JCTC Journal of Chemical Theory and Computation

AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA

Raviprasad Aduri, Brian T. Psciuk, Pirro Saro, Hariprakash Taniga, H. Bernhard Schlegel, and John SantaLucia, Jr.*

Department of Chemistry, Wayne State University, Detroit, Michigan 48202

Received November 7, 2006

Abstract: Classical molecular dynamics (MD) simulations are useful for characterizing the structure and dynamics of biological macromolecules, ultimately, resulting in elucidation of biological function. The AMBER force field is widely used and has well-defined bond length, bond angle, partial charge, and van der Waals parameters for all the common amino acids and nucleotides, but it lacks parameters for many of the modifications found in nucleic acids and proteins. Presently there are 107 known naturally occurring modifications that play important roles in RNA stability, folding, and other functions. Modified nucleotides are found in almost all transfer RNAs, ribosomal RNAs of both the small and large subunits, and in many other functional RNAs. We developed force field parameters for the 107 modified nucleotides currently known to be present in RNA. The methodology used for deriving the modified nucleotide parameters is consistent with the methods used to develop the Cornell et al. force field. These parameters will improve the functionality of AMBER so that simulations can now be readily performed on diverse RNAs having post-transcriptional modifications.

1. Introduction

Ribonucleic acids (RNA) play important roles in diverse biological functions including protein synthesis, gene silencing, and in the regulation of gene expression.[1−3] RNA is initially synthesized as a phosphodiester polymer of four nucleosides namely adenosine, guanosine, cytidine, and uridine, which are called the "common" nucleosides. In addition to the four common nucleosides, there are many modified nucleosides found in RNA.[4] These nucleoside modifications are formed post-transcriptionally. Presently there are at least 107 modifications that have been discovered in natural RNA.[5−8] Modified nucleosides are found in almost all tRNAs, ribosomal RNAs of both the small and large subunits of the ribosome, mRNAs, snoRNA, and other functionally important RNA molecules.[5] Currently, the biological functions of most modifications are unknown, though some roles are beginning to be elucidated.[9−11] The most commonly occurring modification is pseudouridine, in which the C5 of uracil is covalently attached to the sugar C1′, resulting in a C−C glycosidic bond instead of the usual C−N glycosidic bond.[12] The next most common modification found in RNA is the methylation of the 2′-O position of the ribose sugar. The lifetimes of base pairs involving certain modified nucleosides are reported to be longer than the typical Watson−Crick base pairs, making these modifications essential for the viability of extremophiles.[5] Owing to the ubiquitous presence of the modified nucleosides in RNA, it is essential to develop accurate and reliable force field parameters for these modifications that enable the simulation of molecular dynamics of RNA with or without modifications.[13] Stable MD simulations require uniformity in the force field parameter sets for modified nucleosides to be consistent with the present force field for the common nucleosides.

Molecular mechanics (MM) and molecular dynamics (MD) are useful for revealing dynamics and structure of biomacromolecules thereby elucidating biological function. There are several MM force fields available for performing simulations of biomolecules including CHARMM,[14] AMBER,[15] XPLOR,[16] and others.[17] Armed with an increasing amount of computational resources, researchers have successfully incorporated more accuracy and elegance to force

* Corresponding author phone: (313)577-0101; fax: (313)577-8822; e-mail: jsl@chem.wayne.edu.

10.1021/ct600329w CCC: $37.00
Published on Web 06/12/2007

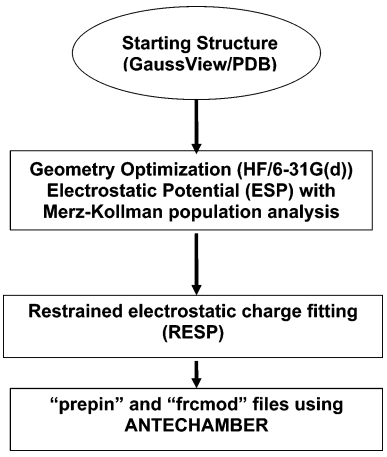

Figure 1. Flowchart of the protocol used in generating the parameters for modified nucleotides.

fields including polarizable functions,[18] lone pairs, coupled stretching and bending modes, and sophisticated models of solvation and electrostatics.[19] AMBER is one of the most widely used force fields in the simulation of biological molecules possessing the necessary parameters for the common nucleosides and amino acids. Recently, AMBER force field parameters were developed for phosphorothioate nucleic acids[20] as well as for various polyphosphates.[21] Presently, force field parameters are available for modifications found in $tRNA^{Phe}$ (http://pharmacy.man.ac.uk/amber/nuc/tRNA_inf.html) and some of the 2′ sugar modifications.[22] Some groups have reported parameters for a few modifications present in the anticodon stem loop of the tRNA.[23−25] However, parameters for the other naturally occurring modified nucleosides are not contained within the AMBER suite.

An expanding knowledge surrounding the role of RNA in various biological processes and the presence of a large variety of modified nucleosides provide an important demand for the development of force field parameters for modified nucleosides suitable for use with the well-established AMBER force field. Herein, we report the development of force field parameters for the known 107 modified nucleosides found in natural RNA.[4] The modified RNA force field parameters have been developed to be consistent with the Cornell et al. force field[26] of AMBER.

2. Methods

2.1. Parametrization Strategy. The strategic approach used for developing AMBER force field parameters for the 107 modifications in RNA is summarized in Figure 1. The parametrization protocol developed by Cornell et al.[26] was followed to be consistent with the AMBER force field. Atom-centered partial charges were calculated using the RESP methodology. The electronic structure calculations were carried out at the Hartree−Fock level of theory using the 6-31G(d) basis set despite improvements in computing resources that would have enabled us to perform calculations at higher levels of theory. In this way, the calculations in this work are consistent with the procedure followed in the original development of the AMBER Cornell et al. force field.[26] To obtain the charge constraint for the sugar moiety, QM calculations were performed on the four common nucleosides, A, C, G, and U with both C3′endo and C2′endo ribose sugars. In both these cases, the sugar atoms among all four nucleosides were equivalenced. The phosphate group and O3′ and O5′ charges were obtained using dimethyl phosphate (DMP) as the model system as shown in Figure 2. RESP charge fitting was done with all the four nucleosides with either C2′ endo or C3′ endo sugar. C2′endo and C3′endo nucleosides were also fit together during the RESP procedure (data not shown). The modifications may play a role altering the sugar pucker, but the sugar pucker preferences for the modified nucleosides are not well understood.[27,28] Since RNA predominantly contains a C3′ endo ribose sugar conformation, and it is the conformation of the sugar used in the initial development of AMBER parameters, we decided to use the charge obtained from the RESP fitting of only the C3′-endo sugar containing nucleosides. The ribose sugar charge was calculated by multiequivalencing the four natural nucleosides A, G, C, and U with C3′ endo sugar conformation as described in Cieplak et al.[29] The charges obtained for the common nucleosides in the C3′-endo conformation are given in Table 1. The charges obtained for C2′-endo ribose, C3′-endo ribose sugar, and 2′-*O*-methyl ribose are given in Table 2. The ribose sugar charges are relatively insensitive to sugar pucker conformations. The standard deviation of charges for comparison of C3′endo vs C2′endo riboses is 0.0269e which is less than the systematic error of the RESP methodology itself, and thus there is no need for separate parametrization of C3′and C2′-endo sugar puckers.

2.2. Ab Initio Calculations. AMBER force field parameters were developed by performing ab initio calculations at the Hartree−Fock level of theory using the 6-31G(d) basis set using the GAUSSIAN03[30] suite of programs. To test that our calculations followed the Restrained ElectroStatic Potential (RESP) charge fitting methodology[31] procedure outlined in the Cieplak et al.,[29] we performed computations on the four commonly occurring nucleosides A, C, G, and U. The charges reported by Cieplak et al. are in excellent agreement (with a standard deviation of 0.0362e, see Table 4) with those determined here, thereby validating our approach. The modular nature of the RESP as well as of the structure of RNA itself allowed us to split the nucleosides into separate base, sugar, and phosphate moieties resulting in the reduction of the computational burden. To account for the phosphate charge, dimethyl phosphate (DMP) was used as the model system. Nucleosides with modifications in the base moiety were modeled by replacing the sugar with a methyl group. Conversely, nucleosides with modifications in the sugar moiety were modeled by replacing the base with a methyl group (Figure 2). The RESP procedure developed by Kollman and colleagues allows a modular approach to recombine sugar and base moieties by "equivalencing".[31] This strategy not only reduces the number of atoms in each ab initio computation but also allows portability of parameters so that different bases and sugars can be appropriately constructed. For example, once computations for 2′-*O*-methyl ribose are complete, the results can be combined with a variety of bases. Conversely, a modified base can be recombined with differing sugars (e.g., ribose, deoxyribose,

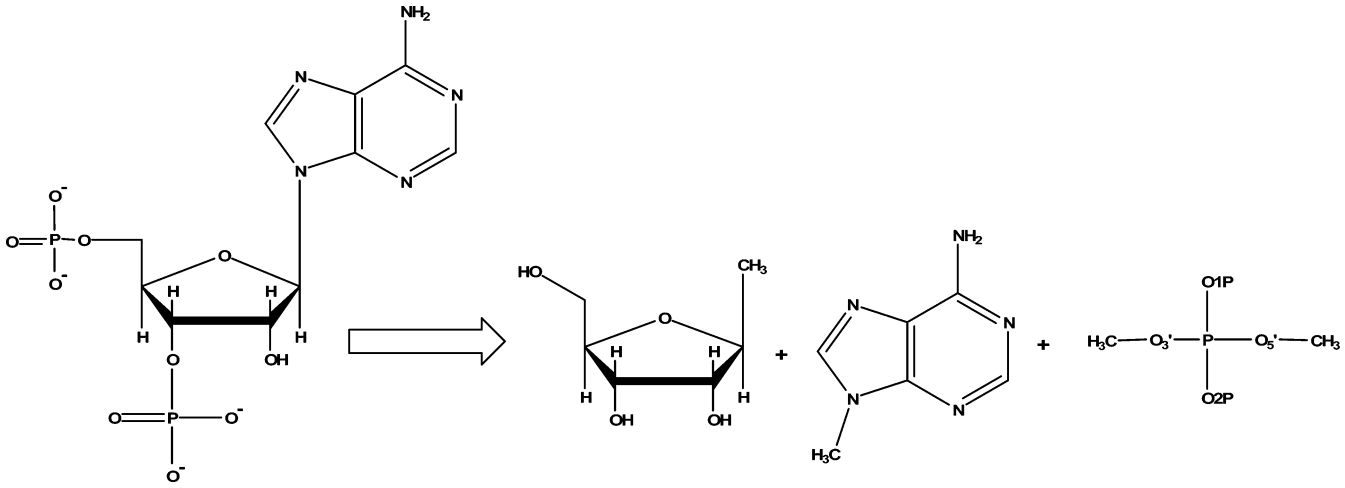

Figure 2. The charge fitting method used to generate the charges for the common nucleosides A, G, C, and U, using the modular nature of RNA to reduce the computational time. The charges for O1P, O2P, O3′, O5′, and P were obtained by using dimethyl phosphate (DMP) as a model system. See text for explanation.

Table 1. Charge Values Obtained in this Work for the Common Nucleosides A, G, C, and U

adenosine		guanosine		cytidine		uridine	
N9	0.0172	N9	0.0268	N1	−0.2152	N1	0.1110
C8	0.1299	C8	0.1066	C2	0.8867	C2	0.4539
N7	−0.5850	N7	−0.5575	O2	−0.6560	O2	−0.5407
C6	0.7111	C6	0.5316	N3	-0.8128	N3	−0.3681
N6	−0.9386	O6	−0.5483	C4	0.9020	C4	0.6022
C5	0.0586	C5	0.1513	N4	−0.9919	O4	−0.5652
C4	0.3050	C4	0.1563	C5	−0.5972	C5	−0.3135
N3	−0.6835	N3	−0.5959	C6	0.1262	C6	−0.2320
C2	0.5741	C2	0.7191	H5	0.2023	H5	0.1697
N1	−0.7536	N2	−0.9044	H6	0.1875	H6	0.2557
H8	0.1749	N1	−0.5287	NH1	0.4251	N3H	0.3087
H2	0.0467	H8	0.1767	NH2	0.4251		
HN1	0.4125	N2H1	0.3968				
HN2	0.4125	N2H2	0.3968				
		N1H	0.3546				

2′-*O*-methyl ribose, etc.). Anticipating the future discovery of new modified RNA, this strategy will allow for many nucleosides to be modeled that have not yet been found in nature or artificially synthesized. For example, deoxy pseudouridine is not found in nature, but it could be constructed from the parameters presented here for the pseudouridine along with the deoxyribose sugar parameters. The physiological pH of 7.0 was used in deciding the protonation states of all functional groups. Other protonation states observed at different pH were not considered in this study.[32] In addition, only the lowest energy tautomeric state was considered. The starting geometries for ab initio calculations were obtained from the PDB database, when available. When a suitable crystal structure could not be retrieved, the structure was generated using GaussView and GAUSSIAN03. Hydrogens were added to the PDB structures using an automated feature in GaussView. Each nucleoside was manually inspected to ensure the proper valence of each heavy atom. The generic names, three-letter codes, starting geometries, and, where available, the RNA in which they occur for all modified nucleotides are summarized in Table 3.

Table 2. Charge Values Obtained for the Three Common Sugars in RNA, C3′-Endo, C2′-Endo, and 2′-*O*-Methyl Ribose Sugars

atom name	C3′-endo	C2′-endo	2′O methyl ribose
P	1.0878	1.0825	1.0878
O1P	−0.7667	−0.7655	−0.7667
O2P	−0.7667	−0.7655	−0.7667
O5′	−0.4713	−0.5036	−0.4725
C5′	0.0635	0.0292	0.1289
C4′	0.0386	0.0625	0.1522
O4′	−0.3272	−0.3851	−0.4652
C3′	0.2125	0.2165	0.0675
O3′	−0.4890	−0.4649	−0.4878
C2′	0.0775	0.1064	0.0405
O2′	−0.5913	−0.6198	−0.3277
C1′	0.0460	0.1096	0.3686
H5′	0.0689	0.0823	0.0426
H5″	0.0689	0.0823	0.0426
H4′	0.1168	0.1215	0.0394
H3′	0.0825	0.0858	0.1460
H2′	0.0929	0.0659	0.0904
H1′	0.1643	0.1462	0.0417
OH2′	0.4101	0.4210	na[a]
CM2	na[a]	na[a]	−0.0385
HM′1	na[a]	na[a]	0.0651
HM′2	na[a]	na[a]	0.0651
HM′3	na[a]	na[a]	0.0651

[a] na − not applicable.

2.3. Electrostatic Potential Calculations. After geometry optimization, the electrostatic surface potential (ESP) was fit using the electrostatic charge computing method developed by Merz and Kollman,[33] which uses a Connolly surface algorithm to calculate a number of shells with radii of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radius of the constituent atoms in the molecule. A Levenberg−Marquardt nonlinear optimization procedure was then used to compute the set of atom-centered point charges that best reproduce

Table 3. Generic Names, Three-Letter Codes, Source of Starting Geometry, and the Occurrence of Different Modifications[a]

generic name	three-letter code[b]	alternate codes[c]	source[d]	occurrence
1-methyladenosine	1MA		(1EHZ)	tRNA
2-methylthio-*N*^{6}-hydroxynorvalyl carbamoyladenosine	26A			tRNA
2-methyladenosine	2MA		1EFW	tRNA
2′-*O*-ribosylphosphate adenosine	2RA		1YFZ	tRNA
N^6-methyl-*N*^{6}-threonylcarbamoyladenosine	66A			tRNA
N^6-acetyladenosine	6AA			tRNA
N^6-glycinylcarbamoyladenosine	6GA			tRNA
N^6-isopentenyladenosine	6IA			tRNA
N^6-methyladenosine	6MA			tRNA
N^6-threonylcarbamoyladenosine	6TA			tRNA
N^6,*N*^{6}-dimethyladenosine	DMA	M2A		16S rRNA
N^6-(*cis*-hydroxyisopentenyl)adenosine	HIA			tRNA
N^6-hydroxynorvalylcarbamoyladenosine	HNA			tRNA
1,2′-*O*-dimethyladenosine	M2A			tRNA
N^6,2′-*O*-dimethyladenosine	MMA			
2′-*O*-methyladenosine	MRA	A2M		tRNA
N^6,*N*^{6},*O*-2′-trimethyladenosine	MTA			
2-methylthio-*N*^{6}-(*cis*-hydroxyisopentenyl) adenosine	SIA			tRNA
2-methylthio-*N*^{6}-methyladenosine	SMA			tRNA
2-methylthio-*N*^{6}-isopentenyladenosine	SPA	MIA	1B23	tRNA
2-methylthio-*N*^{6}-threonyl carbamoyladenosine	STA	12A	1FIR	tRNA
2-thiocytidine	2SC			tRNA
3-methylcytidine	3MC		3MCT	tRNA
N^4-acetylcytidine	4AC			tRNA,rRNA
N^4-methylcytidine	4MC			
5-formylcytidine	5FC			tRNA
5-methylcytidine	5MC		1EHZ	tRNA,16S rRNA
5-hydroxymethylcytidine	HMC			
lysidine	K2C			tRNA
N^4-acetyl-2′-*O*-methylcytidine	MAC			tRNA,rRNA
5-formyl-2′-*O*-methylcytidine	MFC			tRNA
5,2′-*O*-dimethylcytidine	MMC			tRNA
2′-*O*-methylcytidine	MRC	OMC	1EHZ	tRNA
N^4,2′-*O*-dimethylcytidine	M4C			rRNA
N^4,*N*^{4},2′-*O*-trimethylcytidine	MTC			rRNA
1-methylguanosine	1MG		2ASY	tRNA
N^2,7-dimethylguanosine	27G			
N^2-methylguanosine	2MG		1EHZ	tRNA,rRNA
2′-*O*-ribosylphosphate guanosine	2RG			tRNA
7-methylguanosine	7MG	G7M	1EHZ	tRNA,rRNA
under modified hydroxywybutosine	BUG	UBG		tRNA
7-aminomethyl-7-deazaguanosine	DAG		1EFZ	tRNA
7-cyano-7-deazaguanosine	DCG			tRNA
N^2,*N*^{2}-dimethylguanosine	DMG	M2G	(1EHZ)	tRNA
4-demethylwyosine	DWG			tRNA
epoxyqueuosine	EQG			tRNA
hydroxywybutosine	HWG			tRNA
isowyosine	IWG			tRNA
N^2,7,2′-*O*-trimethylguanosine	M7G			tRNA
N^2,2′-*O*-dimethylguanosine	MMG			tRNA
1,2′-*O*-dimethylguanosine	M1G			tRNA
2′-*O*-methylguanosine	MRG	OMG	1EHZ	tRNA
N^2,*N*^{2},2′-*O*-trimethylguanosine	MTG			tRNA
N^2,*N*^{2},7-trimethylguanosine	N2G			
peroxywybutosine	PBG			tRNA
galactosyl-queuosine	QGG			tRNA
mannosyl-queuosine	QMG			tRNA
queuosine	QUG	QUO		tRNA

Table 3. (Continued)

generic name	three-letter code[b]	alternate codes[c]	source[d]	occurrence
archaeosine	RCG			tRNA
wybutosine	WBG	YG	1EHZ	tRNA
methylwyosine	WMG			tRNA
wyosine	WYG			tRNA
2-thiouridine	2SU	SUR		tRNA
3-(3-amino-3-carboxypropyl)uridine	3AU			tRNA
3-methyluridine	3MU			rRNA
4-thiouridine	4SU	S4U	1B23	tRNA
5-methyl-2-thiouridine	52U			tRNA
5-methylaminomethyluridine	5AU			tRNA
5-carboxymethyluridine	5CU			
5-carboxymethylaminomethyluridine	5DU			tRNA
5-hydroxyuridine	5HU			tRNA
5-methyluridine	5MU		1EHZ	tRNA
5-taurinomethyluridine	5TU			tRNA
5-carbamoylmethyluridine	BCU			tRNA
5-(carboxyhydroxymethyl)uridine methyl ester	CMU			tRNA
dihydrouridine	DHU	H2U	1EHZ	tRNA
5-methyldihydrouridine	DMU			
5-methylaminomethyl-2-thiouridine	ESU			tRNA
5-(carboxyhydroxymethyl)uridine	HCU			tRNA
5-(isopentenylaminomethyl)uridine	IAU			tRNA
5-(isopentenylaminomethyl)-2-thiouridine	ISU			tRNA
3,2′-*O*-dimethyluridine	M3U			
5-carboxymethylaminomethyl-2′-*O*-methyluridine	MAU			tRNA
5-carbamoylmethyl-2′-*O*-methyluridine	MCU			tRNA
5-methoxycarbonylmethyl-2′-*O*-methyluridine	MEU			tRNA
5-(isopentenylaminomethyl)-2′-*O*-methyluridine	MIU			tRNA
5,2′-*O*-dimethyluridine	MMU	2MU	1FIR	tRNA
2′-*O*-methyluridine	MRU			tRNA
2-thio-2′-*O*-methyluridine	MSU			tRNA
uridine 5-oxyacetic acid	OAU			tRNA
5-methoxycarbonylmethyluridine	OCU			tRNA
uridine 5-oxyacetic acid methyl ester	OEU			tRNA
5-methoxyuridine	OMU			tRNA
5-aminomethyl-2-thiouridine	SAU			tRNA
5-carboxymethylaminomethyl-2-thiouridine	SCU			tRNA
5-methylaminomethyl-2-selenouridine	SEU			tRNA
5-methoxycarbonylmethyl-2-thiouridine	SMU			tRNA
5-taurinomethyl-2-thiouridine	STU			tRNA
pseudouridine	PSU		1EHZ	tRNA,rRNA
1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine	13P			28S rRNA
1-methylpseudouridine	1MP			tRNA
3-methylpseudouridine	3MP			23S rRNA
2′-*O*-methylpseudouridine	MRP			tRNA
inosine	INO			tRNA
1-methylinosine	1MI			tRNA
1,2′-*O*-dimethylinosine	MMI			tRNA
2′-*O*-methylinosine	MRI			

[a] The PDB reference is given for nucleosides where available. GaussView was used for generating the starting geometry wherever the PDB source is not mentioned. [b] Three-letter code proposed in this study. [c] Alternate three-letter codes used previously. [d] Source refers to where we obtained the coordinates for starting geometries of modified nucleosides. Values in parenthess indicate that the modification occurs in that PDB file, but it was not used in this work. If no PDB source is given or if in parentheses, then the starting geometry was generated using GaussView.

the surface charges that were derived quantum mechanically.[34] Because of differences in convergence criteria, the optimized geometry of the molecule may also differ slightly based on the QM program used, which would alter charge values. The grid size (i.e., the number of shells of points and the density of points on the shells) used to compute the electrostatic potential slightly influences the atom centered point charges in the ESP calculation. It is well-known that the atomic charges derived from using a grid of electrostatic potentials computed by quantum mechanical calculations

Table 4. Comparison of Adenosine Charges Computed in This Work with the Charges Available in PARM99 of AMBER

atom name	adenosine whole nucleoside	adenosine PARM 99	adenosine modular fit
N9	0.0172	−0.0251	−0.0503
C8	0.1299	0.2006	0.1060
N7	−0.5850	−0.6073	−0.5725
C6	0.7111	0.7009	0.6394
N6	−0.9386	−0.9019	−0.8963
C5	0.0586	0.0515	0.0553
C4	0.3050	0.3053	0.4499
N3	−0.6835	−0.7615	−0.7282
C2	0.5741	0.5875	0.5587
N1	−0.7536	−0.6997	−0.7354
H8	0.1749	0.1553	0.1734
H2	0.0467	0.0473	0.0579
NH1	0.4125	0.4115	0.4122
NH2	0.4125	0.4115	0.4122
SD from PARM99	0.0362	N/A	0.0529
SD from whole nucleoside	N/A	0.0362	0.0506

depend slightly on the rotational orientation of the molecule.[35] This effect is due to the finite grid used for the sampling of the electrostatic potential (every 1 Å^2 in this study) in the ESP calculation of point charges. To overcome these charge differences due to geometrical orientation of the molecule, multiorientation charge fitting can be utilized.[36] This allows for sampling of many orientations of the molecule, which reduces the round-off errors in atom charges. To test the effect of multiorientation on the charge fitting, we used pseudouridine (PSU) as a model system. The R.E.D. II[36] code provides a good platform for fitting the charges by using a rigid-body reorientation algorithm to make multiple orientations of the molecule. The R.E.D. code allows for the random selection of three different heavy atoms, which are used to orient the molecule. Due to the small standard deviation (∼0.016e, see the Supporting Information) in charge values due to the orientation effect and the laborious computations and file manipulations required to implement multiorientation on the 107 modifications we decided not to perform R.E.D. on the modified nucleosides (see below for discussion).

2.4. Restrained Electrostatic Potential Charges. RESP charge fitting was carried out as described by Cieplak et al.[29] The modular nature of nucleotides allowed for restraining the charge of a methyl group to replace either sugar or base moiety during the ab initio calculation. In the case of base modifications, the total methyl group charge was restrained to the total charge of the sugar (0.118186e) obtained from the common nucleoside calculations during the first stage of RESP fit. When fitting the 2′-*O*-methyl ribose sugar to acquire the charges for this modified sugar, the methyl group replacing the base was restrained to an equivalent and opposite charge value obtained for the normal sugar (i.e., −0.118186e). All equivalent and polar hydrogens, such as hydrogens in an amino group, were equivalenced during the first stage of the RESP fit; whereas, the nonpolar equivalent hydrogens, as in the case of methyl group and H5′ and H5″ of the sugar, were equivalenced in the second stage of the fit. We used ANTECHAMBER Ver 1.24 module of AMBER to do the RESP charge fitting.[37]

2.5. Generating the Parameters. Figure 3 shows a schematic diagram of the protocol followed for generating the charges for the modified nucleosides. A common problem in developing force field parameters is that output text files from one program are incompatible with the input format required for the program used in the next step. For a single RESP computation on a modified nucleotide one could perform such file manipulations manually. For this project, however, performing such manual manipulations on 107 nucleosides is impractical. Thus, we developed several automated text format conversion programs to accomplish this task. Since the ab initio calculations were carried out using the modular approach, we were unable to use the NEWZMAT module of GAUSSIAN to convert the three check point files into a single PDB file for a complete nucleoside. The program gjf2pdb.exe was written to convert GAUSSIAN job files (gjf) into PDB format. This program and others are available on our group home page (http://ozone3.chem.wayne.edu). The nucleoside coordinates were generated by combining the optimized geometry of the modified base with the C3′-endo sugar in GaussView. These Gaussian files were then converted into a single PDB file using gjf2pdb.exe. The resulting PDB files were then used to generate the "ANTECHAMBER" format files using ANTECHAMBER Ver 1.24. Once the ANTECHAMBER files were generated, the charges obtained from the RESP fit were input into the ANTECHAMBER files accordingly. To reduce the development of new atom types, we used the Generalized Amber Force Field (GAFF)[38] to assign the atom types for the modified nucleosides. GAFF contains atom types for all atoms present in the modified nucleosides studied except selenium. In the selenium case, we temporarily decided to assign atom type "SS" to selenium, since the chemical nature of selenium closely resembles sulfur. SS originally represented a thione functional group which is similar in character to the C═Se group found in the modified base 5-methylaminomethyl-2-selenouridine (SEU). The bond lengths, bond angles, and dihedral values used for selenium were similar to atom type "SS". We are in the process of determining the force constants, equilibrium distances, and equilibrium angles for selenium. Once these parameters are available, there may be a need to introduce a new atom type for selenium in GAFF. Once the atom types were assigned, the preparatory file "prepin" and force field file "frcmod" were generated using ANTECHAMBER V.1.24.

3. Naming Convention

We were unable to find a literature consensus in the naming convention used for the modified nucleosides found in RNA. For example, 5,6-dihydrouridine can be found as H2U[39] or DHU.[40] Consequently, we were compelled to develop a consistent three-letter code indicating the nature of the modification as clearly as possible without conflicting with amino acid or other names. In this naming

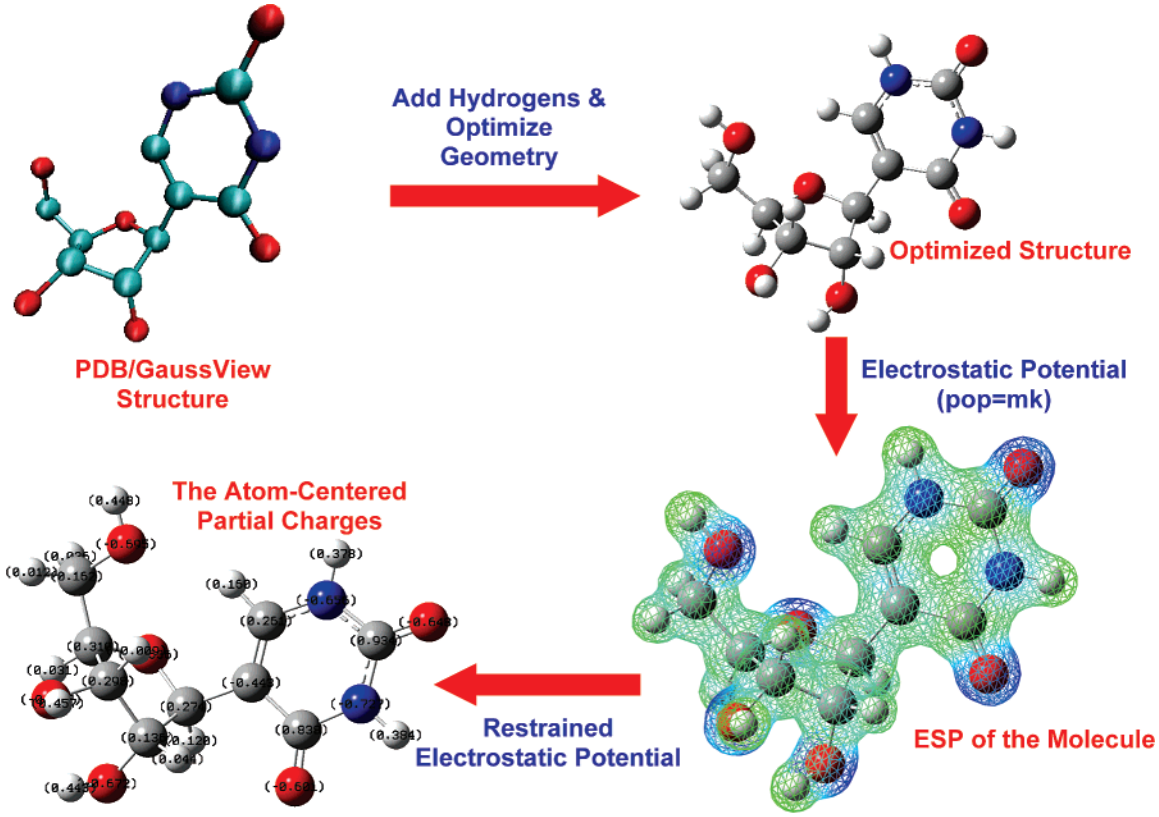

Figure 3. Protocol for the determination of atom-centered partial charges. The starting structures were obtained either from a PDB file or created using GaussView. Hydrogen atoms were added using GaussView. Geometry optimization was done using Gaussian03. The electrostatic potential was computed using Merz−Kollman population analysis, and charges were produced by fitting the ESP using RESP as explained in the text.

convention, the last letter signifies the closest common nucleoside associated with the modification (i.e., the transcribed base encoded in the genomic DNA). For example, wybutosine is named WBG and not Y base, which would conflict with the IUPAC nomenclature for a pyrimidine.[41] Other examples are shown in Table 3. The pseudouridine modification uses "P" as the last letter, and modifications involving inosine were given the letter "I". Additionally, we verified that the three-letter codes used for modified nucleotides did not interfere with any of the letter codes that were already used in AMBER. In the present naming convention, the nature of modification is explicitly used to form the three-letter code when ever possible. For example, 1MA stands for 1-methyladenosine, whereas, 5FC is the code for 5-formylcytidine, and MRX was used to indicate the presence of a 2′-*O*-methyl group on the ribose sugar (e.g. MRA, MRP). We also avoided using A, C, G, and U as the starting letter to escape confusion with the one-letter codes that are still used for the common nucleosides, particularly for sequence alignment algorithms. Thus, the presence of a character other than A, C, G, or U indicates that the three characters in a sequence denote a single modified nucleotide. The generic names along with their three-letter codes for all the 107 modifications are given in Table 3. We hope our naming convention will be widely adopted by the community.

4. Web Site for AMBER Parameters for Modified Nucleosides

Optimized geometries, electrostatic potentials, RESP input and output files, and format conversion executables are available on our Web site http://ozone3.chem.wayne.edu. The Web site also contains the "prepin" and "frcmod" files and the protocols needed to implement the modified nucleoside parameters into AMBER. The optimized geometries of the modifications allow for the opportunity to reproduce the charges obtained in the present study. The modifications are classified according to their closest common nucleotide. For example, 1-methyladenine will be found under the "adenosine modifications" section. The Web site allows users to download parameters for one modification at a time or download parameters for all 107 modifications at once in a compressed file. The Web site also includes other information regarding each particular modification and links to the McCloskey group "RNA Modification database" Web site (http://library.med.utah.edu/RNAmods).[4] Apart from the force field parameters for the modified nucleosides, the Web site also contains the monomer optimized geometries for all the 107 modification. The modified nucleoside parameters have also been made available on the AMBER contributed parameters Web site (http://pharmacy.man.ac.uk/amber).

5. Results and Discussion

The functional form of the AMBER force field is given in eq 1

$$E = \sum_{\text{bonds}} K_r(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi)] + \sum_{\text{nonbonded}} 4\epsilon_{ij}\left(\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \left[\frac{q_i q_j}{\epsilon R_{ij}}\right]\right) \quad (1)$$

The total internal energy of a molecule is decomposed into energy components representing bond stretching, angle bending, torsional angle twisting, Lennard-Jones potential, and nonbonded coulomb electrostatic terms. The present study is focused on developing the atom-centered partial charges necessary to compute the electrostatic term in eq 1 for the 107 naturally occurring modified nucleosides found in RNA. The force constants and equilibrium distances, bond angles, and dihedral terms were generated using the GAFF.[38] As these modifications may not occur at the 5′ or 3′ termini of RNA, we did not develop the parameters for the 5′ or 3′ terminal modifications. We validated the parameters by conducting molecular dynamics simulations on $tRNA^{Phe}$, which contains 14 modified bases. Figure 4 depicts example chemical structures of some of the modified nucleosides for which AMBER parameters were developed.

Charges obtained for the common nucleosides, A, C, G, and U in C3′-endo conformation are shown in Table 1. These charges are in good agreement with the AMBER force field parameters in PARM99 of AMBER as shown in Table 4. Similar agreement was observed for cytidine, guanosine, and uridine (data not shown). We cannot reproduce the charges exactly because the optimized geometry and orientation of the structures used to generate PARM99 are not available. The ribose sugar charge was obtained by equivalencing the four natural nucleosides. Although C1′ and H1′ atoms were not equivalenced in Cieplak et al.,[29] we did not see any major changes in the charges with or without C1′ and H1′ equivalencing. To provide sugar charges that are uniform among all the modified nucleotides, we decided to equivalence the C1′ and H1′ atoms along with all other sugar atoms (data not shown). To confirm that our modular fit reproduces the charges generated on whole nucleosides, we compared the results of a QM calculation on a whole adenosine nucleoside versus an adenine with a methyl replacing the ribose. Table 4 shows the comparison between charges obtained with RESP on the nucleoside vs the methylated free base. Charges from the two methods agree with a standard deviation of 0.0506e for adenosine (0.0594e in the case of guanosine), suggesting that our modular approach is a faithful way of obtaining the charges for these large molecules. As mentioned above, the advantage of using modular approach is to combine different kinds of sugars with different kinds of modified bases thereby avoiding expensive computational calculations. In addition, the largest deviations are observed on the quaternary carbons C8, C6, and C4, which are well-known to be difficult to determine accurately.[20] When we compared our charges generated for adenosine with the charges reported in PARM99, there was good agreement with a standard deviation of 0.0362e. Once the modular approach was tested, it was used to produce the atom centered partial charges for all 107 naturally occurring modified nucleotides found in RNA. Table 5 contains the charges obtained for pseudouridine, inosine, and 5-methylcytosine.

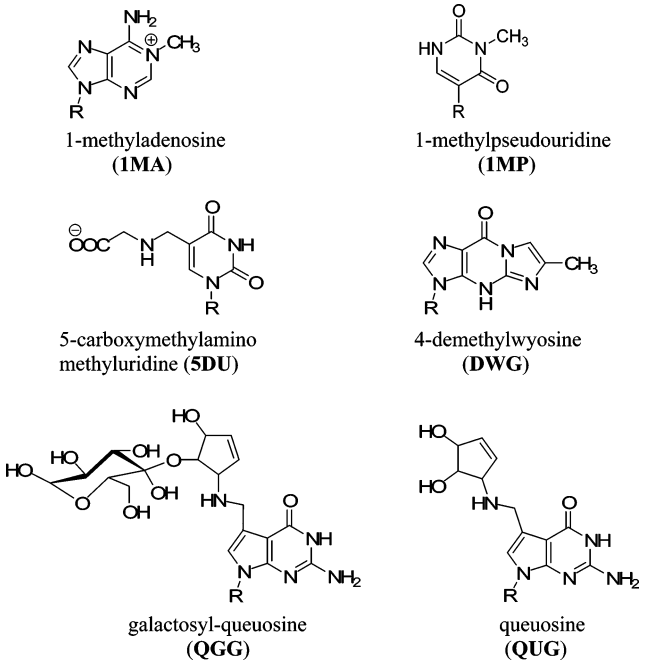

Figure 4. Examples of modified nucleosides present in RNA. These modifications range from simple methylation, as in the case of 1MA, to more complex carbohydrate containing compounds such as QGG. The generic names and their three-letter codes are also given. Only the hydrogen atoms on the polar atoms are shown for clarity. In each case, the only lowest energy tautomer was considered for the protonation that exists at pH 7.

Table 5. Charges Obtained for Pseudouridine, Inosine, and 5-Methylcytosine

pseudouridine		inosine		5-methylcytosine	
C5	−0.2218	N9	−0.0112	N1	−0.0674
C4	0.6913	C8	0.0627	C2	0.7939
O4	−0.5851	N7	−0.5341	O2	−0.6289
N3	−0.4208	C5	0.1198	N3	−0.7268
C2	0.5871	C6	0.5805	C4	0.6304
O2	−0.5729	O6	−0.5538	N4	−0.8933
N1	−0.3019	N1	−0.5208	H41	0.4095
C6	−0.1208	C2	0.3594	H42	0.4095
H6	0.2061	N3	−0.6184	C5	−0.0510
HN1	0.3084	C4	0.3503	C6	−0.1962
HN3	0.3121	H2	0.1223	H6	0.2158
		H1	0.3461	C10	−0.2707
		H8	0.1791	H20	0.0856
				H21	0.0856
				H22	0.0856

To confirm the effect of multiorientation on charge derivation, we used R.E.D. II to apply multiorientation methodology and generate the atom-centered partial charges. Pseudouridine was used with the 5′ and 3′ oxygens capped with hydrogen to reduce the computational burden. As there is no literature available on the optimum number of orientations necessary to get reproducible charges, we decided to perform 4, 8, 12, and 20 different orientations, which are

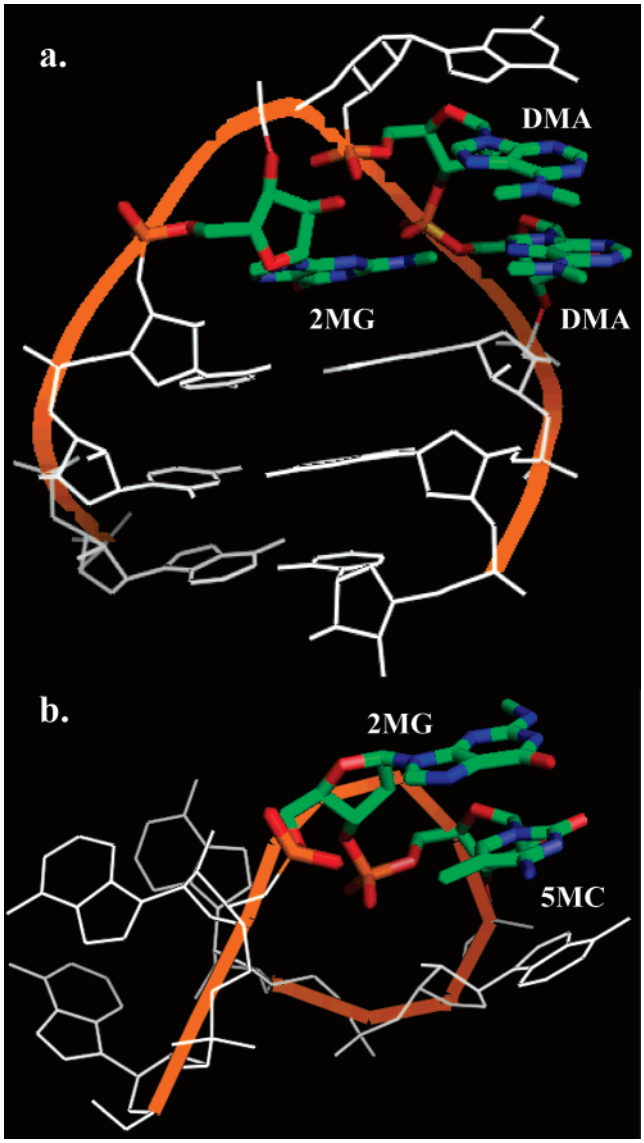

Figure 5. The effect of modifications in the stability and functioning of 16S rRNA of 30S ribosome (1J5E). (a) The dimethylated adenines (DMA) in the "dimethyl A loop" of 16S rRNA help in the stabilization of the loop through stacking interactions and forms a hydrophobic pocket with 2MG. (b) The methylated 966 and 967 positions of 16S rRNA increase the surface area for stacking and also form a van der Waals contact with the hydrophobic portion of Arg-128 of S9 protein (not shown).

shown as RED_4, RED_8, RED_12, and RED_20 in Supporting Information Table 1. R.E.D. was also performed on the orientation obtained from the GAUSSIAN calculation, which is given as RED_1 in Supporting Information Table 2. The orientations used in RED_4 were retained in the RED_8 case and so on. We did not observe significant changes between the charge values in comparing each case. The standard deviation between the RED_20 to RED_1 is 0.0149e. Thus, the change observed in the charge values from a single orientation to multiple orientations is insignificant. To overcome the multiorientation effect on the charges, we decided to increase the grid size (the number of shells of points as well as the density of points per shell) in calculating the electrostatic potential. Different grid sizes were tested with Merz−Kollman charge fitting methodology as well as the CHELP-G method.[42] We calculated electrostatic potential with four different options. In the first case, we computed the electrostatic potential using four shells and with a density of one point per every square angstrom (MK). We increased the density of points to four per Å^2 in the second case (MK 4,4). We kept the density of ESP points the same and increased the number of shells from four to eight in the third case (MK 4,8) and used eight shells with a density of eight points per Å^2 (MK 8,8). No significant changes in charge values are observed as the number of shells or the density of ESP points are increased using Merz−Kollman charge fitting methodology. We also used the CHELP-G method with four shells and density of one point per Å^2 (ChelpG) and with eight shells and a density of eight points per Å^2 (ChelpG 8,8). We did not see any major fluctuations in the charge values from using MK vs ChelpG methods. The results are summarized in Supporting Information Table 2. Since the goal of the present study is to develop parameters for modified nucleosides that are consistent with Cornell et al.[26] force field, the protocol outlined in Cieplak et al.[29] (i.e., calculating ESP with four layers and a density of one point per Å^2) is sufficient to produce the atom-centered partial charges for the modified nucleosides present in RNA. Using the given optimized geometries (available at http://ozone3.chem.wayne.edu) to perform QM and RESP calculations as described, the charges reported herein can be readily reproduced.

5.1. Testing and Verifying the accuracy of Parameters. Once the parameters for the 107 known modifications in RNA were computed, they were incorporated into AMBER to test the stability of MD trajectories of RNA with modified nucleotides. A molecular dynamics simulation of yeast tRNA^{Phe} containing 14 different modifications[40] was carried out using the crystal structure 1EHZ.pdb for the starting coordinates. The parameters for all the 14 modified nucleosides were successfully incorporated into LEAP, which properly generated the topology and coordinate files for this highly modified tRNA. SANDER[15] was then used to do the energy minimization and molecular dynamics with two different methods: (a) use of implicit solvent with generalized-Born electrostatics[43] and (b) use of explicit solvent with particle-mesh Ewald electrostatics.[44,45] To test the significance of the presence of modifications in the stability and functioning of tRNA^{Phe}, we wanted to study the MD of tRNA^{Phe} without modifications. To build the unmodified version of tRNA^{Phe}, the modified nucleosides were replaced with their respective common nucleosides (e.g., DHU with uridine) using the RNA-123 software suite developed in our lab for the analysis of RNA structures as well as 3D structure prediction of RNA. We performed 1 ns generalized-Born simulations on both the modified as well as the unmodified tRNA^{Phe}. In the case of tRNA^{Phe} with all 14 modifications, the structure remains stable throughout a 1 ns simulation using generalized-Born implicit solvent dynamics (data not shown), implying that the parameters developed can be reliably used in AMBER for simulating RNA with modifications. Further studies on these two systems using explicit

solvent conditions as well as crystallographic conditions will definitely help in understanding the role of modifications in the stability and functioning of tRNA. We are in the process of studying the effects of these modifications in the stability of $tRNA^{Phe}$ by performing long time-scale AMBER molecular dynamics with explicit solvation on the structure with modifications and on the corresponding structure lacking modifications. Simulations of $tRNA^{Phe}$ in the crystalline environment with the periodic boundary conditions present in the crystal are also being done.

Thus far, several modifications have been successfully incorporated into RNA-123. With the availability of the geometries for these modified nucleosides, we were able to model all 12 known modifications in *E. coli* 16S rRNA into *T. thermophilus* 30S ribosome crystal structure (1J5E). Interestingly, all of the modifications were accommodated in the published PDB structure without any steric conflicts. Further, the placement of the modifications suggests functional roles for them in increased stacking or formation of hydrophilic pockets for protein binding (Figure 5). The PDB coordinates for the modified 16S rRNA are available at our group home page http://ozone3.chem.wayne.edu.

5.2. Implementation of the Modified Nucleotide Parameters in Other Force Fields. In addition to using the charges obtained from our study in generating parameters for AMBER, we have also used some of these charge values in CNS,[16] which is based on the CHARMM force field.[14] Parameter files for CNS can be created using programs such as PRODRG and XPLO-2D,[46] but these files do not contain charge values. We introduced the charge values into the parameter files of pseudouridine for CNS. These parameter files, having the charge values from this study, were used in NMR structure calculations for the 1060 hairpin loop of human 18S rRNA, which contains a single pseudouridine residue.[47]

5. Conclusions

We have successfully developed and implemented AMBER force field parameters for the 107 naturally occurring modified nucleosides present in RNA. As the evidence for the versatile functions of RNA in the cell is expanding, it is becoming apparent that modified nucleosides play important roles in achieving these functions. The availability of force field parameters for modified nucleosides enhances the functionality of AMBER and thereby will contribute to understanding how modified nucleosides participate in the function and structural stabilization of RNA. The modified nucleoside parameters described herein allow for AMBER MD simulations and molecular mechanics for all modified RNAs. Further the modular approach allows for many new combinations of base and/or sugar modified nucleotides to be readily computed.

Acknowledgment. The authors would like to thank Dr. Thomas Cheatham, III and Prof. David Case for useful discussions and suggestions. We thank Dr. Jason Sonnenberg for help with the GAUSSIAN calculations. We would also like to thank Larry Clos, Frederick Sijenyi, and Marcus Wood for critical analysis of the manuscript. We also thank Wayne State University Grid Computing Facility for the computational time and support. This work was supported by NIH grant GM073179.

Supporting Information Available: Effects of multiorientation on atom-centered partial charges and the influence of the number of shells and the density of points used in calculating the electrostatic potential. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Noller, H. F. RNA structure: reading the ribosome. *Science* **2005**, *309*, 1508−14.

(2) Sen, G. L.; Blau, H. M. A brief history of RNAi: the silence of the genes. *FASEB J.* **2006**, *20*, 1293−9.

(3) Krutzfeldt, J.; Stoffel, M. MicroRNAs: a new class of regulatory genes affecting metabolism. *Cell Metab.* **2006**, *4*, 9−12.

(4) Rozenski, J.; Crain, P. F.; McCloskey, J. A.; The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **1999**, *27*, 196−7.

(5) Henri Grosjean, R. B. *Modification and Editing of RNA*; ASM Press: 1998.

(6) Murphy, F. V. t.; Ramakrishnan, V.; Malkiewicz, A.; Agris, P. F. The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186−91.

(7) Sumita, M.; Desaulniers, J. P.; Chang, Y. C.; Chui, H. M.; Clos, L., 2nd; Chow, C. S. Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. *RNA* **2005**, *11*, 1420−9.

(8) Agris, P. F. Decoding the genome: a modified view. *Nucleic Acids Res.* **2004**, *32*, 223−38.

(9) Blount, K. F.; Uhlenbeck, O. C. The structure-function dilemma of the hammerhead ribozyme. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 415−40.

(10) Bruno, L. *Conserved ribosomal RNA modification and their putative roles in ribosome biogenesis and translation*; Springer Berlin/Heidelberg: 2005; Vol. 12, pp 263−284.

(11) Alexandrov, A.; Chernyakov, I.; Gu, W.; Hiley, S. L.; Hughes, T. R.; Grayhack, E. J.; Phizicky, E. M. Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* **2006**, *21*, 87−96.

(12) Lane, B. G.; Ofengand, J.; Gray, M. W. Pseudouridine and O2′-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins. *Biochimie* **1995**, *77*, 7−15.

(13) Auffinger, P.; Westhof, E. RNA solvation: A molecular dynamics simulation perspective. *Biopolymers* **2000**, *56*, 266−274.

(14) MacKerell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field,; M. J.; Fischer, S. G.; J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C. M.; S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B. S.; M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J. Y.; D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(15) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−88.

(16) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta. Crystallogr., Sect. D.: Biol. Crystallogr.* **1998**, *54*, 905−21.

(17) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.

(18) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23*, 1515−31.

(19) Halgren, T. A.; Damm, W.; Polarizable, force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−42.

(20) Lind, K. E.; Sherlin, L. D.; Mohan, V.; Griffey, R. H.; Ferguson, D. M. Parameterization and simulation of the physical properties of phosphorothioate nucleic acids. In *Molecular Modeling of Nucleic Acids;* Leontis, N. B.; SantaLucia, J., Eds.; ACS Symposium Series 682; American Chemical Society: Washington, DC, 1998; pp 41−54.

(21) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003**, *24*, 1016−25.

(22) Lind, K. E.; Mohan, V.; Manoharan, M.; Ferguson, D. M.; Structural, characteristics of 2′-O-(2-methoxyethyl)-modified nucleic acids from molecular dynamics simulations. *Nucleic Acids Res.* **1998**, *26*, 3694−3699.

(23) Stuart, J. W.; Koshlap, K. M.; Guenther, R.; Agris, P. F.; Naturally-occurring modification restricts the anticodon domain conformational space of tRNA(Phe). *J. Mol. Biol.* **2003**, *334*, 901−18.

(24) McCrate, N. E.; Varner, M. E.; Kim, K. I.; Nagan, M. C.; Molecular, dynamics simulations of human $tRNA^{Lys}$, 3 UUU: the role of modified bases in mRNA recognition. *Nucleic Acids Res.* **2006**, *34*, 5361−8.

(25) Auffinger, P.; Louise-May, S.; Westhof, E.; Molecular, dynamics simulations of solvated yeast tRNA(Asp). *Biophys. J.* **1999**, *76*, 50−74.

(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−97.

(27) Durant, P. C.; Bajji, A. C.; Sundaram, M.; Kumar, R. K.; Davis, D. R.; Structural, effects of hypermodified nucleosides in Escherichia, coli and human tRNALys, anticodon loop: The effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A. *Biochemistry* **2005**, *44*, 8078−89.

(28) Sierzputowska, G. H.; Sochacka, E.; Malkiewicz, A.; Kuo, K.; Gehrke, C. W.; Agris, P. F. Chemistry and structure of modified uridines in the anticodon, wobble position of transfer RNA are determined by thiolation. *J. Am. Chem. Soc.* **1987**, *109*, 7171−7177.

(29) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357−77.

(30) Frisch, M. J. T.; G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Scalmani, G.; Kudin, K. N.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Li, X.; Hratchian, H. P.; Peralta, J. E.; Izmaylov, A. F.; Brothers, E.; Staroverov, V.; Kobayashi, R.; Normand, J.; Burant, J. C.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Chen, W.; Wong, M. W.; Pople, J. A. *Gaussian DV, Revision E.05*; Gaussian, Inc.: Wallingford, CT, 2006.

(31) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−80.

(32) Steenken, S. Purine bases, nucleosides, and nucleotides: aqueous solution redox chemistry and transformation reactions of their radical cations and e- and OH adducts. *Chem. Rev. (Washington, DC, U.S.)* **1989**, *89*, 503−20.

(33) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−9.

(34) U. Chandra Singh; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(35) Woods, R. J.; Pell, W. M. K.; Moffat, S. H.; Smith, V. H., Jr. Derivation of net atomic charges from molecular electrostatic potentials. *J. Comput. Chem.* **1990**, *11*, 297−310.

(36) Pigache, A.; Cieplak, P.; Dupradeau, F.-Y. In *Automatic, highly reproducible and effective RESP and ESP charge Derivation: Application to the development of programs RED and X RED*, 227th ACS National Meeting, Anaheim, CA, March 28−April 1, 2004; Anaheim, CA, March 28−April 1 2004.

(37) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−60.

(38) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−74.

(39) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Romby, P.; Caillet, J.; Springer, M.; Rees, B.; Ehresmann, C.; Ehresmann, B.; Moras, D. The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **1999**, *97*, 371−81.

(40) Shi, H.; Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* **2000**, *6*, 1091−1105.

(41) NC-IUB, Nomenclature for incompletely specified bases in nucleic acid sequences Recommendations 1984. *Eur. J. Biochem.* **1985**, *150*, 1−5.

(42) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.

(43) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(45) York, D. M.; Darden, T. A.; Pedersen, L. G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald, and truncated list methods. *J. Chem. Phys.* **1993**, *99*, 8345−8348.

(46) Kleywegt, G. J. *XPLO2D, 051002*; Department of Cell and Molecular Biology, Uppsala University: Uppsala, SWEDEN.

(47) Spahn, C. M.; Beckmann, R.; Eswar, N.; Penczek, P. A.; Sali, A.; Blobel, G.; Frank, J.; Structure of the 80S ribosome from Saccharomyces cerevisiae−tRNA-ribosome and subunit-subunit interactions. *Cell* **2001**, *107*, 373−86.

CT600329W