doi:10.1016/j.jmb.2004.11.072 J. Mol. Biol. (2004) xx, 1–16

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Calculation of Standard Atomic Volumes for RNA Cores and Comparison with Proteins: RNA is packed more tightly than Protein

N. R. Voss* and M. B. Gerstein

Molecular Biophysics and Biochemistry, Yale University 260 Whitney Ave, P.O. Box 208114, New Haven, CT 06520 USA

**Corresponding author*

Traditionally, for biomolecular packing calculations research has focused on proteins. Besides proteins, RNA is the other large biomolecule that has tertiary structure interactions and complex packing. No one has yet quantitatively investigated RNA packing nor compared its packing to that of proteins because, until recently, there were no large RNA structures. Here we address this question in detail, using Voronoi volume calculations on a set of high-resolution RNA crystal structures. We do a careful parameterization, taking into account many factors such as atomic radii, crystal packing, structural complexity, solvent, and associated protein to obtain a self-consistent, universal set of volumes that can be applied to both RNA and protein. We report this set of volumes, which we call the NucProt parameter set. Our measured values are consistent across the many different RNA structures and packing environments. However, our volumes are only defined on well-packed atoms, those with sufficient packing neighbors, that typically occur on the interior of RNA molecular and not the unbounded atoms on the surface. When common atom types are compared between proteins and RNA, nine of 12 types show that RNA has a smaller volume and packs more tightly than protein, suggesting that close-packing may be as important for the folding of RNAs as it for proteins. Moreover, calculated partial specific volumes show that RNA bases pack more densely than corresponding aromatic residues from proteins. Finally, we find that RNA bases have similar packing volumes to DNA bases, despite the absence of tertiary contacts in DNA. Programs, parameter sets and raw data are available online at http://geometry.molmovdb.org

Keywords: packing density; RNA volumes; RNA packing

Introduction

Numerous methods have been developed to determine atom radii and volumes for proteins[1–11] and have been applied to DNA.[12] These radii and volumes are necessary in understanding protein structure and particularly for uncovering the relationship between packing and stability. Many studies requiring accurate protein radii and volumes have characterized a number of protein properties including: protein energies,[13] protein–protein interactions,[14] standard residue volumes,[5] internal core packing,[15,16] packing at the water interface,[17,18] protein cavities,[7,8,19] the quality of crystal structures,[20] analysis of volume by amino acid composition,[21,22] macromolecular motions,[23,24] and even measurements of the fit between enzyme and substrate.[25,26] Standard volumes and radii are also important in an indirect sense for the prediction of side-chain packing.[27–29]

Although a standard protein volume set has been available for years[1] and a DNA volume set was produced recently,[12] no attempts have been made to obtain a standard volume set for RNA molecules. This has been primarily due to a lack of RNA structures other than tRNAs and small oligonucleotides, because their crystallization was once thought to be difficult. Within the past decade, it has been shown that RNA structures can be crystallized in the same way as proteins. This has created a new

Abbreviations used: VDW, van der Waals; PSV, partial specific volume; A-DNA, A-form DNA.

E-mail address of the corresponding author: neil.voss@yale.edu

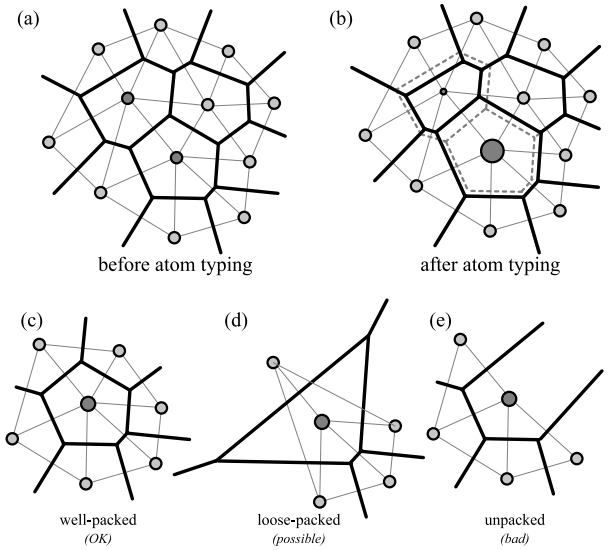

Figure 1. Voronoi constructs and problems. Effect of atom typing on atom volume. (a) Two-dimensional example of the Voronoi construction. Planes are drawn equidistant between any two atoms. The planes are then intersected to get a volume. (b) For atoms of different sizes the planes are no longer placed equidistant between the atoms, but rather as a ratio function of the van der Waals radius of the atoms. So, large atoms are assigned a larger volume and small atoms are assigned a smaller volume. Three major types of Voronoi packing. (c) Well-packed: polyhedron is closed and surface falls under cutoff value. (d) Loose-packed: polyhedron is closed, but due to lack of neighbors the polyhedron has a large surface area above the cutoff value. (e) Unpacked: Voronoi polyhedron is open and no volume can be calculated. Only well-packed are used to determine the volumes of the atoms.

emphasis on solving RNA structures including: ribosomes, self-splicing introns, and many others. Now that there are several structures available, RNA packing can be addressed and analyzed.

To calculate volumes, we employ the traditional Voronoi polyhedra method.[30] In 1908, Voronoi found a way of partitioning all space amongst a collection of points using specially constructed polyhedra. Here we refer to a collection of "atom centers" rather than "points." Bernal & Finney[31] first applied this method to molecular systems and Richards[3] first used it with proteins. The methods used in this work have been previously described by others,[3,31] as well as in our earlier work.[9–11] Figure 1 shows how a Voronoi polyhedron is constructed. This construct partitions space such that all points within a polyhedron are closer to the atom defining the polyhedron than to any other atom. The Voronoi planes are shifted from the original equidistant planes (Figure 1(a)) to the modified set (Figure 1(b)) determined by the relative sizes of the van der Waals (VDW) radii of the atoms, i.e., bigger atoms take up more space in the Voronoi construct than smaller ones. Only atoms whose volumes are well-defined (Figure 1(c)) and not loosely packed (Figure 1(d)) or unpacked (Figure 1(e)) are included.[11] Unpacked and loosely packed atoms usually consist of surface atoms or atoms near cavities and therefore do not have enough neighbors to pack tightly. The Voronoi method provides a good estimate of the true volume of an atom and in turn, reliable, self-consistent values for the comparison of atom volumes. Atoms are assigned VDW radii based on their atom type. The typing follows standard united atom conventions and chemical atom typing. A new technique applied in this study is used to test the contributions of crystal symmetry to surface atoms. Since we are interested in large RNA complexes, and the RNA molecules are typically in close contact within the crystal form, it naturally follows to use this additional packing to our advantage as long as there is no effect on the final numbers.

Using Voronoi polyhedra, we report the standard volumes (and many other statistics) for all RNA atoms (49 in total) and all four nucleotides. These atoms are arranged into 18 atom type volumes and radii based on the chemical structure. Further, the benefit of using crystal symmetry to increase the size of the data set is presented. Crystal symmetry had no effect on the final volumes, but increased the population of atoms in our set. Also, we locate less defined atoms within packed RNA structures, such as the backbone has a low percentage of well-packed polyhedra and is, therefore, less defined. We measure the dependence of the volume of the nucleotides on different RNA structural categories (e.g. tRNA, small rRNA, or ribosomes). The final RNA nucleotide volumes are then compared against DNA and organic molecule measurements. In order to evaluate the role of water, ions and proteins in RNA packing, we remove solvent and protein atoms from the calculations and look at the results. This had no effect on the final volumes, but the number of well-packed RNA atoms decreases significantly. We also compare proteins to RNA by comparing the atom types they have in common. The atom types of protein are found to run slightly larger than those of RNA. In addition, from these volumes, we can calculate the partial specific volume of the RNA nucleotides and find that RNA packs more densely than protein.

Formalism and Results

Atomic radii calculations

Nomenclature

The atomic groups for the RNA atoms are given a nomenclature of the general form "*XnHmS*", where *X* indicates the chemical symbol; *n*, the number of bonds, which, in most cases, is equivalent to saying sp, sp^2, or sp^3 orbitals; *Hm*, the number (*m*) of hydrogen (*H*) atoms attached to the atom where the H does not change and acts a label for the number (*m*); and *S* the subclassification for the atom type,

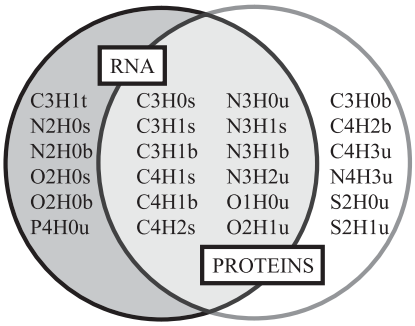

Figure 2. Venn diagram of protein and RNA atom types. Diagramed are all atom types involved in RNA and protein. Types on the right are only involved in RNA while types on the far left are only involved in proteins leaving the central 12 types existing in both RNA and protein.

which is one of the following symbols: *b* (big), *s* (small), *t* (tiny) or *u* (unique). When there are no subclasses for the atom type; *u* (unique) is used. When the atom type needs to be divided into two separate sub-types the type with the larger volume is designated as *b* (big) and the smaller volume *s* (small). In one case (*C3H1*), the atom type requires the addition of a new classification from the previous two subclasses defined previously in proteins.[11] Since the new subclass is smaller than both of the existing *b* (big) and *s* (small) subclasses, its subclass is designated as *t* (tiny). Figure 2 summarizes the 24 different atom types involved in RNA as well as protein and shows which types are common to both.

Voronoi plane positioning method

Voronoi polyhedra were originally developed by Voronoi nearly a century ago.[30] While the Voronoi construction is based on partitioning space amongst a collection of "equal" points, all protein atoms are not equal. Some are clearly larger than others. In 1974, a solution was found to this problem,[3] and since then Voronoi polyhedra have been applied to proteins and DNA. Two principal methods of re-positioning the diving plane have been proposed to make the partition more physically reasonable: method B[3] and the radical plane method.[32] Both methods depend on the radii of the atoms in contact and the distance between the atoms (Figure 1(b)). The simplified method B (or ratio method) divides the plane between the two atoms proportionately according to their covalent radii:

$$d = R + (D - R - r)/2 \quad (1)$$

where d is the distance from the atom to the plane, R, the VDW radius of the atom, r, the VDW of the neighboring atom and D is the distance between the two atoms. This method was accepted for a long time, but it was determined that it had a particular flaw. The flaw is vertex error, where the planes created by neighboring atoms do not perfectly intersect at precise vertices. Vertex becomes a major problem when working with spheres of dramatically different radii. Then the radical plane was introduced which uses a particular quadratic equation to properly divide up the space to obtain precise vertices:

$$d = (D^2 + R^2 - r^2)/2D \quad (2)$$

Because it creates perfect polyhedra, the radical plane method is more pure geometrically than method B. These precise vertices are required for space dividing constructions such as Delaunay triangulations[33] and alpha shapes.[8]

In particular, when comparing the two methods in terms of final volumes there is little difference between the two methods. Even though method B suffers from vertex error, it has been shown to be quite robust for protein calculations, even more robust than the radical plane.[10] In particular, there are two main issues where the methods differ: vertex error and self-consistency. For arbitrary systems with radii of significantly different values, vertex becomes a major issue and the method B is no longer a reasonable approach. However, the radii of proteins atoms do not differ that much and it has been shown that vertex error accounts for one part in 500.[17] In addition, method B has shown to give more self-consistent volumes. It was revealed that the radical plane method actually results in a higher standard deviation than method B, suggesting that it places the plane in a less consistent manner.[10] Further, method B is has been thoroughly tested over the years, while the radical plane is a more recent approach. In addition, the current standard volume set in proteins uses method B for its calculations, so to make direct comparison we will need to have an RNA volume set under the same methodology.

While method B suffers from vertex error, it was reported that this only accounts for one part in of the total volume primarily due to having radii.[17] There are also a two caveats associated with the radical plane method. First, all prior Voronoi research in proteins is based on the method B technique, therefore using radical planes for RNA makes it difficult to draw parallels to protein. Second, volumes calculated by the radical plane result in overall higher standard deviations.[10] Furthermore, in this study, the average standard deviation of the atom types rises from 1.24 to 1.32 for the radical plane method.

Despite this, we report the base volumes for both methods (with little difference) but we use the more traditional method B in all figures, comparisons to protein and radii refinement. The raw data sets and histograms for both methods are also available on the web.

Importance of atom typing

Described in more detail by Tsai *et al.*,[11] the distance between the atoms and their intersecting

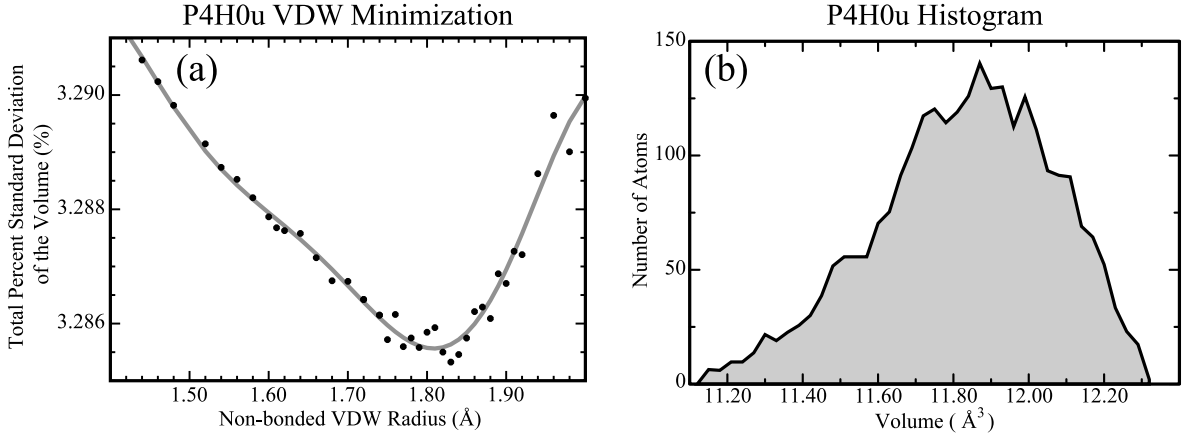

Figure 3. Determining non-bonded VDW radius for the unassigned P4H0u atoms. (a) The normalized standard deviation for the P4H0 atom *versus* its VDW radius. The minimum is found to be 1.82 Å. These values are used for the final Voronoi volume calculations. (b) Histogram of the P4H0u atoms from our final NucProt data set showing one distinct peak.

planes used for Voronoi volume calculation depends on the VDW radius of the atom type. Due to this dependence on atom radius, it becomes increasingly important to obtain accurate atomic classifications and radii. Work done earlier studied the affect of varying the number of atomic classifications and came to the determination that the atom typing system described by *XnHmS* nomenclature was the best balance between over and under fitting for accurate measurements of the volumes for the atoms.[11]

VDW radii taken from protein set

The VDW radii for several of the atomic groups involved in RNA structures have analogous atoms in proteins. Several papers have been published on the VDW radius of protein atoms.[5,9] For these overlapping groups, the non-bonded VDW radii of RNA atom groups are simply transferred from their corresponding protein atom groups using the radii defined by the ProtOr set.[9] Whenever there is a small or big designation for the group, the atom group is compared by volume to the protein atoms, e.g., guanine *N1*, of chemical type *N3H1*, is more similar in volume to *N3H1s* than *N3H1b*. Despite vast differences, RNA structure contains only three new atomic groups that completely lack a protein analog, namely *O2H0*, *N2H0* and *P4H0*. *N2H0*, though a new type, is found to be very similar to *N3H1*. Assignment of all the RNA atoms to groups is for the most part straightforward; the only complication came from assignment of the *N2H0* nitrogen atoms and two remaining missing types.

Adjusting the bonded VDW radii

Because this new NucProt data set is to include RNA and protein, an investigation into the bonded radii is undertaken to make the values more accurate for both protein and RNA. Using the defined bond length from CNS,[34] the bond radii are varied for each atom type (grouping small and big subtypes into one type) in order to minimize the sum over all squared bond differences (the bond length − the VDW radius of both atom types bonded) in RNA and protein together. These new bonded radii are not significantly different from the previously published radii,[11] but give a better account of the atom types. For example, the *O1H0u* bonded radius drops the most (from 0.66 to 0.52) reflecting a smaller oxygen atom size due to its double-bonded character, whereas ten of the 24 types change by less than 0.01 Å. These newly adjusted bonded radii should provide a more self-consistent volume data set.

New atom types for RNA

Next we need to determine the modal behavior for the unassigned atoms, i.e., do they require small and big subgroups or are they a unique type. *P4H0* only contains one RNA atom and cannot be subdivided further unless the phosphorus atom attached to a guanosine is packed differently than a uridine phosphorus, which is not the case. Therefore, *P4H0* is given a *P4H0u* designation. Further, the *P4H0u* atom type produces a tight histogram (Figure 3(b)), confirming its behavior as a unimodal distribution. The *O2H0* and *N2H0* atom types contain three atoms and six atoms, respectively, of which neither follows a simple distribution. The *O2H0* atom consists of the 3′, 4′ and 5′ sugar oxygen atoms. Individual volume calculations show *O4′* is significantly smaller than both of its type-equivalents *O3′* and *O5′*. The histogram of the *O2H0* atoms (Figure 4(f)) shows a bimodal distribution confirming this assessment. Hence, we design two species of *O2H0* atom: a big class, *O2H0b* (*O3′* and *O5′*), and a small class, *O2H0s* (*O4′*). The *N2H0* atom type is more complex and contains six different types: *ADE-N1*, *ADE-N3*, *ADE-N7*, *GUA-N3*, *GUA-N7*, and *CYT-N3*. From Figure 4(d), the *N3* and *N7* atoms from both purines are grouped into a large

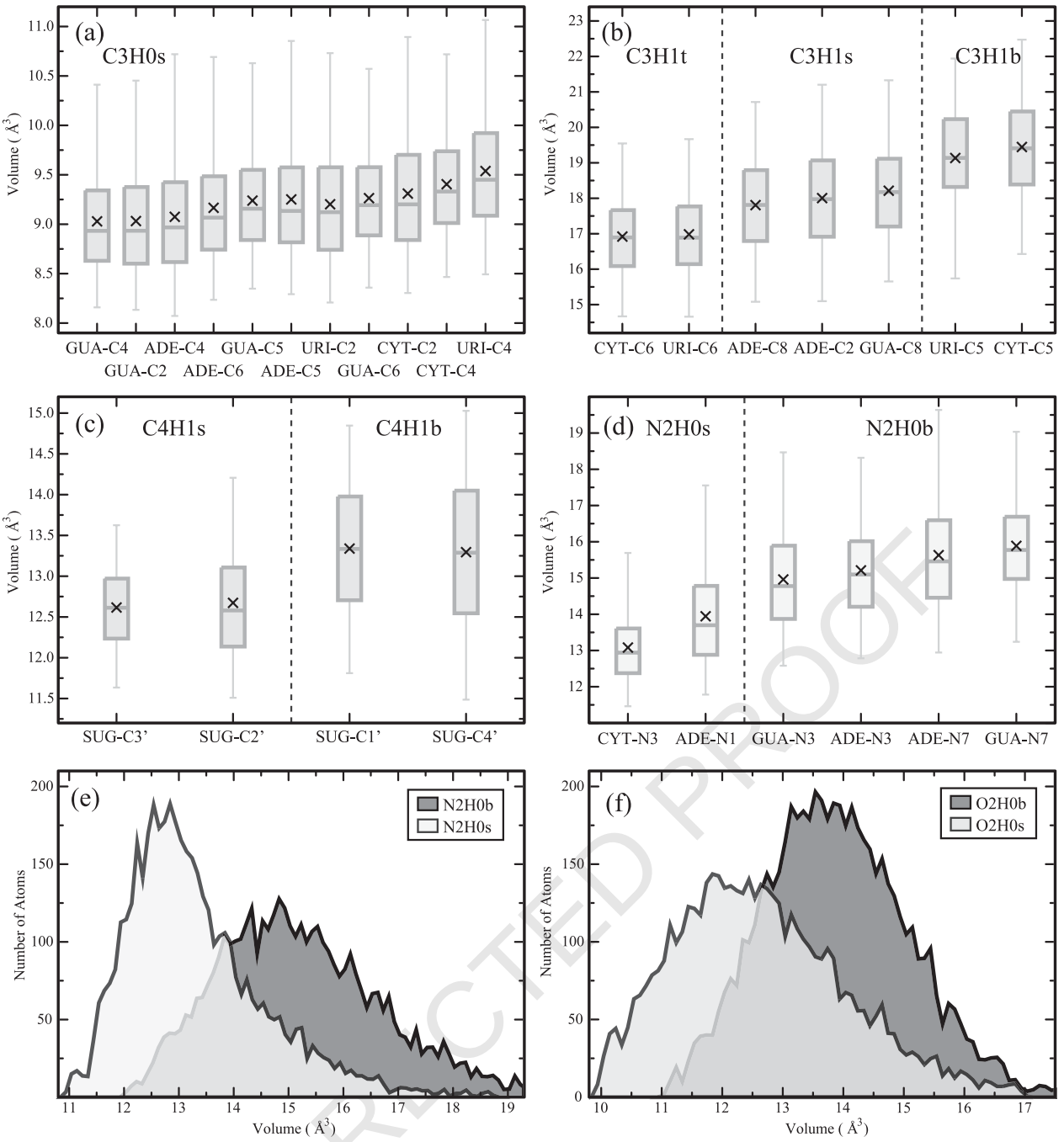

Figure 4. Distributions of atom type volumes. (a) Distribution of all atoms composing the *C3H0* group showing one distinct volume. (b) Distribution of all atoms composing the *C3H1* group, suggesting three distinct groups: tiny, small and big. (c) Distribution of all atoms composing the *C4H1* group, suggesting two distinct groups: small and big. (d) Distribution of all atoms composing the *N2H0* group, suggesting two distinct groups: small and big. (e) Volume distribution of the two *N2H0* groups, small and big. (f) Volume distribution of the two types of *O2H0* atoms, small and big.

set and while the *CYT-N3* is significantly smaller than all of the other atoms, it is grouped with the only slightly larger *ADE-N1*. After grouping, a good separation between the small and big subgroups is found (Figure 4(e)).

Determining non-bonded VDW radius of new types

For the unassigned atom groups (*O2H0*, *N2H0*, and *P4H0*, from above), a non-bonded VDW radius needs to be determined. The bonded VDW radii were assigned when the bonded radii were adjusted for all the atom types. All nitrogen-containing atom groups (*N3Hx*, *N4Hx*) in the ProtOr set[10] for proteins are defined as having the same bonded and non-bonded atomic radius, so we felt the *N2H0* should have the same values as its sister atom types because its volume is the same as *N3H1* types. The non-bonded VDW radii for the *P4H0* and *O2H0* do not have existing values and so

their non-bonded VDW radii are determined by varying the non-bonded VDW radius of the atom in question and minimizing the sum of the percent standard deviation of volume (standard deviation of the volume divided by the mean volume) over each atom in RNA. The standard deviation gives an unfair bias to minimizing the error of atoms with larger volumes due to their larger deviations, so by taking the standard deviation divided by the mean this bias is reduced. This method for calculating the missing non-bonded VDW radii results in the most self-consistent set of volumes.

As shown in Figure 3(a), the standard deviation of the volume for phosphorus atom, *P4H0* volume gave a convex curve when its radius was varied. The curve is then fit to a tenth-degree polynomial (only to smooth out the noise without loss of generality) and the *P4H0u* radius is taken to be the minimum of the polynomial fit, which is 1.82 Å. This final value gave an extremely tight unimodal distribution (Figure 3(b)). Likewise, a two-dimensional optimization is employed for the *O2H0* types due to its bimodal distribution (Figure 4(e)), by simultaneously varying the radius of both subtypes. The global minimum of the percent standard deviation is determined to be 1.50 Å for the *O2H0s* and 1.62 Å for *O2H0b* (data not shown).

Determination of volumes

Brief description of Voronoi method

The volumes of the atoms are determined with the same Voronoi method as published earlier.[9–11,17] For every pair of atoms, a plane is constructed approximately equidistant from both of the atoms (in actuality the distance is adjusted by the VDW radii of each atom type) and orthogonal to the bond between the atoms (Figure 1). The planes are then intersected, leaving an enclosed polyhedron for each atom. While not every atom has a closed polyhedron, the majority of the polyhedra is closed.

Assembling the structure set

Structures were obtained from the NDB[35] by searching for nucleic acids structures containing RNA, with strand lengths greater than 26 nt, to avoid small synthesized RNA molecules, and resolution better than 5 Å. The cutoff value of 5 Å was chosen to include all four ribosomal subunit structures, including both low and high-resolution versions. We compare the high-resolution only data to the entire set and find no difference in the final volumes. After determining our criteria, the search results in a raw structure set of 125 RNA structures. Most of the found structures are redundant (e.g., 50 S ribosomal subunits soaked with various complexes or tRNA with and without synthetases) and some structures contained DNA base-paired with RNA in a complex. After removing the DNA hybrid structures and duplicates, taking care to use the most accurate and detailed structure within each redundant set, a final set is created consisting of 50 unique structures. For comparison purposes, the final sets are broken down into five smaller disjoint

Table 1. Summary of structure sets

Set name	Number of PDB files	Number of RNA atoms	%OK no symm	%OK with symm	% of total atoms	% of total "OK" atoms	PDB Ids
Disjoint subsets							
SRP	6	10,137	35.0	37.9	3.7	3.6	1hq1, 1jid, 1lng, 1mfq, 1e8o, 1l9a
Small ribo	13	19,234	36.6	40.3	6.9	7.3	483d, 1msy, 1jbs, 1i6u, 1mms, 1mzp, 1mji, 1dk1, 1g1x, 1qa6, 430d, 364d, 357d
tRNA	14	27,379	33.2	34.8	9.9	9.0	1f7u, 1ehz, 1fir, 1qf6, 1il2, 1h4s, 1b23, 1qtq, 1ser, 1ffy, 1i9v, 1ttt, 1ivs, 2fmt
Small RNA	13	31,438	33.7	37.2	11.3	11.0	1l2x, 437d, 1et4, 1m5o, 1hr2, 1duh, 1cx0, 1kxk, 1f1t, 1l3d, 1hmh, 1kh6, 1nbs
Ribosomes	4	188,911	38.9	39.0	68.2	69.2	1jj2, 1i94, 1n32, 1nkw
All	50	277,099	37.4	38.4	100.0	100.0	1b23, 1cx0, 1dk1, 1duh, 1e8o, 1ehz, 1et4, 1f1t, 1f7u, 1ffy, 1fir, 1g1x, 1h4s, 1hmh, 1hq1, 1hr2, 1i6u, 1i94, 1i9v, 1il2, 1ivs, 1jbs, 1jid, 1jj2, 1kh6, 1kxk, 1l2x, 1l3d, 1l9a, 1lng, 1m5o, 1mfq, 1mji, 1mms, 1msy, 1mzp, 1n32, 1nbs, 1nkw, 1qa6, 1qf6, 1qtq, 1ser, 1ttt, 2fmt, 357d, 364d, 430d, 437d, 483d
Additional sets							
hi-res	9	11,281	49.4	58.7	4.1	6.2	1jbs, 1msy, 483d, 1et4, 437d, 1l2x, 1jid, 1hq1, 1f7u, 1ehz
RNA only	19	33,782	33.7	38.6	12.2	12.2	1ehz, 1fir, 1i9v, 1l2x, 437d, 1et4, 1hr2, 1duh, 1kxk, 1f1t, 1l3d, 1hmh, 1kh6, 1nbs, 483d, 1msy, 430d, 364d, 357d

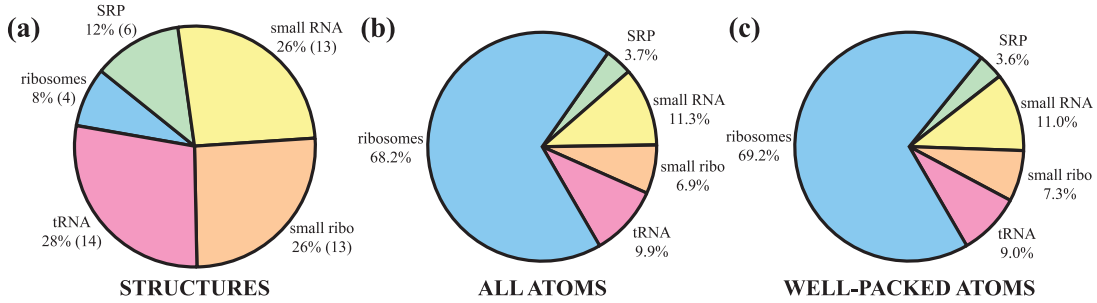

Figure 5. Distribution of atoms within PDB set. Pie charts showing how the structure set breaks down into the five major categories. (a) Number of structures for each subset. (b) Number of atoms for each subset. (c) Number of sufficiently packed or "OK" atoms for each subset. Though the four ribosomal structures account for only 8% of the structures of the pdb set, they account for 69.2% of the atoms used in the final calculations. In the text it is shown that ribosomal and non-ribosomal RNA have the same final base volumes.

subsets: (i) *SRP*—RNA structures involved with the Signal Recognition Particle, (ii) *small-ribo*—small ribosomal RNA fragments, such the 5 S rRNA structures, (iii) *tRNA*—transfer RNA with and without synthetases, (iv) *small-RNA*—the other remaining small RNA molecules including ribozymes and self-splicing introns, and (v) *ribosomes*—complete ribosomal subunits (Table 1). These structure sets are unfortunately heavily weighted towards ribosomal data (69% of atoms), because of their immense size, despite only being only four of the 50 structures in the set (Figure 5, Table 1). This effect will be addressed later.

Generation of final volume set

Surface atoms (as well as loosely packed interior atoms) sometimes lack closed polyhedra or have extended polyhedra and give rise to indeterminate or inflated volumes, respectively. These two special cases of atoms need to be removed from the set of atom volumes in order to obtain a self-consistent data set. The first case of loosely packed atoms occurs when the Voronoi shell is heavily extended (Figure 1(d)). Loosely packed atoms are distinguished from well-packed atoms by their surface area. Atoms above a certain surface area cutoff are characterized as "possible," meaning that they have a volume, but it is unsure whether it is relevant. The loosely packed atoms are not used in our final NucProt data set due to their indefinite character. The second case, an atom having insufficient neighbors, leaves the Voronoi shell open ended and produces an indeterminate volume (Figure 1(e)). These unclosed polyhedra are easy to identify for they have no volume and are designated "bad" by the software.[11] Consequently, only atoms with closed polyhedra and small surface areas are then labeled as "ok" atoms. In should also be noted that all atoms (protein, RNA, ions, water, organic molecules and also modified nucleotides and amino acid residues) within the PDB file are taken into account for Voronoi plane positioning. Unfortunately, modified bases volumes are not reported due their small population within our set, thus making it almost impossible to provide any reasonable statistics. There are only 29 pseudouridines within our PDB set and given that at most half the atoms are well-packed, it would be a very unreliable volume for general use.

Despite applying these standard methods for generating the final volume set, extreme atom volumes existed for each RNA atom. Therefore, as an additional measure, the extreme atom volumes are removed from the ends of each RNA atom distribution such that the average range for each distribution drops in half. Dropping the distribution range in half is chosen, because it provides

Table 2. Summary of effect of other atoms on the packing calculations

	50 S Ribosomal subunit (1jj2)				
Crystal symm	+	−	−	−	−
Protein	+	+	−	+	−
Ions/water	+	+	+	−	−
Base volumes (Å^3)					
GUA	145.9	145.9	145.7	146.4	146.2
ADE	140.0	140.0	139.4	138.9	138.2
CYT	115.5	115.5	115.3	115.6	115.0
URI	110.8	110.8	110.6	110.9	110.2
SUG	176.1	176.1	175.4	179.2	177.4
Addition information					
Count	33,245	33,204	28,996	23,007	20,409
%OK	54.0	53.9	47.1	37.3	33.1
%Closed	93.5	93.1	91.2	90.7	84.9
Mean %SD	6.89	6.89	6.78	7.10	6.55

Table 3. Base volumes across several different structure sets

	Standard Disjoint Sets					*Additional Sets*			ALL SETS	radical	*Other Published Sets*		
Base Volume (Å^3)	SRP	ribosome	small ribo	small RNA	tRNA	hi-res	allrna	nosymm	complete	plane	organic[a]	A-DNA[c]	B-DNA[c]
GUA	141.3	146.4	144.0	144.7	142.7	144.2	144.0	146.1	145.9	144.5	157.7	145.4	143.8
ADE	134.1	139.7	136.4	137.9	136.2	138.7	137.3	139.4	139.2	137.8	148.8	138.3	136.1
CYT	110.2	115.6	111.8	113.4	111.6	113.2	113.1	115.2	115.0	113.2	122.1	113.9	113.2
URI	102.2	111.3	107.4	106.1	107.9	106.6	106.9	111.0	110.8	109.4	119.2	132.9[d]	132.6[d]
Backbone Vol (Å^3)													
SUG	172.9	176.0	175.8	175.2	174.5	176.8	175.7	175.9	176.1	179.3	133.6[b]	181.8	174.8

[a]Values converted from the work done by Lee & Chalikian.[62] Volumes require conversion of units from cm/mol to Å^3/residue. [b]This value is from a nucleoside, not a nucleotide, and lacks a phosphate group with a volume of approximately 43 Å^3 (depending on oxidation state). Calculated sugar volume is averaged over three base volumes subtracted from the nucleoside volumes. [c]Values taken from the work done by Nadassy *et al.*[12] [d]Thymine values are used in place of uracil. Thymine should be approximately 27 Å^3 greater in volume than uracil, based on atom type volumes in thymine.

the best balance between data loss and reduction of the range. This method is very effective because to drop the range in half only 1.25% of the data is removed from the set. In essence, the central 97.5% of the data has half the range of the complete set of data, highlighting some of the extreme values resulting from over-packed atoms due to structural overlap errors or loosely packed atoms missed by the surface area cutoff. After applying these methods, only well-packed atoms are then used in final volume calculations for self-consistency, making it important to maximize the number of well-packed atoms for a good sample size.

Effect of surface molecules

The treatment of surface atoms plays an important role in calculating Voronoi volumes because Voronoi volumes rely on neighboring atoms to create polyhedra surrounding each atom. By increasing the number of neighboring atoms it is possible to have more well-packed atoms. To explore these problematic surface atoms, the effects of crystal symmetry, bound proteins, and solvent atoms on structure are examined for their significance. All three factors had little effect on the final volumes, but all make a significant contribution to the number of observations for each atom (Table 2). In the final volume set, atoms from both high and low-resolution structures, both protein containing and protein free RNA structures, and only crystal symmetry generated structures are integrated into our final NucProt data set.

Final volumes

We now can provide final volumes. Since we are actually calculating distributions of volumes, i.e. the probability of a volume given an atom type, we provide both histograms and mean values. In particular, we show a sample distributions in Figures 4 and 8†. It is also useful to have explicit mean values for the volumes. The final volumes for the RNA bases are 145.9 Å^3 for guanine, 139.2 Å^3 for adenine, 115.0 Å^3 for cytosine, and 110.8 Å^3 for uracil (Table 3). All four RNA sugar backbones are approximately the same size and so we report only one value of 176.1 Å^3 (Table 3). The nucleotide volumes are 322.6 Å^3 for guanosine, 315.0 Å^3 for adenosine, 290.7 Å^3 for cytosine, and 285.5 Å^3 for uridine.

† The rest is available on http://geometry.molmovdb.org/NucProt

Discussion

Effects on calculations

RNA backbone and base packing

From the standard deviations and packing percentages, we are able to locate areas within RNA structure that are not well-packed or less defined. In our NucProt data set, only 26.8% of the sugar–phosphate backbone atoms are packed sufficiently to make a volume measurement, which is 20.2% less than the worst base, uracil (at 47.0% well-packed). Further, several backbone atoms have low percentages of well-packed Voronoi polyhedra (Figure 6(b)) and high standard deviations (Figure 6(d)). These results suggest that atoms located in the bases benefit from the tight ring structure of purines and pyrimidines providing inherent packing neighbors as well as base-pairing and base-stacking interactions common in RNA structure. In addition, the atoms located in major groove edge of the RNA bases also have high standard deviations, high packing densities and low percentages of well-packed atoms (Figure 6). This presumably implies a less packed major groove in RNA structures, but it is more likely due to no inherent neighbors in an A-form helix. Therefore, our results suggest that the backbone-sugar regions and major groove atoms are less packed than the interior base atoms.

Role of crystal symmetry in volume size

Small RNA structures, in general, consist of a single helix and hence lack helix-helix packing, resulting in poorly packed backbones. In an effort to

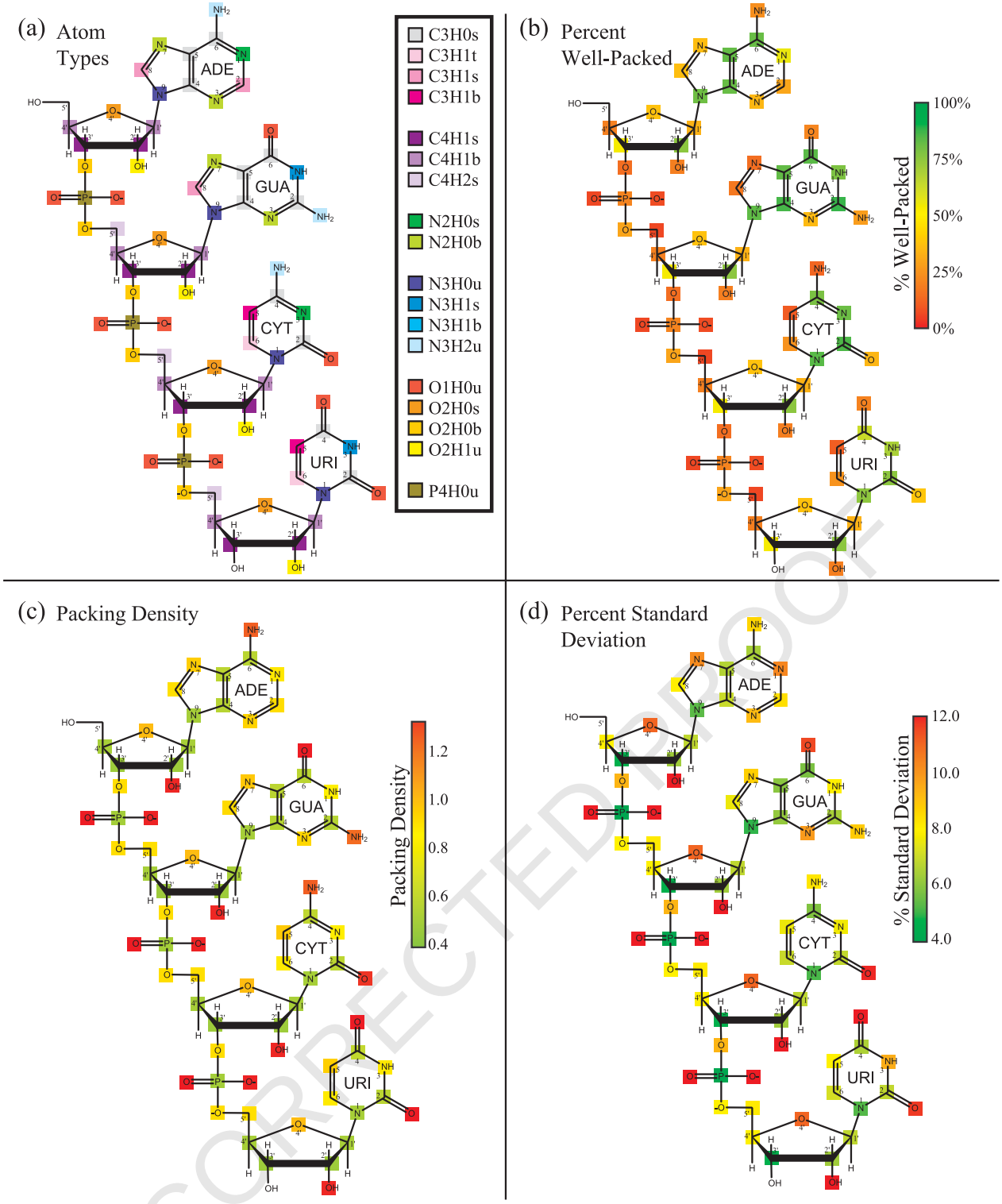

Figure 6. Graphical display of different atomic packing measurements. For each atom in RNA. (a) The atom typing is shown to show what decisions are made to classify the various atoms. (b) The percent well-packed atoms shows that the backbone and extensions off the rings are in general less defined. (c) The packing density (Voronoi volume divided by VDW volume) measures the how tightly each atoms packs. This number tends to be biased by the number of hydrogen atoms bonded to an atom, but still provides insight as another measures of packing. (d) The percent standard deviation of the volume (standard deviation of the volume divided by the mean volume) highlights the unbiased error involved in the volume measurements.

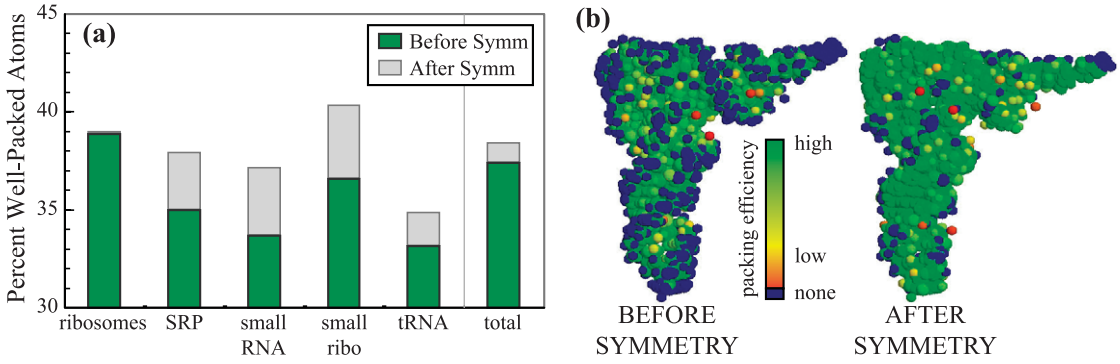

Figure 7. Effects of crystal symmetry. (a) Effect of the crystal symmetry on each subset of structures. Ribosomes saw little to no effect, while small RNA molecules, including ribozymes and other small RNAs, see a large jump in their percentage of well-packed atoms. (b) Example of one structure (1ehz), which is by no means the best, where crystal packing helps increase the number of well-packed atoms. Shown is the packing efficiency, i.e., the Voronoi volume of an individual atom divided by the mean volume for the atom. Blue represents atoms that have unclosed Voronoi polyhedra. Packing before crystal symmetry is shown on the left and after is on the right. You can see the dramatic effect of crystal symmetry on obtaining information for surface atoms. The final volumes show no difference between data sets with and without crystal symmetry.

prevent this single helix dilemma, we need to utilize the crystal symmetry contained in the PDB file. Crystal symmetry neighbors have additional relevant packing interactions from their presence within the crystal. Alas, any software found for generating crystal symmetric neighbors is not applicable or does not work well for our purpose. All are incapable of outputting the information to a file or do not have the facility to generate all symmetry neighbors within a given distance of the target structure. Fortunately, matrix information on crystal symmetric rotations and translations are contained within most PDB headers (as well as in the online PDB format description, Appendix 1[36]) and once recognized, it was simple to implement a small script to achieve the additional symmetry neighbors.

As shown in Table 2, crystal symmetry had little to no effect on the final volumes, but does contribute a significantly larger number of acceptable atoms for making calculations (Figure 7). The different disjoint subcategories of structures show that even though the set of ribosomes have little to no effect on the number of well-packed atoms, all other sets containing the smaller structures increase the percentage of well-packed atoms by a dramatic amount. The additional number of well-packed atoms created from the symmetry neighbors not only gives us more data for error analysis, but helps increase the amount of information from atoms involved in backbone packing that would normally have unclosed polyhedra.

Roles of different RNA structural categories

As noted, our data set consists of 69% ribosomal atoms. *A priori* it is unjustified to assume short double-stranded RNAs have the same packing properties as large macromolecular complexes such as the ribosome. The ribosome is large enough in all three dimensions to truly have an interior while short double-stranded RNA is completely exposed to solvent. Further, we want to confirm that our data set, containing 69% ribosomal atoms, is representative of small RNA molecules as well. There is little difference in nucleotide volume (Table 3) or atom size (data not shown) among the different RNA types. Table 3 clearly shows that the base size, sugar backbone, and entire nucleotide volumes differ by less than 9 $Å^3$ from the smallest to largest values, which is within the standard deviation. Ribosomal RNA volumes run slightly larger than the other structural categories. This could be due to the larger size of the complexes and their inherent problem of packing helices against other helices. Despite slight variations between the structural categories, the final outcome of the volume calculations suggests that all RNA packs in a universal way.

Role of water, ions and proteins in RNA structures

To test the role water, ion, and protein atoms play in RNA structures, we took the largest RNA structure (the refined *Haloarcula marismortui* 50 S ribosomal subunit, 1jj2) and conduct packing tests by systematically removing each kind of atom (Table 3). Crystal symmetry plays a small role in the 50 S subunit, because there is more interior than surface. On the other hand, when the solvent is removed (leaving the RNA and protein), the percentage of well-packed atoms differs by 16.7% from the original value. Similarly, when the protein is removed (leaving the RNA and solvent), the percentage differs by 6.9%. Further, when only the RNA is used the final percentage differs by 20.9%. The final difference of 20.9% is very close to the sum of the other differences of 23.9%, suggesting an independence of the two atom classes. In addition, the solvent has a much larger affect on the loss of well-packed atoms than does the protein. This indicates that RNA atoms pack tightly against the

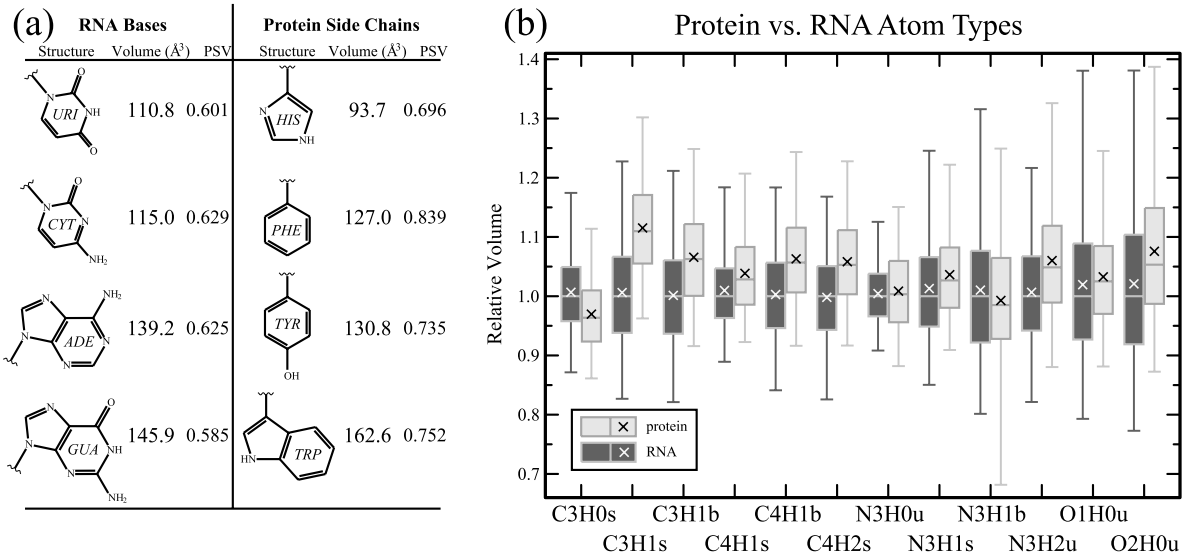

Figure 8. Comparison of RNA volumes to protein. (a) The comparison of aromatic protein residues to RNA bases. On average RNA has a smaller PSV than protein, suggesting RNA packs more densely than protein. (b) Comparing protein to RNA atoms using relative volume. Relative volume is volume divided by the median RNA volume for that atom type. All 12 atoms in the intersection of the protein and RNA atom types are shown for comparison.

solvent. Despite these major differences in the percentages of well-packed atoms, there is no change in the final RNA volumes from the removal of solvent and protein atoms, reinforcing the idea that our set is self-consistent.

Comparison to proteins and DNA

Partial specific volumes

To address the effect of RNA and protein in packing, the partial specific volume (PSV) is computed by taking the calculated volume divided by the atomic mass of the RNA molecule and then changing the units of cubic Ångströms per Dalton ($Å^3$/Da) to the classical form of milliliters per gram (ml/g) (using conversion factor of 1 $Å^3$/Da$=N_A\times 10^{-24}=0.6022$ ml/g). We now provide a new online tool for calculating volumes and PSV for any sequence†.

RNA is found to have an average calculated PSV of 0.569 ml/g, which is significantly more dense than the published protein calculated average of 0.728 ml/g over 13 protein structures.[5] RNA bases are more loosely packed than their complete nucleotide form with an average calculated PSV of 0.610 ml/g and it follows that the sugar and phosphate backbone is more tightly packed with an average calculated PSV of 0.544 ml/g. These values compare well to the published experimental value of 0.540 ml/g.[37] For proteins, it is shown that calculated values are on average 0.5% less than experimental values.[5] Our average PSV is about 5.4% greater than this experimental value, but Durchschlag explains that the experimental RNA PSVs depended heavily on solvent content and the values were difficult to obtain and also may fluctuate greatly.[37] Our results thus conclude that the calculated PSV using Voronoi volumes for RNA is a good estimate for the experimental PSV.

Atom types

There are 18 atom types each in RNA and in protein, but when intersected they only share 12 common atom types. Figure 8(b) analyzes the 12 common types in more detail. Though proteins and RNA for the most part share similar distributions, most protein atom types (nine of 12) run slightly larger (Table 4, Figure 8(b)). The three exceptions to this rule are *C3H0s*, *N3H1b* (which run smaller) and *N3H0u* is almost exactly the same size. The most dramatic effect is shown by *C3H1s*, which consists of aromatic ring carbon atoms in protein and purine ring carbon atoms in RNA (Table 4). For *C3H1s*, the 25th percentile of the protein is greater than the 75th percentile of the RNA, suggesting two distinctly different values (Figure 8(b)). One of the reasons why RNA atom types are smaller in volume than equivalent protein types, is their built-in chemical structure. Proteins are chains that have few atoms per residue and pack against one another to achieve tight packing, while RNA contains more than 18 atoms per residue and, therefore has inherent neighboring packing interactions. In fact, the worst packed atoms in RNA are either attached to the phosphorus atom or an extension off the ring structure (e.g. sugar O2′, guanine O6, purine O2). In essence, the atom type data shows that RNA is more tightly packed than protein atoms.

Similar protein residues

An interesting question to ask is how does the volume of RNA compare to protein amino acid residues of similar chemical structure. Namely, how

Table 4. Summary of atom types in proteins and RNA

Atom type	RNA NucProt Set						Protein NucProt Set (ProtOr)						Comparison, %vol change
	Number	%OK	Count	Mean	Standard deviation	%SD	Number	%OK	Count	Mean	Standard deviation	%SD	
C3H0s	11	81.6	30,479	9.21	0.60	6.5	20	76.1	12,097	8.77	0.63	7.2	-5.0
C3H0b							13	49.4	4418	9.77	0.77	7.9	–
C3H1t	2	24.7	1387	16.95	1.19	7.0							–
C3H1s	3	26.7	2766	17.98	1.44	8.0	8	43.5	1888	20.59	1.81	8.8	12.7
C3H1b	2	10.4	586	19.32	1.51	7.8	8	55.3	2181	21.37	1.90	8.9	9.6
C4H1s	2	64.1	16,554	12.65	0.72	5.7	18	53.2	7227	13.26	1.01	7.6	4.6
C4H1b	2	26.3	6795	13.32	0.97	7.3	6	54.8	3747	14.44	1.33	9.2	7.7
C4H2s	1	6.1	790	21.74	1.77	8.1	20	25.6	4468	23.45	2.34	10.0	7.3
C4H2b							7	27.3	1137	24.42	2.14	8.8	–
C4H3u							9	37.5	3673	36.92	3.25	8.8	–
N2H0s	2	68.7	4461	13.41	1.26	9.4							–
N2H0b	4	29.4	4285	15.33	1.51	9.8							–
N3H0u	4	81.0	10,465	8.79	0.45	5.1	1	74.1	592	8.82	0.66	7.5	0.4
N3H1s	2	74.8	4811	13.63	1.17	8.6	20	62.5	10,356	13.82	1.20	8.7	1.4
N3H1b	4	29.8	4352	15.43	1.67	10.8	4	29.0	500	15.87	2.21	13.9	2.7
N3H2u	3	21.7	2322	22.10	1.94	8.8	4	9.7	286	23.38	2.77	11.8	5.5
N4H3u							1	1.2	12	21.21	1.85	8.7	–
O1H0u	6	13.4	5052	16.29	2.09	12.8	27	36.1	8273	16.17	1.59	9.8	-0.7
O2H0s	1	37.5	4850	12.73	1.40	11.0							–
O2H0b	2	22.3	5753	13.98	1.18	8.4							–
O2H1u	1	19.1	2468	17.39	2.28	13.1	3	20.0	619	18.60	2.45	13.2	6.5
P4H0u	1	20.5	2645	11.86	0.23	1.9							–
S2H0u							2	50.1	280	29.17	2.81	9.6	–
S2H1u							1	51.6	63	34.60	5.73	16.6	–

does the volume of an RNA purine compare to tryptophan and how does an RNA pyrimidine compare to phenylalanine, histidine, and tyrosine. Figure 8(a) reports the protein volumes for the side-chains (calculated from the residue volume subtract the volume of glycine) of tryptophan, tyrosine, phenylalanine, and histidine.[17] While these values for the volume are all relatively close to the RNA base volumes (Table 3), the PSV tells a different story (Figure 8(a)). The average PSV for the RNA bases is 0.609 ml/g while the four protein side-chains have an average PSV of 0.755 ml/g, suggesting that the RNA bases are much more dense than protein aromatic side-chains (Figure 8(a)). Further, if we divide the total volume by the number of atoms (including the hydrogen atoms), we get an average volume of 9.83 Å^3 per atom for the RNA bases and 11.25 Å^3 per atom for the protein residues. Though the volume per atom numbers are biased by the atom type volumes, this also highlights that RNA seems to pack more tightly than protein. These results may be due to nucleotide base rings containing more nitrogen atoms than the amino acid aromatic rings. In addition, the RNA rings have more atoms attached to them, creating a large number of inherent neighbors. In addition, RNA duplex base stacking may contribute favorably to achieve this tighter packing. Though it is difficult to directly compare these vastly different chemical structures, we find that the RNA bases are more tightly packed than the aromatic protein residues.

DNA volumes

In 2001, Nadasssy *et al.* published the standard atomic volumes of double-stranded DNA.[12] Comparing the volume of RNA in large macromolecular structures to that of A-form DNA (A-DNA) we see a small deviation (Table 3). RNA bases: adenine, guanine and cytosine are larger by only 0.9 Å^3, 0.5 Å^3, and 1.1 Å^3, respectively, to that of A-DNA. Since we cannot directly compare uracil to DNA, we compare its volume to the volume for thymidine and they are within the expected difference of 27 Å^3, due to the extra methyl group. We found that in RNA structures about half of the base volumes are within a standard deviation of the A-DNA base volumes (Table 3), suggesting similar packing of the bases. The sugar–phosphate backbone on the other hand reports a slightly larger difference. The A-DNA sugar plus phosphate reported by Nadasssy *et al.* is 5.7 Å^3 larger than the RNA backbone reported here (Table 3). Though one should expect the backbone atoms of A-DNA to be 8.3 Å^3 smaller (based on our atom type volumes) due to the additional volume taken up by the 2′ oxygen, this is not the case. Furthermore, in DNA the 2′-carbon volume is reported as 18.0 Å^3, while we report a volume of 12.67 Å^3; this drop is expected because of the loss of the hydrogen. But if we take the 2′-oxygen volume of 17.39 Å^3 into account, we now have a total volume of 30.07 Å^3 to fit into the space of 18.0 Å^3. RNA structure must accommodate for this additional occupied space. In summary, the published A-DNA volumes are approximately equal for the bases and differ slightly for the backbone where A-DNA is packed less tight than RNA.

Implications in RNA packing

Early results for proteins showed protein interiors are more tightly packed than amino acid crystals.[3] These results also indicated that tight packing and detailed interactions are important in protein folding. RNA tends to be seen as a loosely packed molecule, held together primarily by base-pairing and electrostatic interactions through backbone alterations and metal ion coordination. This is borne out by a survey of a number of prominent papers in RNA structure and folding.[38–55] These papers mention electrostatics and hydrophobic effects as important factors in RNA folding, but none of them mention the importance close packing. For instance, Doudna & Doherty argue that the hydrophobic effect, hydrogen bonding, metal ion coordination and VDW forces all contribute to the formation of compact structures.[51] They say that hydrophobic effects in RNA occur mainly at the level of secondary structure, making a contribution to vertical base stacking. Additionally, they assert that RNA folding is opposed by electrostatic repulsion from the negatively charged phosphate backbone.

However, our results show, surprisingly, that RNA is actually packed more tightly than proteins. In essence, we demonstrate that close packing is as important for RNA folding as for proteins. This suggests a number of interesting energetic calculations that might be worthwhile doing. To emphasize this point, we have modified the text as shown below and changed the title to: "Calculation of Standard Atomic Volumes for RNA Cores and Comparison with Proteins: RNA is packed more tightly than protein".

Another interesting aspect of RNA packing illuminated by our volume calculations concerns the 2′-carbon atom. In DNA the 2′-carbon volume is reported as 18.0 Å^3,[12] while we report a volume of 12.67 Å^3; this drop is expected because of the loss of the hydrogen. But if we take the 2′-oxygen volume of 17.39 Å^3 into account, we now have a total volume of 30.07 Å^3 to fit into the space of 18.0 Å^3. Therefore, RNA structure must accommodate this additional occupied space.

Practical applications

We now point out in the paper how our parameter set is useful for RNA studies and, in fact, directly increases our understanding of RNA structure. Many applications of our volumes and radii come to mind. We provide three new data sets: a set of atomic RNA volumes, a set of RNA VDW radii and a variety of annotated sets of large,

non-redundant, RNA-containing PDB structures. Many programs used for structure solving and model refinement use VDW radii. Any RNA informatics endeavor requires begins with annotated sets of PDB structures, which we provide.

Packing density

Structure–function research can involve the atomic radii and volumes of RNA in order to locate non-standard regions and possibly functional areas. In particular, the volumes can be used to measure the local packing density, the ratio of a given atom to its expected volume within a particular region. The local packing density can be used to determine more and less packed regions within a particular structure. Second, we can use the packing density to locate atoms with extreme volumes. This is useful in evaluating the quality of a crystal structures by locating areas that are packed too loosely or too tightly. Additionally, regions with extreme volumes may pinpoint active sites or other functional features. Finally, packing density is an accepted method for measuring the tightness of fit between RNA and a substrate, such as polymerases, RNases and other RNA-binding molecules.

Volume and PSV calculation

Before our volume results, calculating molecular volumes for RNA containing macromolecules was limited. Two techniques have existed for determining the volume of unknown particles: electron microscopy and small-angle X-ray scattering.[56] Both methods are problematic. Previous studies of 50 S ribosomal subunits to determine their volume did so with a large range of 1.8–4.4 million cubic Angstroms.[56]

Using our published volume set, we can estimate the molecular volume of any RNA based solely on its sequence. For example, the Voronoi volume of 50 S small subunit structure is 1,374,538 Å^3. Based on the actual three-dimensional coordinates of the solved ribosome structure, the Richards' rolling probe method[16] calculates the molecular volume to be 1,400,281 Å^3. This slight difference of 1.8% is reasonable considering we are only using sequence information. Therefore, in essence, we can get a good estimate for the volume without knowing three-dimension coordinates. Further, if the sequence is known then the mass is readily calculated to obtain a partial specific volume for any unknown structure. Using our volumes, we calculate the partial specific volume of the 50 S subunit to be 0.617 ml/g which compares well to the published value of approximately 0.592 ml/g.[56,57]

We have also built a web tool† to perform this calculation of volume and PSV on an arbitrary RNA or protein sequence. For instance, application of the tool to the U65 snoRNA 172 nt consensus sequence,[58] which currently has an unknown structure, shows it to have a volume of 41,700.4 Å^3 and a PSV of 0.569 ml/g.

Exploration of the ribosome

One immediate future application of our parameter sets is the analysis of the ribosome.[59] The ribosome has an extremely complex intertwined folding of protein and RNA that is currently not fully understood. It is an open question how this large macromolecule packs together. We can now use our volume and radii parameters to analyze internal solvent volumes. In a similar sense we can evaluate which helices within the ribosome structure interact with which other helices. This is similar in spirit to work done on membrane proteins.[60,61] Finally, the exit tunnel is a site of antibiotic binding; using our new parameter sets we can trace out the volume and diameter as a function of distance from the active site to better understand how these molecules are functioning to block translation.

Conclusions

In this study, we performed a careful parameterization of currently available RNA structures to obtain a universal, self-consistent set of volumes, denoted as the NucProt parameter set. This composite set can be applied to both RNA and protein. In addition, several factors such as crystal symmetry, structural complexity and protein and solvent interactions are taken into account for their influence on the final results. Using two measures, the percentage of well-packed atoms and final volumes, the impact of each factor was assessed on the data. While all the factors affected the percentage of well-packed atoms, none of them had any affect on the final volumes. From these volume calculations, it is immediately apparent that the RNA backbone is not as tightly packed as its base as determined by its standard deviation and also its percentage of well-packed atoms. For RNA, the calculated partial specific volume corresponded well with its experimental value. When compared to proteins, RNA is found to be more dense, because its partial specific volume is smaller. Comparing common atom types between protein and RNA showed that in nine of 12 cases, RNA has a smaller volume and is therefore packed tighter. Further, when comparing aromatic protein side-chains to the RNA bases, the partial specific volume for RNA bases was again smaller than the protein side-chains as well as their average volume per atom. Thus, RNA packs more tightly than protein, but based only on well-packed atoms. A-form DNA, on the other hand, has approximately the same base volumes as RNA, though the backbones differ by more than the bases it is within the standard deviation of the total volume. In conclusion, RNA packs more tightly than protein and approximately the same as DNA.

Location of files, programs, scripts and statistics.

Further details on parameter sets, additional statistics, perl and shell scripts, packaged program files, and the raw volume data are provided online†.

References

1. Bondi, A. (1964). Van der Waals volumes and radii. *J. Phys. Chem.* **68**, 441–451.
2. Chothia, C. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature*, **248**, 338–339.
3. Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1–14.
4. Finney, J. L. (1975). Volume occupation, environment and accessibility in proteins. The problem of the protein surface. *J. Mol. Biol.* **96**, 721–732.
5. Harpaz, Y., Gerstein, M. & Chothia, C. (1994). Volume changes on protein folding. *Structure*, **2**, 641–649.
6. Li, A. J. & Nussinov, R. (1998). A set of Van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins: Struct. Funct. Genet.* **32**, 111–127.
7. Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V. & Subramaniam, S. (1998). Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. *Proteins: Struct. Funct. Genet.* **33**, 18–29.
8. Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V. & Subramaniam, S. (1998). Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins: Struct. Funct. Genet.* **33**, 1–17.
9. Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999). The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **290**, 253–266.
10. Tsai, J. & Gerstein, M. (2002). Calculations of protein volumes: sensitivity analysis and parameter database. *Bioinformatics*, **18**, 985–995.
11. Tsai, J., Voss, N. & Gerstein, M. (2001). Determining the minimum number of types necessary to represent the sizes of protein atoms. *Bioinformatics*, **17**, 949–956.
12. Nadassy, K., Tomas-Oliveira, I., Alberts, I., Janin, J. & Wodak, S. J. (2001). Standard atomic volumes in double-stranded DNA and packing in protein–DNA interfaces. *Nucl. Acids Res.* **29**, 3362–3376.
13. Chothia, C. (1975). Structural invariants in protein folding. *Nature*, **254**, 304–308.
14. Janin, J. & Chothia, C. (1990). The structure of protein–protein recognition sites. *J. Biol. Chem.* **265**, 16027–16030.
15. Janin, J. (1979). Surface and inside volumes in globular proteins. *Nature*, **277**, 491–492.
16. Richards, F. M. (1985). Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol.* **115**, 440–464.
17. Gerstein, M., Tsai, J. & Levitt, M. (1995). The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **249**, 955–966.
18. Gerstein, M. & Chothia, C. (1996). Packing at the protein–water interface. *Proc. Natl Acad. Sci. USA*, **93**, 10167–10172.
19. Hubbard, S. J. & Argos, P. (1995). Detection of internal cavities in globular proteins. *Protein Eng.* **8**, 1011–1015.
20. Pontius, J., Richelle, J. & Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**, 121–136.
21. Gerstein, M., Sonnhammer, E. L. & Chothia, C. (1994). Volume changes in protein evolution. *J. Mol. Biol.* **236**, 1067–1078.
22. Gerstein, M. (1998). How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold. Des.* **3**, 497–512.
23. Gerstein, M. & Krebs, W. (1998). A database of macromolecular motions. *Nucl. Acids Res.* **26**, 4280–4290.
24. Krebs, W. G. & Gerstein, M. (2000). The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucl. Acids Res.* **28**, 1665–1675.
25. David, C. W. (1988). Voronoi polyhedra as structure probes in large molecular systems. *Biopolymers*, **27**, 339–344.
26. Finney, J. L. (1978). Volume occupation, environment, and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* **119**, 415–441.
27. Dunbrack, R. L., Jr (1999). Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL. *Proteins Suppl.* **3**, 81–87.
28. Koehl, P. & Delarue, M. (1997). The native sequence determines sidechain packing in a protein, but does optimal sidechain packing determine the native sequence? *Pac. Symp. Biocomput.*, 198–209.
29. Lee, C. & Levitt, M. (1997). Packing as a structural basis of protein stability: understanding mutant properties from wildtype structure. *Pac. Symp. Biocomput.*, 245–255.
30. Voronoi, G. F. (1908). Nouveles applications des paramétres continus á la théorie de formas quadratiques. *J. Reine. Angew. Math.* **134**, 198–287.
31. Bernal, J. D. & Finney, J. L. (1967). Random close-packed hard-sphere model II. geometry of random packing of hard spheres. *Disc. Faraday Soc.* **43**, 62–69.
32. Gellatly, B. J. & Finney, J. L. (1982). Calculation of protein volumes: an alternative to the Voronoi procedure. *J. Mol. Biol.* **161**, 305–322.
33. Tsai, J., Gerstein, M. & Levitt, M. (1997). Simulating the minimum core for hydrophobic collapse in globular proteins. *Protein Sci.* **6**, 2606–2616.
34. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D: Biol. Crystallogr*, **54**, 905–921.
35. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T. *et al*. (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.
36. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al*. (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
37. Durchschlag, H. (1986). Specific volumes of biological macromolecules and some other molecules of biological interest. In *Thermodynamic Data for Biochemistry and Biotechnology* (Hinz, H.-J., ed.), pp. 45–128, Springer, Berlin.
38. Draper, D. E. (2004). A guide to ions and RNA structure. *RNA*, **10**, 335–343.
39. Su, L. J., Brenowitz, M. & Pyle, A. M. (2003). An

† http://geometry.molmovdb.org

alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *J. Mol. Biol.* **334**, 639–652.
40. Pyle, A. M. (2002). Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* **7**, 679–690.
41. Misra, V. K. & Draper, D. E. (2002). The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317**, 507–521.
42. Kim, H. D., Nienhaus, G. U., Ha, T., Orr, J. W., Williamson, J. R. & Chu, S. (2002). Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc. Natl Acad. Sci. USA*, **99**, 4284–4289.
43. Silverman, S. K., Deras, M. L., Woodson, S. A., Scaringe, S. A. & Cech, T. R. (2000). Multiple folding pathways for the P4-P6 RNA domain. *Biochemistry*, **39**, 12465–12475.
44. Hanna, R. & Doudna, J. A. (2000). Metal ions in ribozyme folding and catalysis. *Curr. Opin. Chem. Biol.* **4**, 166–170.
45. Ryder, S. P. & Strobel, S. A. (1999). Nucleotide analog interference mapping of the hairpin ribozyme: implications for secondary and tertiary structure formation. *J. Mol. Biol.* **291**, 295–311.
46. Rook, M. S., Treiber, D. K. & Williamson, J. R. (1999). An optimal Mg(2+) concentration for kinetic folding of the tetrahymena ribozyme. *Proc. Natl Acad. Sci. USA*, **96**, 12471–12476.
47. Batey, R. T. & Doudna, J. A. (1998). The parallel universe of RNA folding. *Nature Struct. Biol.* **5**, 337–340.
48. Strobel, S. A. & Doudna, J. A. (1997). RNA seeing double: close-packing of helices in RNA tertiary structure. *Trends Biochem. Sci.* **22**, 262–266.
49. McConnell, T. S., Herschlag, D. & Cech, T. R. (1997). Effects of divalent metal ions on individual steps of the tetrahymena ribozyme reaction. *Biochemistry*, **36**, 8293–8303.
50. Doudna, J. A. & Cate, J. H. (1997). RNA structure: crystal clear? *Curr. Opin. Struct. Biol.* **7**, 310–316.
51. Doudna, J. A. & Doherty, E. A. (1997). Emerging themes in RNA folding. *Fold. Des.* **2**, R65–R70.
52. Cate, J. H., Hanna, R. L. & Doudna, J. A. (1997). A magnesium ion core at the heart of a ribozyme domain. *Nature Struct. Biol.* **4**, 553–558.
53. Draper, D. E. (1996). Strategies for RNA folding. *Trends Biochem. Sci.* **21**, 145–149.
54. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A. *et al*. (1996). RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.
55. Pyle, A. M. & Green, J. B. (1995). RNA folding. *Curr. Opin. Struct. Biol.* **5**, 303–310.
56. Van Holde, K. E. & Hill, W. E. (1974). General physical properties of ribosomes. In *Ribosomes* (Nomura, M., Tissieres, A. & Lengyel, P., eds), pp. 53–91, Cold Spring Harbor Press, Cold Spring Harbor, NY.
57. Hill, W. E., Rossetti, G. P. & Van Holde, K. E. (1969). Physical studies of ribosomes from *Escherichia coli*. *J. Mol. Biol.* **44**, 263–277.
58. Ganot, P., Bortolin, M. L. & Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
59. Steitz, T. A., Moore, P. B. (2004).
60. Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O. & Fleming, P. J. (2000). Internal packing of helical membrane proteins. *Proc. Natl Acad. Sci. USA*, **97**, 5796–5801.
61. Gerstein, M. & Chothia, C. (1999). Perspectives: signal transduction. Proteins in motion. *Science*, **285**, 1682–1683.
62. Lee, A. & Chalikian, T. V. (2001). Volumetric characterization of the hydration properties of heterocyclic bases and nucleosides. *Biophys. Chem.* **92**, 209–227.

Edited by D. E. Draper

(Received 19 August 2004; received in revised form 24 November 2004; accepted 24 November 2004)