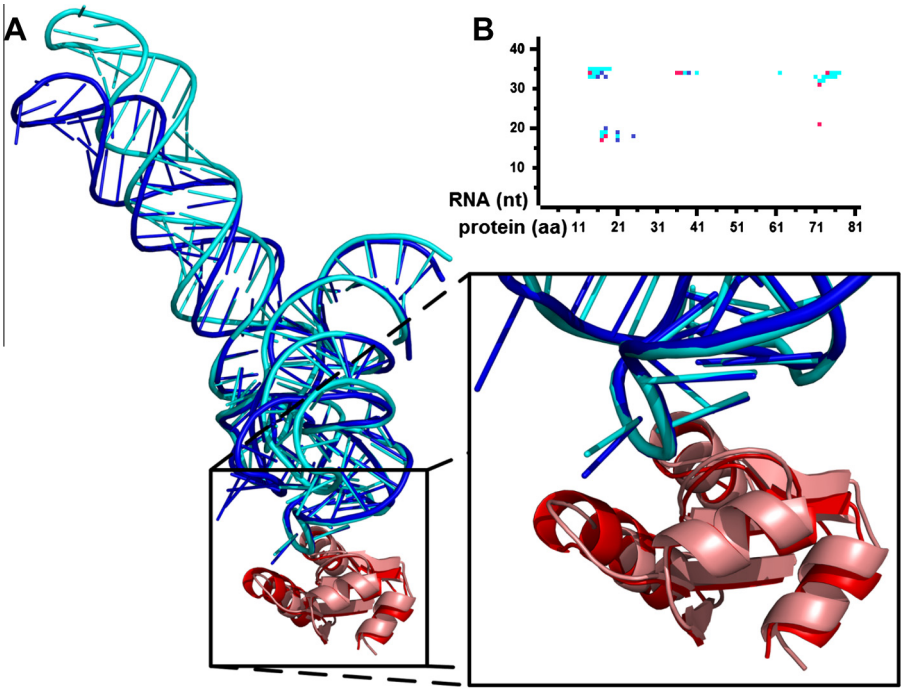

Fig. 1. Comparison of the predicted model of the YbxF–SAM-I riboswitch complex with the experimentally determined structure. (A) Superposition of structures. The protein and RNA moieties are shown in salmon and cyan in the model, and in red and blue in the experimentally determined structure. (B) Comparison of protein–RNA contacts (defined as nucleotide and amino acid residues having at least one pair of non-hydrogen atoms at a distance ⩽ 5 Å) in the model and in the experimentally determined structure: correctly predicted contacts are shown in cyan, contacts present only in the experimentally determined structure (not observed in the model) are shown in red, and contacts present only in the model are shown in blue. Residue numbers are indicated on the axes: *x* (protein) and *y* (RNA). Maps of protein–RNA contacts were calculated and visualized using RNAmap2D, a program developed in the authors' laboratory [84].

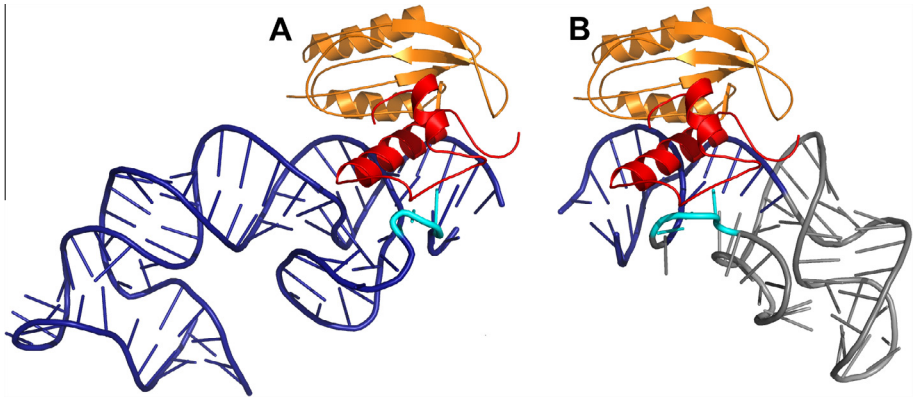

Fig. 2. Left: crystal structure of *T. thermophilus* S6:S18 ribosomal proteins complex (S6 and S18 subunits are colored in orange and red, respectively) bound to the three way junction in 16S rRNA (PDB code: 1g1x). Right: model of the *E. coli* S6:S18 ribosomal proteins complex bound to a structural motif present in the 5′ untranslated region of its own mRNA. The interface region modeled based on *T. thermophilus* structure as a template is shown in dark blue, with the "CCR" pattern shown in cyan. The region modeled in template-free manner is shown in grey.

untranslated region of its own mRNA, and thus presumably regulates translation of the encoded genes. We also found that the mRNA motif contains a highly conserved "CCR" sequence pattern, identical to the one that contributes base-specific contacts in the rRNA-S6:S18 complex interface, and we predicted that the mRNA motif structurally mimics the protein binding site on the ribosome. Thus, we decided to build a model of the S6:S18-mRNA complex, using known structure of the S6:S18-rRNA complex as a template.

The structure of *Escherichia coli* S6:S18 protein complex was taken from the crystal structure of the ribosome (PDB code: 3i1m). The mRNA motif was modeled using a combined template-based and template-free approach, previously proven to be successful in the RNA-puzzles experiment [77]. The interface region of the target RNA containing the "CCR" pattern was modeled by a comparative approach using ModeRNA [34]; the three-way junction of the 16S rRNA bound by S6:S18 complex from *Thermus thermophilus* (PDB code: 1g1x) was used as a template. This template was also essential to define the relative arrangement of the protein and RNA components of the model, i.e. the core of the protein–RNA interface involving the CRR trinucleotide was assumed to be identical between the rRNA–protein complex (the template) and the mRNA–protein complex (the target to be modeled). In this case we have not attempted protein–RNA docking. The rest of the RNA structure was modeled in a template-free manner, with restraints on secondary structure, using SimRNA, a method developed in the authors' laboratory, which uses a coarse-grained representation and a statistical potential [16]. Our model of S6:S18-mRNA complex (Fig. 2) remains to be verified experimentally.

5. Conclusion

Over recent years interest in studying protein–RNA complexes has been rapidly growing. The number of structures of protein-nucleic acid complexes deposited in the PDB database increases each year (269 in 2007 vs. 459 in 2012), as well as more publications associated with the term "RNA-binding proteins" appear in PubMed (1874 in 2007 vs. 2226 in 2012). However, due to the difficulty in the structure determination of protein–RNA complexes, there is also an increasing demand for the development of computational methods for predicting such structures either from structures of the components or directly from sequences. The recognition of importance of modeling RNP complexes is reflected in their recent inclusion as targets both in CAPRI and RNA Puzzles experiments.

All the presented approaches for modeling RNP complexes can provide practically useful predictions, but they also suffer from various limitations. One problem specific to RNA modeling is the relative paucity of experimentally determined RNA and RNP structures that can be used as templates. It is hoped that the significance of this problem will wane with the growing interest in structural biology of RNA and correspondingly increasing number of structures solved each year. Another serious problem is at the stage of docking: the existing computational methods seldom take into account conformational changes that may occur upon binding, both in protein and RNA components. One of the possible solutions is to combine the existing tools that enable template-free modeling of the protein and RNA components, with scoring functions for the assessment of intermolecular contacts. The first step in this direction that has been already made is the development of modeling techniques that start with preliminary models obtained by e.g. rigid body docking of "unbound" protein and RNA structures, and carry out re-folding of protein and RNA fragments that participate in their mutual interactions. The conceptual similarity of successful algorithms for protein and RNA 3D structure modeling [16] suggests that their combination into unified modeling methods is feasible. Synergy is also expected to appear from the combination of theoretical predictive methods with low-resolution experimental analyses. It has been suggested that the structure of many RNP complexes, such as the spliceosome, may be modeled using cryo-EM maps as molecular envelopes into which structures of individual components could be fitted, using restraints from biochemical experiments and other bioinformatics-based predictions [82]. This requires the development of new multiresolution modeling methods and new ways of encoding experimental data (review: [83]). We hope that the recent surge of interest in protein-RNA interactions will encourage both biologists to use bioinformatics tools to obtain structural insight into the systems, for which they have obtained experimental data, and the developers of tools to propose new algorithms and their user-friendly implementations.

Acknowledgements and Funding

The research on protein–RNA interactions in the Bujnicki laboratory was funded primarily by the Foundation for Polish Science (FNP, grant TEAM/2009-4/2) and by the European Research Council (ERC, StG grant RNA + P = 123D). The development of our servers for protein bioinformatics was funded mainly by the Polish Ministry of Science and Higher Education (MNiSW, grant POIG.02.03.00-00-003/09). J.M.B. was additionally supported by the "Ideas for Po-

land" fellowship from the FNP. S.D.-H., J.M.K. and G.Ch. were additionally supported by the National Science Centre (NCN, grants 2011/03/D/NZ8/03011 to S.D.-H. 2012/05/N/NZ2/01652 to J.M.K. and 2011/01/D/NZ1/00212 to G.Ch.). S.D.-H. also acknowledges support from the Polish Ministry of Science and Higher Education (MNiSW, fellowship for outstanding young scientists). I.T. was also supported by MNiSW (Iuventus grant number 0570/IP1/2011/71) and by FNP (START fellowship). L.P.K. was additionally supported by MNiSW (Iuventus grant number 0301/IP1/2013/72). We thank Grzegorz Lach for participation in the modeling of the SAM-I riboswitch RNA structure.

References

[1] Y. Chen, G. Varani, FEBS J. 272 (2005) 2088–2097.
[2] T. Glisovic, J.L. Bachorik, J. Yong, G. Dreyfuss, FEBS Lett. 582 (2008) 1977–1986.
[3] J. Cheng, P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt, V. Sementchenko, A. Piccolboni, S. Bekiranov, D.K. Bailey, M. Ganesh, S. Ghosh, I. Bell, D.S. Gerhard, T.R. Gingeras, Science 308 (2005) 1149–1154.
[4] I. Lasa, A. Toledo-Arana, A. Dobin, M. Villanueva, I.R. de los Mozos, M. Vergara-Irigaray, V. Segura, D. Fagegaltier, J.R. Penades, J. Valle, C. Solano, T.R. Gingeras, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 20172–20177.
[5] K.E. Lukong, K.W. Chang, E.W. Khandjian, S. Richard, Trends Genet. 24 (2008) 416–425.
[6] T.A. Cooper, L. Wan, G. Dreyfuss, Cell 136 (2009) 777–793.
[7] S. Cammer, C.W. Carter Jr., Bioinformatics 26 (2010) 709–714.
[8] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, Science 289 (2000) 905–920.
[9] A. Ke, J.A. Doudna, Methods 34 (2004) 408–414.
[10] L.G. Scott, M. Hennig, Methods Mol. Biol. 452 (2008) 29–61.
[11] H. Steen, O.N. Jensen, Mass Spectrom. Rev. 21 (2002) 163–182.
[12] D. Wichadakul, J. McDermott, R. Samudrala, Methods Mol. Biol. 541 (2009) 101–143.
[13] I.S. Moreira, P.A. Fernandes, M.J. Ramos, J. Comput. Chem. 31 (2010) 317–342.
[14] J.M. Bujnicki, Prediction of Protein Structures, Functions and Interactions, Wiley & Sons, 2008.
[15] C. Laing, T. Schlick, J. Phys.: Condens. Matter 22 (2010) 283101.
[16] K. Rother, M. Rother, M. Boniecki, T. Puton, J.M. Bujnicki, J. Mol. Model. 17 (2011) 2325–2336.
[17] A. Kryshtafovych, K. Fidelis, J. Moult, Proteins 79 (Suppl. 10) (2011) 196–207.
[18] C. Chothia, A.M. Lesk, EMBO J. 5 (1986) 823–826.
[19] J. Soding, M. Remmert, Curr. Opin. Struct. Biol. 21 (2011) 404–411.
[20] C. Chothia, M. Gerstein, Nature 385 (579) (1997) 581.
[21] S. Kumar, B. Ma, C.J. Tsai, N. Sinha, R. Nussinov, Protein Sci. 9 (2000) 10–19.
[22] M.A. Kurowski, J.M. Bujnicki, Nucleic Acids Res. 31 (2003) 3305–3307.
[23] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389–3402.
[24] M. Rother, K. Milanowska, T. Puton, J. Jeleniewicz, K. Rother, J.M. Bujnicki, Bioinformatics 27 (2011) 2441–2442.
[25] T. Rognes, Nucleic Acids Res. 29 (2001) 1647–1652.
[26] E.K. Freyhult, J.P. Bollback, P.P. Gardner, Genome Res. 17 (2007) 117–125.
[27] E.P. Nawrocki, D.L. Kolbe, S.R. Eddy, Bioinformatics 25 (2009) 1335–1337.
[28] F. Eggenhofer, I.L. Hofacker, C. Honer Zu Siederdissen, Nucleic Acids Res. 41 (2013) W499–503.
[29] J.A. Cruz, E. Westhof, Nat. Methods 8 (2011) 513–521.
[30] W. Kladwang, F.C. Chou, R. Das, J. Am. Chem. Soc. 134 (2012) 1404–1407.
[31] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779–815.
[32] M.C. Peitsch, Bio/Technology 13 (1995) 658–660.
[33] S.C. Flores, Y. Wan, R. Russell, R.B. Altman, Pac. Symp. Biocomput. (2010) 216–227.
[34] M. Rother, K. Rother, T. Puton, J.M. Bujnicki, Nucleic Acids Res. 39 (2011) 4007–4022.
[35] R. Das, D. Baker, Annu. Rev. Biochem. 77 (2008) 363–382.
[36] A. Roy, A. Kucukural, Y. Zhang, Nat. Protoc. 5 (2010) 725–738.
[37] R. Das, D. Baker, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 14664–14669.
[38] M. Popenda, M. Szachniuk, M. Antczak, K.J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R.W. Adamiak, Nucleic Acids Res. 40 (2012) e112.
[39] M. Parisien, F. Major, Nature 452 (2008) 51–55.
[40] S.C. Flores, R.B. Altman, RNA 16 (2010) 1769–1778.
[41] F. Jossinet, E. Westhof, Bioinformatics 21 (2005) 3320–3321.
[42] F. Jossinet, T.E. Ludwig, E. Westhof, Bioinformatics 26 (2010) 2057–2059.
[43] H.M. Martinez, J.V. Maizel Jr., B.A. Shapiro, J. Biomol. Struct. Dyn. 25 (2008) 669–683.
[44] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.
[45] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. 14 (33–38) (1996) 27–38.
[46] S. LLC., The PyMOL Molecular Graphics System, Version 1.5.0.4.
[47] N. Guex, M.C. Peitsch, Electrophoresis 18 (1997) 2714–2723.
[48] M.J. Skwark, A. Elofsson, Bioinformatics 29 (2013) 1817–1818.
[49] L.J. McGuffin, M.T. Buenavista, D.B. Roche, Nucleic Acids Res. 41 (2013) W368–W372.
[50] M. Pawlowski, M.J. Gajda, R. Matlak, J.M. Bujnicki, BMC Bioinformatics 9 (2008) 403.
[51] M. Pawlowski, A. Bogdanowicz, J.M. Bujnicki, Nucleic Acids Res. 41 (2013) W389–397.
[52] E. Capriotti, T. Norambuena, M.A. Marti-Renom, F. Melo, Bioinformatics 27 (2011) 1086–1093.
[53] T. Norambuena, J.F. Cares, E. Capriotti, F. Melo, Bioinformatics (2013), http://dx.doi.org/10.1093/bioinformatics/btt1441.
[54] J. Bellay, S. Han, M. Michaut, T. Kim, M. Costanzo, B.J. Andrews, C. Boone, G.D. Bader, C.L. Myers, P.M. Kim, Genome Biol. 12 (2011) R14.
[55] L.P. Kozlowski, J.M. Bujnicki, BMC Bioinformatics 13 (2012) 111.
[56] T. Puton, L. Kozlowski, I. Tuszynska, K. Rother, J.M. Bujnicki, J. Struct. Biol. 179 (2012) 261–268.
[57] R.R. Walia, C. Caragea, B.A. Lewis, F. Towfic, M. Terribilini, Y. El-Manzalawy, D. Dobbs, V. Honavar, BMC Bioinformatics 13 (2012) 89.
[58] M. Terribilini, J.D. Sander, J.H. Lee, P. Zaback, R.L. Jernigan, V. Honavar, D. Dobbs, Nucleic Acids Res. 35 (2007) W578–584.
[59] O.T. Kim, K. Yura, N. Go, Nucleic Acids Res. 34 (2006) 6450–6460.
[60] L. Perez-Cano, J. Fernandez-Recio, Proteins 78 (2010) 25–35.
[61] H. Zhao, Y. Yang, Y. Zhou, Nucleic Acids Res. 39 (2011) 3017–3025.
[62] S. Vajda, D.R. Hall, D. Kozakov, Proteins (2013), http://dx.doi.org/10.1002/prot.24343.
[63] K. Bastard, A. Saladin, C. Prevost, Int. J. Mol. Sci. 12 (2011) 1316–1333.
[64] M.F. Lensink, S.J. Wodak, Proteins 78 (2010) 3073–3084.
[65] M. Zacharias, Curr. Opin. Struct. Biol. 20 (2010) 180–186.
[66] J. Janin, Mol. BioSyst. 6 (2010) 2351–2362.
[67] Y. Huang, S. Liu, D. Guo, L. Li, Y. Xiao, Scientific reports 3 (2013) 1887.
[68] C. Dominguez, R. Boelens, A.M. Bonvin, J. Am. Chem. Soc. 125 (2003) 1731–1737.
[69] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 2195–2199.
[70] D.W. Ritchie, G.J. Kemp, Proteins 39 (2000) 178–194.
[71] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, Nucleic Acids Res. 33 (2005) W363–367.
[72] H.A. Gabb, R.M. Jackson, M.J. Sternberg, J. Mol. Biol. 272 (1997) 106–120.
[73] E. Karaca, A.S. Melquiond, S.J. de Vries, P.L. Kastritis, A.M. Bonvin, Mol. Cell. Proteomics 9 (2010) 1784–1794.
[74] M.J. Gajda, I. Tuszynska, M. Kaczor, A.Y. Bakulina, J.M. Bujnicki, Bioinformatics 26 (2010) 2986–2987.
[75] I. Tuszynska, J.M. Bujnicki, BMC Bioinformatics 12 (2011) 348.
[76] D. Russel, K. Lasker, B. Webb, J. Velazquez-Muriel, E. Tjioe, D. Schneidman-Duhovny, B. Peterson, A. Sali, PLoS Biol. 10 (2012) e1001244.
[77] J.A. Cruz, M.F. Blanchet, M. Boniecki, J.M. Bujnicki, S.J. Chen, S. Cao, R. Das, F. Ding, N.V. Dokholyan, S.C. Flores, L. Huang, C.A. Lavender, V. Lisi, F. Major, K. Mikolajczak, D.J. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K.M. Weeks, C. Waldsich, M. Wildauer, N.B. Leontis, E. Westhof, RNA 14 (2012) 610–625.
[78] N. Eswar, D. Eramian, B. Webb, M.Y. Shen, A. Sali, Methods Mol. Biol. 426 (2008) 145–159.
[79] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, J. Mol. Biol. 247 (1995) 536–540.
[80] S.C. Agalarov, G. Sridhar Prasad, P.M. Funke, C.D. Stout, J.R. Williamson, Science 288 (2000) 107–113.
[81] D. Matelska, E. Purta, S. Panek, M.J. Boniecki, J.M. Bujnicki, S. Dunin-Horkawicz, RNA (2013), http://dx.doi.org/10.1261/rna.038794.038113.
[82] M.S. Jurica, Curr. Opin. Struct. Biol. 18 (2008) 315–320.
[83] S.C. Flores, J. Bernauer, S. Shin, R. Zhou, X. Huang, Brief. Bioinform. 13 (2012) 395–405.
[84] M.J. Pietal, N. Szostak, K.M. Rother, J.M. Bujnicki, BMC Bioinformatics 13 (2012) 333.