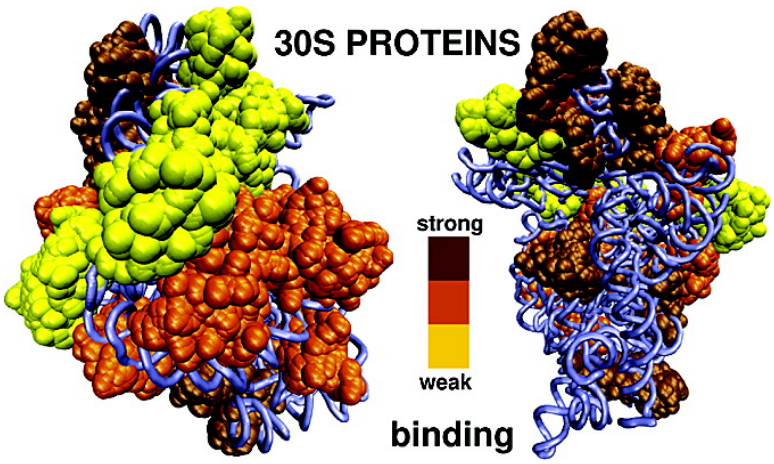
Article

Exploring Assembly Energetics of the 30S Ribosomal Subunit Using an Implicit Solvent Approach

Joanna Trylska, J. Andrew McCammon, and Charles L. Brooks

J. Am. Chem. Soc., **2005**, 127 (31), 11125-11133• DOI: 10.1021/ja052639e • Publication Date (Web): 16 July 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 07/16/2005

Exploring Assembly Energetics of the 30S Ribosomal Subunit Using an Implicit Solvent Approach

Joanna Trylska,*[,†,‡,§] J. Andrew McCammon,[†,‡,||] and Charles L. Brooks III[‡,⊥]

Contribution from the Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0365, Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0374, Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University, 02-106 Warsaw, Poland, Howard Hughes Medical Institute and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365, and Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received April 22, 2005; E-mail: joanna@icm.edu.pl

Abstract: To explore the relationship between the assembly of the 30S ribosomal subunit and interactions among the constituent components, 16S RNA and proteins, relative binding free energies of the *T. thermophilus* 30S proteins to the 16S RNA were studied based on an implicit solvent model of electrostatic, nonpolar, and entropic contributions. The late binding proteins in our assembly map were found not to bind to the naked 16S RNA. The 5′ domain early kinetic class proteins, on average, carry the highest positive charge, get buried the most upon binding to 16S RNA, and show the most favorable binding. Some proteins (S10/S14, S6/S18, S13/S19) have more stabilizing interactions while binding as dimers. Our computed assembly map resembles that of *E. coli*; however, the central domain path is more similar to that of *A. aeolicus*, a hyperthermophilic bacteria.

Introduction

Ribosomes are ribonucleoprotein assemblies that are responsible for the translation of the genetic code. They are composed of small and large asymmetric subunits, named according to their sedimentation coefficients, 30S and 50S for bacterial ribosomes. The subunits are held together by a number of intermolecular noncovalent interactions. The 30S subunit consists of the 16S ribosomal RNA and 21 proteins (named S1, S2, etc.) which contain globular domains with extended loops or tails buried deep inside the ribosomal RNA. Other molecules bind during ribosome activity, including messenger RNA and transfer RNA. The small subunit's most important role is to maintain the translation fidelity.

Since it was found that the *E. coli* small ribosomal subunit can reassemble in vitro from the 16S RNA and a mixture of the 30S proteins[1,2] and form an active particle, the pathway and the mechanism of the assembly have been of significant interest (for a review see ref 3). The in vitro small subunit complexation is a sequential and ordered process that occurs in a cooperative manner. Therefore, all the information needed for the small subunit assembly is present in the ribosomal RNA and protein components. The assembly map of Nomura and co-workers[1,2] gives the order of association of each protein to 16S RNA, dividing the proteins into primary (bind directly and independently to the 16S RNA), secondary (need at least one of the primary proteins to be bound to 16S RNA prior to associating), and tertiary binders (require at least one protein each from the primary and secondary binding set to be able to bind). Also, a map of the assembly kinetics has been determined,[4] which classifies the proteins into early, middle, middle-late, and late binders (see Figure 1). The kinetic map suggests that the assembly proceeds roughly from the 5′, through central, to the 3′ domain of 16S RNA even though in vitro this process is not coupled with transcription. Three major pathways may be indicated in the assembly process: the S4/S8 pathway, the S15 pathway, and the S7 pathway. The pathways correspond to the 16S RNA domain classification with the S4/S8 pathway including proteins binding to the 5′ domain, the S15 pathway including proteins binding to the central domain, and the S7 pathway including proteins that bind to the 3′ domain. The pathways also confirm that the RNA domains fold independently as separate entities in a 5′ to 3′ direction. These maps serve as a model of the ordered assembly of *E. coli* 30S subunits; however, one has to be aware that this binding classification is somewhat arbitrary and was based on the experiments performed

† Department of Chemistry and Biochemistry, University of California at San Diego.
‡ Center for Theoretical Biological Physics, University of California at San Diego.
§ Warsaw University.
|| Howard Hughes Medical Institute and Department of Pharmacology, University of California at San Diego.
⊥ The Scripps Research Institute.

(1) Mizushima, S.; Nomura, M. *Nature (London)* **1970**, *226*, 1214−1218.
(2) Held, W. A.; Ballou, B.; Mizushima, S.; Nomura, M. *J. Biol. Chem.* **1974**, *249*, 3103−3111.
(3) Culver, G. M. *Biopolymers* **2003**, *68*, 234−249.
(4) Powers, T.; Daubresse, G.; Noller, H. F. *J. Mol. Biol.* **1993**, *232*, 362−374.

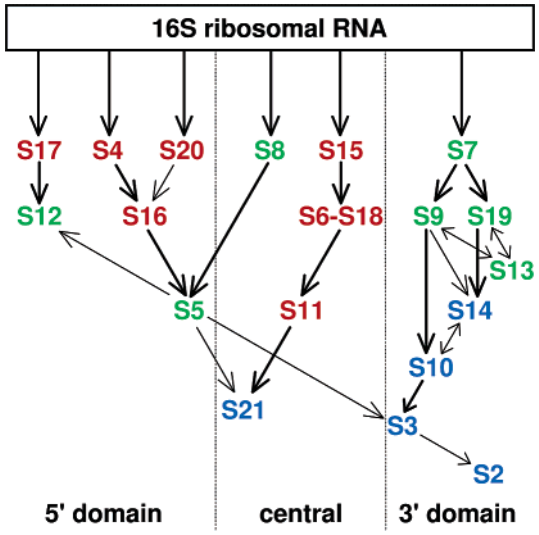


Figure 1. In vitro assembly kinetics map of the *E. coli* 30S ribosomal subunit based on the work of Powers et al.[4] The early kinetic class of proteins are colored in red, middle and middle-late, in green, and late or delayed binders, in blue. Arrows indicate interactions among proteins from the studies of the Nomura laboratory[2] and based on the review of G. Culver.[3] Proteins are shown interacting with their appropriate 16S RNA 5′, central, and 3′ domains. Top row proteins correspond to primary binders, tertiary proteins are in blue, and all the remaining proteins were classified as secondary in the Nomura map.

in vitro for the mixture of 16S RNA and 30S proteins. Moreover, the division depends on the kind of experimental data taken into account. For example, S11 was classified as a secondary binding protein in the sequence map[2] and as an early one in the kinetics map.[4] Comparing the kinetic with the sequence assembly maps, only the late and tertiary binding proteins are the same. This suggests that there may be multiple paths of the small subunit assembly. It has been recently shown[5] that the central domain assembly path for the hyperthermophilic bacterium *A. aeolicus* is not identical to that of *E. coli*. Therefore, the assembly paths may also differ among the species.

The information which emerged from the crystal structures of the small ribosomal subunit confirmed that, overall, the 30S proteins interact exclusively with individual 16S RNA domains. For the protein-RNA contact map, see ref 6. The two proteins that represent exceptions and have significant dual interaction are the S20, which binds to the 5′ and 3′ domains, and S16, which binds to the 5′ and central domains. Two proteins have been identified as assembly initiators: S4 of the 5′ domain and S7 of the 3′ domain. They are both primary binding proteins, but S4 is classified as an early one and S7, as a middle one in the kinetics-based map of *E. coli*.[4]

There have been a large number of experimental studies devoted to the process of the 30S assembly (for details see review of G. Culver[3] and references therein). However, due to the size and complexity of the system, theoretical approaches have not been numerous.[6] The aim of this work was to study the relative binding free energies of the proteins forming the 30S complex with 16S RNA and to investigate to what extent the continuum models with empirical parameters applied to a rigid structure relate to the experimentally determined order or kinetics of binding. The results could help determine if the assembly process, which is known to be cotranscriptional, is dependent on the 16S RNA folding and predict the differences in assembly maps for *E. coli* and *T. thermophilus*. Furthermore, we investigated which proteins favor binding as dimers and which can bind as separate entities. We calculated the binding free energy differences with respect to binding to the naked 16S RNA and to the whole 30S subunit and to various intermediates in the domain assembly paths.

The binding free energy of each protein was estimated using the PB/SA method, by solving the Poisson−Boltzmann (PB) equation in continuum solvent and by calculating the amount of solvent accessible surface area (SASA) buried upon binding. The values of configurational, translational, and rotational entropy costs upon complexation were estimated based on the empirical formulas and experimental binding affinities available for proteins S4, S7, S8, and S15.[7−11]

The question of how the ribonucleoprotein complexes assemble has been a matter of intensive studies for many years, yet many details of the assembly remain unclear. This problem remains of great importance because it has recently been found that aminoglycoside antibiotics such as paromomycin and neomycin inhibit not only the translation itself but also the 30S subunit formation.[12] The question of whether we are able to explain the small subunit assembly based on the static crystal structure and relative binding free energy calculations performed with the extended PB/SA model was raised and explored in this work.

Methods

Binding Free Energy Calculations. Rigorous, fully atomistic simulations to obtain the absolute binding free energies of the 30S proteins are not currently feasible, due to the large size of the system. Therefore, we applied an implicit solvation model and determined the relative binding free energies for all the proteins binding to 16S RNA in the small subunit. Even with the implicit solvent approach, the huge size of the subunit enables us to perform the calculations only for a single conformation, and we do not account directly for the flexibility of the binding species.

Following Froloff et al.,[13] the free energy of association of two molecules may be written in the form

$$\Delta G^{\mathrm{bind}}_{\mathrm{calc}} = \Delta G_{\mathrm{elec}} + \Delta G_{\mathrm{np}} + \Delta G_{\mathrm{strain}} - T\Delta S_{\mathrm{conf}} - T\Delta S_{\mathrm{trans+rot}} \quad (1)$$

where ΔG_{elec} is the electrostatic contribution to binding, ΔG_{np} is the nonpolar term, $T\Delta S_{\mathrm{conf}}$ describes the loss of configurational main- and side-chain entropy upon complexation, $T\Delta S_{\mathrm{trans+rot}}$ represents the loss of rigid-body translational and rotational degrees of freedom and a possible change in vibrational motions, and $\Delta G_{\mathrm{strain}}$ represents the reorganizational cost and distortions upon complexation of the binding species. In this study, the electrostatic component was calculated using the continuum PB model (for details of the theory see refs 14 and 15), and the nonpolar term was estimated based on the surface area model.

(5) Recht, M. I.; Williamson, J. R. *J. Mol. Biol.* **2004**, *344*, 395−407.
(6) Stagg, S. M.; Mears, J. A.; Harvey, S. C. *J. Mol. Biol.* **2003**, *328*, 49−61.
(7) Gerstner, R. B.; Pak, Y.; Draper, D. E. *Biochemistry* **2001**, *40*, 7165−7173.
(8) Rassokhin, T. I.; Golovin, A. V.; Petrova, E. V.; Spiridonova, V. A.; Karginova, O. A.; Rozhdestvenskii, T. S.; Brosius, J.; Kopylov, A. M. *Molecular Biology* **2001**, *35*, 527−535.
(9) Tishchenko, S. V.; Vassilieva, J. M.; Platonova, O. B.; Serganov, A. A.; Fomenkova, N. P.; Mudrik, E. S.; Piendl, W.; Ehresmann, C.; Ehresmann, B.; Garber, M. B. *Biochemistry (Moscow)* **2001**, *66*, 1165−1171.
(10) Gruber, T.; Kohrer, C.; Lung, B.; Shcherbakov, D.; Piendl, W. *FEBS Lett.* **2003**, *549*, 123−128.
(11) Recht, M. I.; Williamson, J. R. *J. Mol. Biol.* **2001**, *313*, 35−48.
(12) Mehta, R.; Champney, W. S. *Antimicrob. Agents Chemother.* **2002**, *46*, 1546−1549.
(13) Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−1301.

The nonpolar term includes the differences in the formation of the repulsive cavity in the solvent and the differences in the solute−solvent van der Waals dispersion interactions between separate molecules and the complex. A linear dependence of ΔG_{np} on the change in the solvent accessible surface area ($\Delta SASA$) buried upon binding was assumed with γ, the microscopic surface tension, as the proportionality coefficient. The nonpolar term always favors binding, but its magnitude depends on the values of γ used.

The absolute binding free energies are impossible to obtain, due to the approximate semiempirical approach and also because it is not trivial to evaluate correctly all the entropic contributions. However, analyzing relative binding energies for different proteins associating with 16S RNA is sufficient for the kind of studies we perform. We estimate one of the entropic terms, the protein side-chain conformational entropy loss upon binding to 16S RNA, based on the empirical scale of Pickett and Sternberg.[16] To do so, we calculate the number of protein residues that get buried upon binding and assume a linear correspondence with the side-chain entropy loss.

Also, due to the size and complexity of the system, it is not possible to accurately estimate each of the other entropic contributions, translational, rotational, vibrational, and the protein and RNA reorganization energy. They sum up to a substantial value; therefore, all other missing terms are grouped into a constant which is fitted, together with other parameters, to experimental data obtained for proteins S4, S7, S8, and S15.[7−10] Experimentally determined values of $K_d = 1/K_a$ and the relation $\Delta G = -RT \ln K_a$, where R is the gas constant, T, temperature, and K_a, the apparent association constant, were used to calculate ΔG_{expt} and to perform the fitting.

The final estimate of the binding free energy used in this study is given by the following formula:

$$\Delta G_{calc}^{bind} \approx \Delta G_{elec} + \gamma \Delta SASA + B \cdot NR_{buried\ res} + C \qquad (2)$$

where $NR_{buried\ res}$ is the number of residues buried upon binding for each protein, and B and C are constants. The parameters were subject to least-squares fit to experimental data for four proteins (see next section), and the values obtained for the dielectric constant of 4 are $\gamma = 0.058$ kcal/mol·Å^2, $B = 0.82$, and $C = 120$ kcal/mol. A similar value of γ was also applied before in the literature.[13,17] The value of B is in the range of the empirical scale of Pickett and Sternberg,[16,18] where the loss of configurational entropy of the side chain upon folding contributes between 0.5 and 2 kcal/mol to the free energy change, depending on the residue type.

Experimental Data on Binding Affinity. The apparent dissociation constant (K_d) of 6.6 nM for the binding of the S8 protein from *T. thermophilus* to 16S RNA 42 nucleotide fragment was measured.[9] For the same protein a similar value of 3.5 nM was obtained by different authors.[10] The affinity of S4 for ribosomal RNA from another thermophilic organism, *B. stearothermophilus*, is in the nanomolar range as well.[7] The hybrid of S7 *T. thermophilus* protein with an *E. coli* 16S RNA fragment gives a K_d of 36 nM.[8] Other studies[11] obtained K_d equal to 4.9 nM for the binding of S15 protein from *A. aeolicus* to the whole 16S ribosomal RNA. S15 is from a different species, but it binds to the same domain. Early studies for *E. coli* obtained values of K_d in the range $10^{-7}-10^{-8}$ M for proteins S4, S7, S8, S15, S17, and S20.[19] However, a 10-fold higher affinity was shown for binding of thermophilic proteins to their 16S RNA in comparison with their mesophilic counterparts.[10] This suggests that all the protein-RNA interactions are at least 10-fold stronger for the thermophilic species than for the mesophilic organisms, which function at lower temperatures.

Preparation of the System. The conformation of the 30S subunit from *T. thermophilus* obtained to 3.05 Å resolution[20] was taken from the Protein Data Bank (entry code 1j5e). This particular structure was chosen because of the best available resolution and the fact that it was crystallized as a native 30S complex without any bound ligands. It does not contain the S1 protein, but omission of S1 does not reduce the ribosome function or prevent the 30S assembly.[21] The structure contains proteins S2 to S20, which correspond to those of *E. coli*, and a small peptide Thx. The structure of *T. thermophilus* lacks protein S21.

The structure was protonated using the HBUILD[22] facility of CHARMM.[23,24] The positions of the hydrogens were energy-minimized with 1000 steps of the steepest descents method. The C- and N-terminal residues, Asp, Glu, Lys, and Arg were kept charged. The net charge of His and all other protein residues were set to 0. The two zinc atoms were assigned a radius of 1.4 Å and a +2e charge. The unknown atoms present in the structure which were positioned in close proximity to the 16S RNA were treated as divalent ions and assigned a radius of 1.73 Å and a charge of +2e, representing the net charge of magnesium. $\Delta SASA$ was calculated with the *acc* utility of the APBS package[25] using the CHARMM27 van der Waals radii set, with a probe sphere radius of 1.4 Å.

Electrostatic Calculations. The electrostatic contribution, ΔG_{elec}, was calculated by solving the linearized form of the PB equation with the Adaptive Poisson−Boltzmann Solver[25] for the complex and binding molecules and by taking the difference in their electrostatic energies. We have previously shown that it is an accurate approach and gives similar results as when calculating the solvation and Coulombic contributions separately but is less time-consuming.[26] Partial charges were assigned according to the CHARMM27 force field. Two sets of atomic radii to describe the solute−solvent boundary were examined: those from CHARMM27 and the radii developed by Nina et al.[27] for amino acids and Banavali and Roux[28] for nucleotides (further on referred to as the Nina and Banavali set). The latter set of radii was parametrized with CHARMM partial charges to best reproduce the solvation energies of single amino acids and nucleotides obtained by free energy perturbation and molecular dynamics simulations with explicit solvent. To eliminate numerical errors, for each PB calculation the same grid center and grid spacing were used. The size of the 30S complex is 210 × 180 × 200 Å^3, and the grid dimensions were set to 360 × 300 × 340 Å^3 and the PB equation was solved in parallel on 300 NPACI Data Star or 800 Blue Horizon processors, respectively, to 0.2 Å resolution. This high resolution is needed for the grid-based PB energy values to converge. The solution of the nonlinear PB equation for this highly charged system is available only to 0.6 Å grid resolution which is too coarse for binding energy calculations. However, the effect of the nonlinear term on the calculated order of binding was subject to tests. The electrostatic contributions obtained from the solution of the linear and nonlinear PB equation show similar trends apart from proteins S10 and S16, and the differences are less than 10%. It does not change the relative characteristics of binding and the order of binding of proteins, and because we study relative binding energies, even for such

(14) Sharp, K. A.; Honig, B. *Ann. Rev. Biophys. Chem.* **1990**, *19*, 301−332.
(15) Baker, N. A.; McCammon, J. A. In *Structural Bioinformatics*; Weissig, M.; Bourne, P. E., Eds.; John Wiley & Sons: New York, 2003; p 427.
(16) Pickett, S. D.; Sternberg, M. J. E. *J. Mol. Biol.* **1993**, *231*, 825−839.
(17) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281−296.
(18) Doig, A. J.; Sternberg, M. J. E. *Protein Sci.* **1995**, *4*, 2247−2251.
(19) Schwarzbauer, J.; Craven, G. R. *Nucleic Acids Res.* **1981**, *9*, 2223−2237.
(20) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M., Jr.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327−339.
(21) Nomura, M. *Science* **1973**, *179*, 864−873.
(22) Brunger, A. T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 148−156.
(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(24) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *The Encyclopedia of Computational Chemistry*, v. R. Schleyer, P., et al., Eds.; John Wiley & Sons: Chichester, 1998; Vol. 1, pp 271−277.
(25) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(26) Konecny, R.; Trylska, J.; Tama, F.; Zhang, D.; Baker, N. A.; Brooks, C. L., III; McCammon, J. A. Submitted.
(27) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 5239−5248.
(28) Banavali, N. K.; Roux, B. *J. Phys. Chem.* **2002**, *106*, 11026−11035.

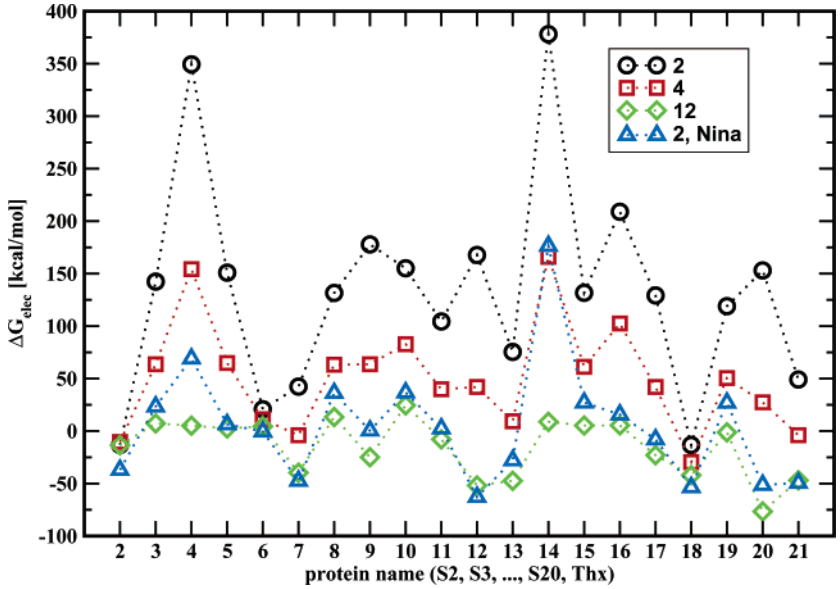


Figure 2. Electrostatic contribution to binding energies of the 30S *T. thermophilus* proteins for various dielectric constants assigned to the solute interior. Calculations with $\epsilon = 2$, 4, and 12 were for the CHARMM27 charge and radii set and with $\epsilon = 2$ for CHARMM27 charges and the Nina and Banavali radii set (see Methods). *X* axis number denotes S protein name (S2,S3,...,S20); number 21 denotes Thx.

a highly charged system, the linearized approximation of the electrostatic contribution seems sufficient. The somewhat surprising utility of the linearized PB equation for highly charged solutes has a theoretical basis in the two-fluid model.[29] Moreover, the linearized form of the PB equation has been successfully applied before in the studies of binding of aminoglycosides to the small ribosomal subunit.[30] To define the dielectric boundary, the molecular surface definition was used with the CHARMM radii set and van der Waals surface definition for the Nina and Banavali set, in accord with the conditions of the parametrization of the latter.[27,28] For the boundary conditions, the focusing method was applied.[31,32]

Energy Dependence on Parameters. The sensitivity of the electrostatic contribution on the parameters used in the solution of the PB equation was also examined. Doubling the value of the ionic strength from $I = 150$ mM to $I = 330$ mM or changing the temperature from 25 to 40 °C changes the electrostatic term by less than 3%. The removal of the 180 explicit ions located in proximity of the phosphates in the 16S RNA makes the electrostatic contribution for the 30S proteins between 15 and 30% less favorable; therefore, one needs to include the explicit ions to decrease the strong repulsion between the binding entities. The positions of the ions were determined in the whole 30S subunit; however, we were unable to include contributions from any ions that were released during the formation of the complex. Based on the above tests and to reproduce most closely the physiological conditions, in all further calculations the explicit divalent ions were used, the ionic strength was set to 150 mM, and the temperature was set to 298 K.

The biggest effect in determining the ΔG_{elec} comes from the change in the interior dielectric constant (ϵ). Therefore, we performed calculations for three values of ϵ (2, 4, and 12) with the CHARMM radii set and for $\epsilon = 2$ for the Nina and Banavali radii set. The $\epsilon = 2$ is believed to account for the electronic polarizability, and $\epsilon = 4$, to account additionally for the flexibility and structural reorganization of the systems. Some authors suggest much higher values of $\epsilon = 10-12$ to be used in the case of proteins.[33−35] The dielectric constant of the exterior solvent was set to 78.5 in all of the calculations.

To test the sensitivity of the binding energy results on the placement of the dielectric boundary, two definitions of the surface were used: the molecular surface obtained with a probe radius of 1.4 Å and the surface enclosed by the van der Waals radii. Two sets of radii were used: the CHARMM radii and the radii developed by Nina et al.[27] and Banavali and Roux.[28] Although the results depend on the placement of the dielectric boundary and the surface definition, we are not interested in this study in absolute binding free energies but in their relative values for the whole set of proteins. Based on numerous tests we chose to perform electrostatic calculations with CHARMM27 charges and radii and the molecular surface definition and also with CHARMM27 charges with Nina et al.[27] and Banavali and Roux[28] radii with a van der Waals surface definition. We believe that this particular way of calculating the electrostatic contribution, together with other energy terms, enables us to study the macroscopic and relative properties of the assembly. Such an approach is sufficient and serves well for the purposes of comparing the electrostatic contribution for different proteins.

Results and Discussion

One of the contributions to the binding free energy of 30S proteins to 16S RNA originates from electrostatic interactions. Figure 2 shows the dependence of the electrostatic contribution (ΔG_{elec}), obtained by solving the PB equation as the difference for the complex and each protein and 16S RNA separately, on dielectric constants assigned to the solute. Even though the electrostatic contribution determined by the PB continuum model is very sensitive to the interior dielectric constant, it shows similar characteristics. Because we were able to fit the microscopic surface tension parameter to estimate the nonpolar contribution, it is enough that the ΔG_{elec}, calculated with different dielectric constants, shows similar trends for all the studied proteins to enable us to compare the relative binding energies for various proteins. It is worth noting that, in most

(29) Lau, A. W. C.; Pincus, P. *Phys. Rev.* **2002**, *66*, 041501.
(30) Ma, C.; Baker, N. A.; Joseph, S.; McCammon, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 1438−1442.
(31) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327−335.
(32) Yang, A. S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 252−265.
(33) Simonson, T.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1996**, *118*, 8452−8458.
(34) Grycuk, T. *J. Phys. Chem. B* **2002**, *106*, 1434−1445.
(35) Dominy, B. N.; Minoux, H.; Brooks, C. L., III. *Proteins* **2004**, *57*, 128−41.

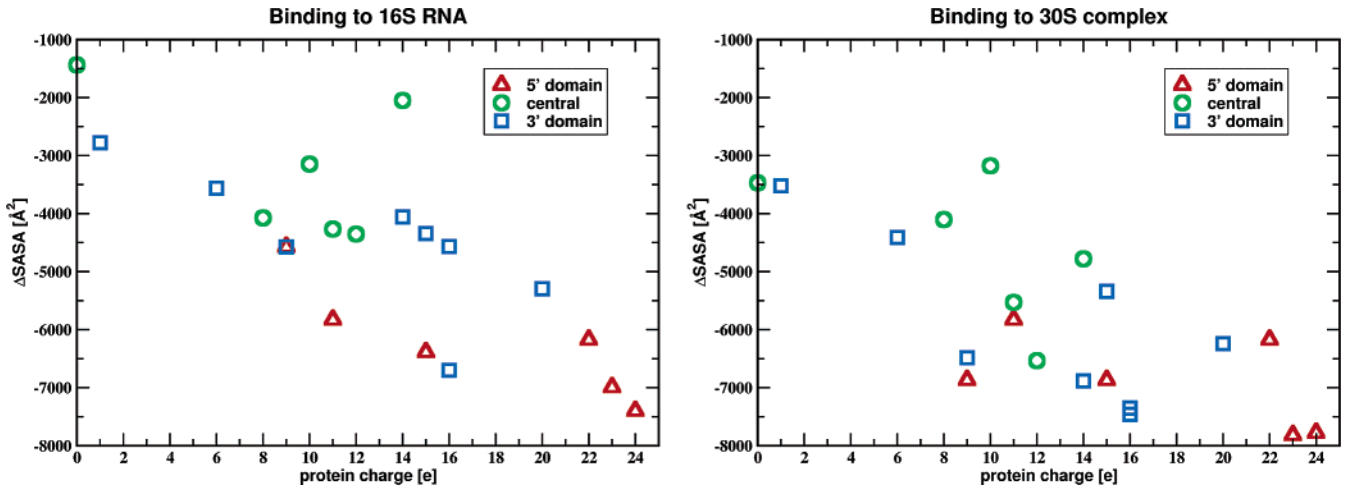


Figure 3. Dependence upon protein charge of the change in SASA upon binding of proteins to the naked 16S RNA and to the whole 30S complex.

Table 1. Calculated Relative Free Energies of Binding (kcal/mol) of the 30S *T. thermophilus* Proteins to 16S RNA and to the Entire 30S Complex Obtained from the Solution of the Linearized PB Equation, with a Dielectric Constant of 4 and CHARMM27 Charges and Radii Set with Molecular Surface Definition[a]

protein	ΔG^{elec}_{calc} to 16S RNA	ΔG^{elec}_{calc} to 30S	$\Delta SASA$ to 16S RNA	$\Delta SASA$ to 30S	NR to 16S RNA	NR to 30S	ΔG^{bind}_{calc} to 16S RNA	ΔG^{bind}_{calc} to 30S
5' domain								
S17	42	53	-6387	-6869	84	90	-140	-151
S20	27	39	-6174	-6174	86	86	-140	-129
S4	154	210	-6992	-7819	120	136	-33	-12
S16	103	127	-5829	-5829	83	83	-48	-23
S12	42	60	-7399	-7777	97	102	-188	-188
S5	65	128	-4573	-6868	80	122	-15	-50
central domain								
S15	61	68	-4074	-4104	55	56	-10	-4
S8	63	94	-4353	-6537	72	108	-10	-77
S6	11	26	-1441	-3473	21	52	65	-12
S18	-30	-7	-2051	-4780	27	62	-6	-113
S11	40	59	-4270	-5531	68	87	-32	-71
3' domain								
S7	-4	19	-4343	-5341	66	82	-82	-104
S9	64	125	-6700	-7461	104	115	-120	-93
S19	51	87	-3563	-4412	47	58	2	-2
S13	9	62	-5295	-6241	70	83	-120	-112
S2	-10	0	-2780	-3520	54	66	-7	-30
S3	64	104	-4057	-6886	70	120	6	-77
S10	83	135	-4286	-6485	57	86	1	-51
S14	166	196	-4568	-7352	47	60	60	-62
THX	-4	26	-3149	-3178	24	24	-47	-18
Dimers								
S10/S14	237	317	-8387	-11029	116	153	-34	-77
S13/S19	61	100	-8660	-8757	123	124	-220	-186
S6/S18	-28	-66	-3245	-6199	56	106	-50	-219

[a] For comparison, proteins are colored according to their assembly kinetics class for *E. coli* with early binders in red, middle or middle-late, in green, and late, in blue. For the definition of column labels see Methods. ΔSASA units are Å^2.

cases, ΔG_{elec} opposes binding. It is more favorable for proteins to interact with the high dielectric solvent, which has a strong ability to reorganize, than with RNA which is more constrained and cannot adjust that much to a protein. That the electrostatic contribution may oppose binding has been noticed before for smaller ligands.[13,36] However, in our case ΔG_{elec} is expected to be much larger in magnitude for proteins than for small ligands.

The nonpolar contribution depends on the amount of SASA buried upon complexation of the molecules. Figure 3 shows the dependence of SASA buried upon binding to the naked 16S RNA and to the 30S complex on the charge of the protein. On average, a roughly linear relationship is evident, meaning that the higher the charge of the 30S protein, the more buried it gets upon binding. Also, the proteins binding to the 5′ domain show an overall higher change in SASA when binding to 16S RNA. The lowest change in SASA upon binding is associated with proteins attaching to the central domain. It is interesting to note that all the proteins get buried to some extent upon binding to 16S RNA, meaning that the tertiary proteins do not bind on the top of the primary or secondary ones but all of them interact with RNA. Moreover, all proteins, except one binding to 16S RNA, bear a positive charge, and the 5′ domain proteins carry on average the highest positive charge.

Table 1 shows the comparison of calculated free energies of binding to the naked 16S RNA and to the entire small subunit. We chose to present the results for $\epsilon = 4$, but similar relative

(36) Baginski, M.; Fogolari, F.; Briggs, J. M. *J. Mol. Biol.* **1997**, *274*, 253−267.

binding is seen with $\epsilon = 2$. We find that the Nina and Banavali radii set and $\epsilon = 12$ do not give an overall qualitative description of binding of 30S proteins which bears resemblance to the assembly maps. The numbers in Table 1 may be only treated as relative, and they do not represent the absolute binding free energies. They give an estimate of which proteins have the most favorable binding and which would not bind alone to a rigid folded 16S RNA structure. We emphasize that the numbers are only a qualitative measure of the binding affinities of each protein. The proteins are categorized according to the domain they bind to. Because proteins S20 and S16 interact with two 16S RNA domains, they are categorized in Table 1 according to the first domain they contact.

Based on our calculations, all the proteins show overall favorable binding to the already formed 30S complex. This is expected and in accord with experiments, because each protein in our model binds to a stable complex with a perfectly folded 16S RNA and in the presence of all other proteins. Late or tertiary binders, S2, S3, S10, and S14, strongly prefer to bind in the presence of all other proteins. S3, S10, and S14 would not bind to the folded but naked 16S RNA. Proteins S4, S15, S16, and S20 both early and/or primary prefer to bind to 16S RNA. Overall, proteins binding to the 5′ domain show the most favorable binding to 16S RNA. Regarding the middle kinetic class proteins, we cannot distinguish them based on this study. In Table 1, S17, even though it is a primary/early binding protein, prefers to bind to the complex rather than to 16S RNA. It has been shown experimentally for *E. coli* subunits that the amount of S17 bound when all other proteins are present in the mixture is higher in comparison to its binding to the naked 16S RNA[2]. Therefore, such may also be the case for *T. thermophilus* species. Moreover, several experimental studies for *E. coli* have shown that most of the proteins would bind, even some only weakly, to the naked 16S RNA. In Table 1, where we present the results for *T. thermophilus* proteins, we see a similar behavior.

Electrostatic contributions oppose binding to a higher extent for all the proteins binding to the whole 30S complex compared with binding to the naked 16S RNA. This is intuitive, since nearly all the proteins that form the 30S subunit are positively charged. It is reasonable to anticipate that their electrostatics oppose binding less for the naked RNA, which is more negatively charged, than for the complex that already contains other proteins. However, the loss of SASA upon binding is higher if proteins were to bind to the complex. Hence, there are two competing contributions which drive the assembly process.

Thx is a small protein characteristic of *T. thermophilus*; thus it was not classified in the in vitro studies on the *E. coli* complex. It resides deep inside the RNA chain and interacts directly only with the 16S RNA in the 3′ domain. Its position in the 3′ domain and the fact that it gets completely buried upon binding, together with the 5′ to 3′ polarity of folding and assembly, suggests it might be a late binding protein.

Early studies of the 30S ribosomal subunit reconstitution found a 21S intermediate which undergoes a rate-limiting unimolecular reaction.[37] The proteins that are required to form this functionally inactive reconstitution intermediate (RI) particle are the following: S4, S8, S16, S7, S15, S17, S11, S18, S9, S19, S5, and S12, with the latter three present only in small amounts. Due to a temperature-dependent RI → RI* transition, the in vitro assembly process possesses a high activation energy. In our simulations, apart from S19, the RI proteins show favorable binding. However, it is surprising that S18 is needed in this intermediate and S6 is not because early experiments have shown that these proteins bind as a dimer in the assembly map.[2,37] This intermediate does not involve tertiary binding proteins S2, S3, S10, and S14, which according to our computations show either no binding or very weak interactions.

The few proteins that show no or weak binding in our calculations were investigated further because from the experimental data it was suspected that they might bind as dimers. In the case of the central domain assembly path, Table 1 predicts no binding of S6 to 16S RNA. In the early assembly map, the binding of S6 could not be observed even in the presence of other sets of proteins.[2] It was detected that, in the *E. coli* subunits, proteins S6 and S18 bind together,[2,4] and recently a similar heterodimer was shown for hyperthermophilic bacteria of *A. aeolicus*.[5,11] The results presented in Tables 1 and 2 show that binding of S6/S18 is favorable but similar both to 16S RNA and S15−16S RNA complex. S15 interacts directly only with 16S RNA; therefore, the change in SASA and in the number of residues buried upon binding of S6/S18 to the naked 16S RNA and to the S15−16S RNA are similar. The *E. coli* experimental assembly map predicts an enhancement of S6/S18 binding to S15−16S RNA in comparison with the 16S RNA alone (Figure 1). But in the central domain assembly path of *A. aeolicus*, which has the extremely high optimal growth temperature of 95 °C, this enhancement is not seen.[5] The crystal

Table 2. Calculated Relative Free Energies of Binding (kcal/mol) of the 30S *T. thermophilus* Proteins Obtained from the Solution of the Linearized PB Equation, with a Dielectric Constant of 4 and CHARMM27 Charges and Radii Set with Molecular Surface Definition[a]

protein	ΔG^{elec}_{calc}	$\Delta SASA$	NR	ΔG^{bind}_{calc}
5′ Domain				
S5 to S8−16S RNA	82	−5892	105	−54
S5 to S4−16S RNA	95	−5322	95	−16
S5 to S16S4−16S RNA	96	−5322	95	−15
S16 to S4−16S RNA	111	−5829	83	−39
S16 to S20−16S RNA	111	−5829	83	−39
S12 to S17−16S RNA	82	−7743	101	−202
Central Domain				
S6/S18 to S15−16S RNA	−27	−3267	56	−51
S8 to S15−16SRNA	66	−4353	72	−8
S15 to S8−16SRNA	64	−4074	55	−7
S11 to S15−16SRNA	41	−4271	68	−31
S11 to S6−16S RNA	41	−4271	68	−31
S11 to S18−16S RNA	50	−4990	79	−55
S11 to S6S18−16S RNA	51	−4990	79	−54
S11 to S6S18S15−16SRNA	52	−4990	79	−53
3′ Domain				
S10 to S7S9−16S RNA	103	−4516	60	10
S14 to S7S9−16S RNA	176	−4598	48	68
S14 to S7S19−16S RNA	170	−4568	47	63
S19 to S7−16S RNA	51	−3563	47	3
S9 to S7−16S RNA	73	−7157	111	−131
S13 to S7−16S RNA	10	−5295	70	−120
S13 to S7S9−16S RNA	14	−5363	71	−119
S13 to S7S9S19−16S RNA	34	−6211	83	−139
S3 to S10S14−16S RNA	83	−6462	113	−80

[a] For the definition of column labels see Methods. ΔSASA units are Å^2.

(37) Held, W. A.; Nomura, M. *Biochemistry* **1973**, *12*, 3273−3281.

structure we are utilizing comes from *T. thermophilus*, also a thermophilic species, which differs from that of *E. coli*, and the *T. thermophilus* bacteria has an optimal growth temperature of 75 °C, twice as high as that of *E. coli*. Therefore, differences between the species are not unlikely.

In our calculations, in the central domain we see direct binding of S11 to the naked 16S RNA, which is not in accord with the experimental *E. coli* assembly map. However, direct binding of S11 is seen for *T. thermophilus* and *A. aeolicus*[5] which agrees with our results. Table 2 shows that the presence of S6 and S15 does not potentiate the binding of S11, but S18, S6/S18 and S6/S18/S15 enhance S11 binding. Reth and Williamson[5] observed only weak cooperativity between the binding of S11 and other central domain proteins. In *A. aeolicus* assembly, binding of S8 is independent of other central domain proteins. We see no enhancement in binding of S8 from S15 and vice versa.

It has also been anticipated that the late binding proteins S10 and S14 of the 3′ domain may bind interdependently (see Figure 1). In our computations, S10 and S14 alone do not show binding to the naked 16S RNA, neither to S7S9−16S RNA nor to S7S19−16S RNA complexes. However, S10/S14 are capable of binding together to the naked 16S RNA and also to S7S9− and S7S19−16S RNA complexes (Tables 1 and 2). Because they are late binding proteins, each of them shows strong binding to the 30S complex. Also, if the S10/S14 dimer was to bind to the 30S complex, it would bind more strongly than S10 or S14 separately. These proteins are in close proximity to each other in the crystal structure sharing an 1780 Å^2 surface area.

In the 3′ domain experimental assembly path of *E. coli* (Figure 1), the binding of S9, S13, and S19 is interdependent. S13 and S19 probably bind together[3] even though it was shown that S13 can bind to 16S RNA alone.[2,38] Therefore, we studied this path in more detail. Our simulations confirm binding of S13 but do not predict binding of S19 alone to either 16S RNA or S7−16S RNA (Tables 1 and 2). If one takes S13/S19 together, the binding of both becomes much more favorable. The same applies to the S9/S19 complex. Moreover, strong S13/S19 binding is also predicted to the 30S complex and to S7−16S RNA. In the 30S subunit, S13 interacts directly only with S19 and vice versa, which may also explain the interdependent binding. S9 binds stronger to S7−16S RNA than to the naked 16S RNA. This is in accord with the assembly map which classifies S7 as the primary binder to the 3′ domain. S13, however, does not show any difference in binding to the naked 16S RNA and to the S7−16S RNA or to S7S9−16S RNA complex. A slight enhancement of binding abilities is seen for binding to the S7S9S19−16S RNA complex, most probably due to its favorable interaction with S19.

It has been found that a single mutation in the S5 protein disrupts assembly.[39,40] It is possible that S5 may play a role in the central domain rearrangements. This protein is in contact with the S4 and S8 proteins, and its binding is dependent on S8 in the assembly map. Indeed, in our calculations S5 binds more tightly to the S8−16S RNA complex than to 16S RNA alone (Table 2). Contrary to the assembly map of *E. coli*, we do not see any enhancement of binding of S5 coming from S16 or S4. The S16 protein from *T. thermophilus* does not form contacts with any other 30S proteins but only with 16S RNA. We also do not see that either S4 or S20 increase the binding free energy of S16. Our computations also indicate that in the 5′ domain S17 facilitates the binding of S12.

Table 3. Difference in the Change of the Solvent Accessible Surface Area of the 30S *T. thermophilus* Proteins upon Binding to the 16S RNA Alone and to the Entire Complex (30S Proteins Plus 16S RNA)[a]

protein	charge [e]	$\Delta\Delta SASA$	interactions
primary binders			
S4	+23	827	S5
S7	+15	998	S9, S11
S8	+12	2184	S5, S2, S12, S17
S15	+8	30	–
S17	+15	482	S12, S8
S20	+22	0	–
secondary binders			
S5	+9	2295	S4, S8
S6	0	2032	S18
S9	+16	761	S7
S11	+11	1261	S7, S18
S12	+24	378	S17, S8
S13	+20	946	S19
S16	+11	0	–
S18	+14	2729	S6, S11
S19	+6	849	S13
tertiary binders			
S2	+1	740	S8
S3	+14	2829	S10, S14
S10	+9	2199	S14, S3
S14	+16	2784	S10, S3
Thx	+10	29	–

[a] Main neighbors contributing to this $\Delta\Delta SASA$ [Å^2] are named for each protein. For comparison, proteins are colored according to the kinetic assembly class for *E. coli* with red as early, green as middle or middle−late, and blue as late binders.

The measure of nonpolar interactions with other proteins was investigated. Table 3 shows the difference in the amount of SASA buried upon binding to 16S RNA alone and to the 30S complex. There are only four proteins for which this difference is negligible; they have no direct protein neighbors in the bound form and interact directly only with 16S RNA. These are S20, S16, S21, and S15, two of which are primary binders. For the proteins that bind to the 5′ domain, this measure of interaction is generally the smallest. For proteins binding to the central and 3′ domains, this measure is on average higher. The difference in $\Delta SASA$ between binding to 16S RNA and to the 30S complex is over 2000 Å^2 for proteins S3, S6, S18, S10, and S14. This might be one explanation as to why these proteins tend to form dimers. The high value of $\Delta\Delta SASA$ in Table 3 confirms earlier suggestions that S3, S10, and S14 form a cluster held by hydrophobic interactions.[20]

All except one of the proteins forming the 30S subunit carry a positive charge (Table 3 or Figure 3). At least for one of the proteins which favor binding in a dimer, this positive charge is not high. For example, the net charge of S6 is zero, and that of S19, +6*e* which is small in comparison with the net charge of most of the 30S proteins. Also, we see no correlation between the charge of the protein and the sign and magnitude of the electrostatic contribution (compare Figure 2 and Table 1).

(38) Zimmermann, R. A.; Muto, A.; Fellner, P.; Ehresmann, C.; Branlant, C. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 1282−1286.
(39) Guthrie, C.; Nashimoto, H.; Nomura, M. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *63*, 384−391.
(40) Nashimoto, H.; Held, W.; Kaltschm, E.; Nomura, M. *J. Mol. Biol.* **1971**, *62*, 121−138.

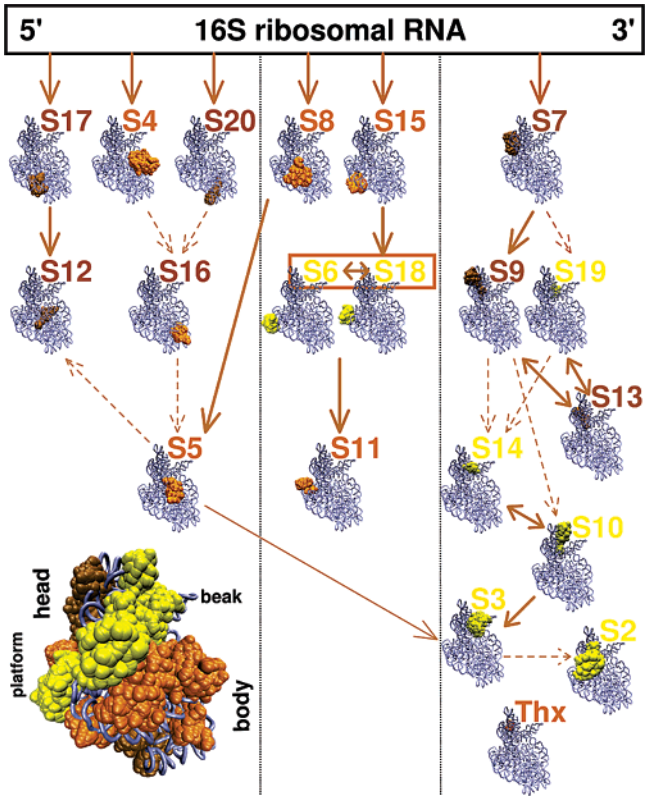


Figure 4. Binding affinity assembly map of the *T. thermophilus* 30S ribosomal subunit based on the binding free energy calculations. Strong binders are shown in dark brown, middle, in orange, and weak, in yellow (colors based on their binding free energies to the naked 16S RNA presented in Table 1). Insets denote the positions of appropriate proteins in the 30S subunit looking from the solvent side. Filled brown arrows indicate detected increase in the binding free energy due to the presence of the protein the arrow originates from. Brown dashed arrows indicate no increase in the binding free energy and are shown to facilitate comparison with the *E. coli* assembly map of Figure 1.

Our calculations are summarized in Figure 4 where we gather the results into a computationally derived binding affinity map for *T. thermophilus*. While the experimental assembly map based on kinetics informs us of how rapidly certain proteins bind, it does not provide direct evidence as to the strength of those interactions. Studies of binding free energies give a different perspective and classification of proteins providing information based on how strongly proteins are bound. Nevertheless, comparison of these two maps (Figures 1 and 4) shows them to be quite similar. It is interesting to note that the proteins with the highest binding affinity bind mostly from the 50S side of the 30S subunit. Also, a tendency for clustering of middle and weak binders from the solvent side of the small subunit is noted and displayed in the lower inset of Figure 4.

Conclusions

We have performed binding energy calculations for the assembly of the twenty small ribosomal subunit proteins with the 16S RNA utilizing the PB implicit solvent model. The PB implicit solvent approach is the only one currently feasible for such a large system. Nevertheless, it suffers from many limitations. One is the single conformation approximation which does not account for the dynamic reorganization of the atomic charges in the rigid system. The other is the choice of parameters such as dielectric constant, partial charges, van der Waals radii, and the placement of the dielectric boundary. Also, the quality of the crystal structure affects the value of the electrostatic contribution. The electrostatic contribution is obtained by solving the linearized PB equation, which in the case of a highly charged system is an approximation. However, the nonlinear solution converges only to a 0.6 Å grid resolution which is too coarse to obtain reasonable results. Other limitations arise from the availability of the experimental data. The assembly maps were obtained for the *E. coli* system, and we utilize the crystal structure of *T. thermophilus*, which for example lacks protein S21 and has twice as high an optimal growth temperature. Very few association constants were determined for the proteins of *T. thermophilus*, and the kinetics of only primary binders to fragments of 16S RNA was studied. From the crystal structure we are able to estimate the binding free energy to the already folded 16S RNA, and in the experiments in which the proteins were added, this might have not always been the case.

The electrostatic contribution to the binding free energy depends strongly on the definition of the dielectric boundary used in the PB calculations. Also, implicit solvent approaches usually overestimate the binding and give more negative values than in reality.[13] Therefore, only the relative binding energies are meaningful and may be analyzed. Additionally, one has to be aware that even though we are using the same 16S RNA conformation, the proteins for which we are comparing the binding energies are different in terms of size and weight, and we estimate some of the energy contributions by a similar constant for every protein. Moreover, it was shown that the binding of the ribosomal proteins in the case of thermophilic organisms may be at least 10-fold higher in affinity compared to the *E. coli* which functions at lower temperatures.[10] Therefore, the range of binding energies that we obtain may be much larger in absolute values.

Even with these limitations of the model, we found reasonable correspondence between predicted and observed early and late binders, as well as general aspects of the assembly paths (compare, for example, Figure 1 and Figure 4). The middle kinetic class of the assembly map is not well delineated by our calculations.

For the majority of the 30S proteins, we find that the calculated electrostatic contribution is positive, meaning that it opposes binding. The proteins lose their favorable electrostatic interactions with the solvent, and they are not compensated by the interactions with the ribosomal RNA. All proteins, with one exception, that bind to 16S RNA carry a net positive charge, and the proteins that bind to the 5′ domain tend to have the highest net charge. Also, there is a trend that the higher the charge of the protein, the more that protein gets buried upon binding. Therefore, proteins binding to the 5′ domain have on average the biggest change in SASA upon complexation and the largest number of residues that get buried upon binding. The nonpolar term favors binding; therefore, if the binding of 30S proteins is coupled with folding, the more favorable contribution should be from the 5′ terminal of 16S RNA.

A majority of the proteins while binding to the 16S RNA alone have a smaller change in SASA than when binding to the whole 30S complex, in which they contact other proteins as well. Proteins S20, S16, S21, and S15 are the only ones that get approximately similarly buried while binding to the 16S RNA alone and to the 30S complex. Therefore, the nonpolar

term for all other proteins would be more favorable if they were to bind to the whole complex and not to the naked 16S RNA.

Simulations show favorable binding in the presence of other 30S proteins. Late and tertiary proteins show no or weak binding to the naked 16S RNA. The early binders which bind to the 5′ domain show on average the most favorable binding to the 16S RNA alone. We also see that proteins that bind from the 50S contact side of the small subunit tend to have the highest binding affinity.

Two proteins from the tertiary binding set (S10 and S14) have positive binding free energies and most probably favor binding as a dimer. S6 and S18 also dimerize; however, according to the calculations S18 is capable of weak binding to 16S RNA itself. Considering the fact that S6 bears no net charge, it is likely that its binding to 16S RNA is accompanied by another charged protein. This is reflected in the favorable binding with S18. The nonpolar contribution for these two proteins (S6 and S18) is much more favorable if they were to bind to the 30S complex and not to the 16S RNA alone. Because they bind to the central domain, they are likely to be among the middle binders. Binding of S19 to 16S RNA alone and to S7−16S RNA is not seen. On the other hand, if S19 is complexed with S13, the S13/S19 dimer favors binding.

By looking at which 16S RNA domain each of the proteins binds, we find that three of the primary binders bind to the 5′ domain, two to the central domain, and only one protein to the 3′ domain. None of the proteins classified as tertiary in *E. coli* assembly maps binds to the 5′ domain. Four tertiary binding proteins associate with the 3′ domain, suggesting that the 30S complex is formed during folding of 16S RNA and that the order of binding is also coupled with folding. Proteins binding to the 5′ domain have on average the least direct interactions with other proteins in the complex.

In our modeling, the central domain assembly path of *T. thermophilus* bears more resemblance to that of *A. aeolicus* than of *E. coli*. This is reasonable since the optimal growth temperature of *T. thermophilus* is closer to that of hyperthermophilic *A. aeolicus* than to *E. coli*.

We interpreted the binding energy results as a qualitative and relative measure of binding because absolute values are difficult to obtain with this model. Other approaches that account for the flexibility of the system are not computationally feasible for such a large macromolecular complex.

Acknowledgment. We thank Dr. Megan Trevathan for her initial work on characterizing the assembly kinetics, which served as a catalysis for our study. Also, Professor Jamie Williamson is acknowledged for sharing data and discussions. Our work is supported in part by the NIH (GM31749 for JAM, RR12255 for CB), NSF (MCB-0071429 for JAM, PHY0216567 for CB), HHMI, W.M. Keck Foundation, SDSC, NBCR, CTBP, and Accelrys Inc. J.T. was also supported by the Polish State Committee for Scientific Research (115/E-343/BST-993/ICM/2004 and 115/E-343/ICM/BST-1076/2005) and by European CoE MAMBA (QLRI-CT-2002-90383). The figures were prepared with VMD,[41] sodipodi, and xmgrace.

JA052639E

(41) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.