2950–2960 *Nucleic Acids Research, 2002, Vol. 30, No. 13* *© 2002 Oxford University Press*

Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome

Franca Fraternali* and Luigi Cavallo[1]

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and [1]Dipartimento di Chimica, Universitá di Salerno via Salvador Allende, I-84081 Baronissi (SA), Italy

Received February 1, 2002; Accepted March 26, 2002

ABSTRACT

We present a new method for the calculation of solvent accessible surface areas at the atomic and residue levels, which we call parameter optimized surfaces (POPS-*A* and POPS-*R*). Atomic and residue areas (the latter simulated with a single sphere centered at the C$^{\alpha}$s atom for amino acids and at the P atom for nucleotides) have been optimized versus accurate all-atoms methods. We concentrated on an analytical formula for the approximation of solvent accessibilities. The formula is simple, easily derivable and fast to compute, therefore it is practical for use in molecular dynamics simulations as an approximation to the first solvation shell. The residue based approach POPS-*R* has been derived as a useful tool for the analysis of large macromolecular assemblies like the ribosome, and is especially suited for use in refinement of low resolution structures. The structures of the 70S, 50S and 30S ribosomes have been analyzed in detail and most of the interactions within the subunits and at their interfaces were clearly identified. Some interesting differences between 30S alone and within the 70S have been highlighted. Owing to the presence of the P-tRNA in the 70S ribosome, localized conformational rearrangements occur within the subunits, exposing Arg and Lys residues to negatively charged binding sites of P-tRNA. POPS-*R* also allows for estimates of the loss of free energy of solvation upon complex formation, particularly useful in designing new protein–RNA complexes and in suggesting more focused experimental work.

INTRODUCTION

In recent years the life sciences have been faced with an explosion of freely available biological information from both gene sequencing and structure determination (1). New research areas like structural genomics (2,3) and functional genomics (4,5) are flourishing at the same time as structural investigation methods are addressing macromolecular assemblages like the RNA polymerases (6), the nucleosome (7) and the ribosome (8–13). In an era in which quantity and quality of data is increasing tremendously, the development of efficient and automated computer-based tools for the analysis and the rationalization of such data is mandatory.

More than 40 years ago, Kauzmann (14) identified the burial of hydrophobic groups as a key driving force for protein folding. Recent analyses revealed that protein–protein interactions are also associated with a significant burial of hydrophobic residues (15,16), and thermodynamic unfolding data showed that ~75% of the variation in the energetics can be accounted for burial of solvent accessible surface area (SASA) (16). Moreover, the evaluation of SASA was demonstrated to play a pivotal role in methods for assigning folds to sequences (17), function to structures (5) and for describing the dynamic behavior of proteins in solution (18,19) as well as their unfolding modes (20,21).

Here, we present a new method to calculate solvent accessibilities at the atomic and residue levels for proteins and nucleic acids, called parameter optimized surfaces (POPS). The analytical formula we used to approximate the SASA (22) was proved to be particularly effective when combined with molecular dynamics (MD) programs (18,21). In order to improve the performance of the formula, we reparametrized it to reproduce atomic (POPS-*A*) and residue (POPS-*R*) SASAs from more accurate methods (23) that do not use an analytical expression. This versatile approach allows us to implement the POPS areas in MD programs and as a weighting factor in structural alignment, threading and structural refinement packages.

The residue level approach is particularly useful for the analysis of large structural assemblies, in order to filter key interactions between molecules due to the burial of surface area. Moreover, it has been designed to characterize the hydrophobic or hydrophilic nature of exposed and buried surfaces and can estimate the loss of solvation free energy upon complex formation. The recently solved structure of the *Thermus thermophilus* 70S ribosome at residue level (C$^{\alpha}$ only and P only for proteins and RNAs) (12) is the ideal candidate to demonstrate the efficiency and the predictive power of POPS-*R*. The 70S ribosome is composed of a small and a large subunit, named 30S and 50S, respectively. The X-ray crystal structure of the 30S subunit comprises 16S rRNA and about 20 proteins (from S2 to S20), while that of the 50S subunit comprises 23S and 5S rRNAs, and about 30 proteins (from L1 to L30). The structure of the 30S isolated subunit was resolved to atomic level (8,10) and it will be used to validate the POPS-*R*

*To whom correspondence should be addressed. Tel: +44 208 959 3666; Fax: +44 208 913 8545; Email: ffranca@nimr.mrc.ac.uk

Table 1. NACS SASAs, in Å^2, and percentual errors with the Hasel, POPS-*A* and POPS-*R* approaches for the molecules in the parameters training set

PDB Code	Descr[a]	NACS SASA	Hasel[b] Error	POPS-*A*[c] Error	POPS-*R*[d] Error
1bfmA	α-P	5896	12	-5	-15
1ehs	α-P	3709	17	3	0
1hry	α-P	7149	11	2	-2
1hyp	α-P	4806	26	3	1
1lrv	α-P	12811	38	4	2
1rfbA	α-P	10937	11	-4	3
1unkA	α-P	5538	29	3	9
1utg	α-P	5198	20	-4	-6
2gfl	α-P	4001	21	-4	-3
2occD	α-P	12658	11	-4	-4
1zymA	α-P	7737	36	5	-3
1aqt	α-P	4766	43	10	-2
1ajz	αβ-P	12741	58	16	11
1avv	αβ-P	5966	31	2	3
1b4v	αβ-P	8289	39	8	3
1geo	αβ-P	8099	44	12	0
1htm	αβ-P	9220	13	-8	-8
1plr	αβ-P	12714	51	11	0
1rot	αβ-P	6898	41	14	0
1tsg	αβ-P	5394	13	-11	-16
1uae	αβ-P	8748	61	15	1
1vil	αβ-P	7584	30	9	-4
1yat	αβ-P	5871	37	5	-4
2bnh	αβ-P	18643	71	23	-2
2dnj	αβ-P	10317	64	16	18
2hgf	αβ-P	6087	22	-0	-4
2vik	αβ-P	7604	26	6	-7
4jdw	αβ-P	13795	72	18	21
9mht	αβ-P	6464	38	15	10
1ata	β-P	5144	14	10	-3
1bw3	β-P	7302	34	16	3
1dlc	β-P	10185	41	10	9
1eur	β-P	13779	62	21	10
1fgp	β-P	4367	17	-3	-9
1gof	β-P	14436	62	17	7
1hcd	β-P	7569	30	13	2
1kap	β-P	10217	49	12	-2
1pdr	β-P	5277	36	8	0
1qcv	β-P	3676	15	-2	4
1tsp	β-P	25484	49	12	-2
1vie	β-P	3958	16	-1	1
1xnb	β-P	7808	47	12	7
2mob	β-P	6089	21	-2	-18
2pec	β-P	13275	63	17	7
5hir	β-P	3389	19	8	-2
1e27	β-P	8013	-21	3	3
1aaf	irr-P	6306	4	6	3
1tiv	irr-P	6427	12	2	16
2ech	irr-P	3416	14	1	-1
2rel	irr-P	4784	11	3	-1
1trlA	P	4104	19	-4	-1
1ehz	RNA	12685	-21	4	-2
1gid	RNA	46789	-15	0	0
2tra	RNA	12216	-21	1	-1
359d	RNA	6989	-23	-4	8
379d	RNA	7034	-24	-5	8
1a60	RNA	8013	-21	3	3
103d	DNA	4294	-19	-2	2
1a83	DNA	3347	-10	6	-4
1da3	DNA	3827	-21	-4	1
352d	DNA	14710	-4	15	-16
467d	DNA	4225	-28	-4	-10
163d	P/RNA	5544	-5	7	18
163d_nu[e]	RNA	5553	-20	4	15
163d_pr[f]	P	2009	0	-8	19
1di2	P/RNA	14818	5	1	-7
1di2_nu[e]	RNA	8172	-30	-6	-5
1di2_pr[f]	αβ-P	8785	25	1	-9
1ec6	P/RNA	14847	18	7	6
1ec6_nu[e]	RNA	8360	-26	2	-11
1ec6_pr[f]	P	10289	30	3	0
1f6u	P/RNA	7640	1	5	7
1f6u_nu[e]	RNA	3750	-26	1	-3
1f6u_pr[f]	Irr-P	5383	5	1	2
1hdw	P/RNA	29659	18	3	-3
1hdw_nu[e]	RNA	5939	-27	4	-5
1hdw_pr[f]	αβ-P	24731	26	2	-3
1hp6	P/RNA	45803	-3	4	2
1hp6_nu[e]	RNA	38313	-19	1	-2
1hp6_pr[f]	αβ-P	11052	40	9	-2
1urn	P/RNA	21567	29	8	9
1urn_nu[e]	RNA	12735	-26	-2	-12
1urn_pr[f]	αβ-P	14189	52	11	9
1hcr	P/DNA	6738	10	-5	7
1hcr_nu[e]	DNA	5030	-17	-6	5
1hcr_pr[f]	α-P	4481	10	-3	14
2stw	P/DNA	11935	15	14	3
2stw_nu[e]	DNA	6632	-16	3	0
2stw_pr[f]	α-P	7218	27	13	3

[a]α-P, mainly α protein; β-P, mainly β protein; irr-P, mainly irregular protein, according to CATH's definition; P, undetermined class protein; P/DNA, protein/DNA complex; P/RNA, protein/RNA complex.

[b]Hasel percentage error calculated as $100 \times (SASA^{Hasel} - SASA^{NACS})/SASA^{NACS}$.

[c]POPS-*A* percentage error calculated as $100 \times (SASA^{POPS\text{-}A} - SASA^{NACS})/SASA^{NACS}$.

[d]POPS-*R* percentage error calculated as $100 \times (SASA^{POPS\text{-}R} - SASA^{NACS})/SASA^{NACS}$.

[e]X_nu, nucleic acid part of the protein/nucleic acid complex with PDB code X.

[f]X_pr, proteic part of the protein/nucleic acid complex with PDB code X.

method. Here we will describe key interactions among complexes in the ribosome detected by POPS-*R* and some, as yet unreported, conformational changes due to the 30S/50S association and to the interaction with P-tRNA.

MATERIALS AND METHODS

Reparametrizing the analytical formula

The total SASA of a molecule composed by *N* atoms is given by:

$$SASA = \sum^{N} A_i ,$$

where A_i is the SASA of the *i*th atom.

The algorithm we used to approximate the A_i is based on the analytical expression proposed by Still and co-workers (22,24) and on the probabilistic method of Wodak and Janin (25). The original formula is:

$$A_i(r^N) = S_i \prod_{i=1}^{N^{atoms}} 1 - \frac{p_i p_j b_{ij}(r_{ij})}{S_i},$$

where $S_i = 4\pi(R_i + R_{solv})^2$ is the SASA of the isolated atom i with radius R_i and a solvent probe with radius R_{solv}.

The term $b_{ij}(r_{ij})$ represents the SASA removed from S_i by the overlap of the atoms i and j at a distance $r_{ij} = |r_i - r_j|$.

If $r_{ij} > R_i + R_j + 2R_{solv}$

$b_{ij}(r_{ij}) = 0;$

while if $r_{ij} < R_i + R_j + 2R_{solv}$

$$b_{ij}(r_{ij}) = \pi(R_i + R_{solv})(R_i + R_j + 2R_{solv} - r_{ij})[1 + (R_j - R_i)r_{ij}^{-1}]$$

The empirical parameter p_i depends on the atom type, while the empirical parameter p_{ij} serves as an additional reducing factor that distinguishes between first and next covalently bound neighbor atoms ($p_{1,2}$ and $p_{1,3}$, respectively) and non-covalently bound atoms ($p \geq_{1,4}$). These parameters were optimized by Hasel *et al.* (22), reproducing the exact SASA of a large number of small molecules.

While in the original parametrization of Hasel *et al.* the p_i parameters depend on the atom hybridization and substitution [e.g. CH(sp3), NH(sp2) and so on, for a total of about 25 parameters], we reparametrized the formula by choosing the parameters p_i dependent on the type of atom in a given residue (e.g. one p_i for the C^β of each standard amino acid, or one p_i for the N1 of each nucleotide, for a total of about 250 parameters) and by splitting off the $p \geq 1,4$ connectivity parameter into two parameters, namely $p_{1,4}$ and $p_{>1,4}$. Moreover, we applied the same algorithm to approximate the A_i at residue level, which means that each amino acid and nucleotide is represented by a single sphere centered at the C^αs atom for amino acids and at the P atom for nucleotides. In this case each parameter p_i corresponds to one amino acid or nucleotide, and R_i is the radius of the sphere that simulates the whole residue, for a total of about 30 parameters

Both the POPS-*A* and POPS-*R* empirical parameters were optimized over the atomic or residue SASAs of a database of 89 specifically chosen biological molecules (proteins, nucleic acids and protein–nucleic acid complexes). The SASA of the atoms of these molecules were evaluated with the program Naccess (NACS) in Hubbard *et al.* (23), and constitute the POPS-*A* training dataset of about 120 000 atoms. The residue NACS SASAs were obtained by adding up the atomic NACS areas, and constitute the POPS-*R* dataset of about 12 000 residues.

The POPS-*A* SASAs of the atoms were fitted to the NACS SASAs through a minimization of the σ^2 variance of POPS-*A* from NACS areas with respect to the empirical parameters p_i and p_{ij}. The atom radii proposed by Hasel *et al.* (22) were adopted. For the POPS-*R* parametrization the same procedure was applied by using the areas in the residues dataset. Besides the parameter p_i, for each residue the radius R_i of the sphere

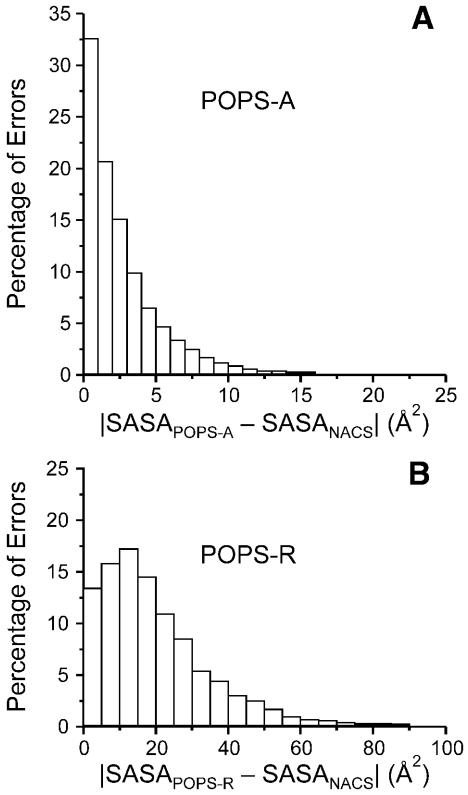

Figure 1. Percentual distribution of absolute atomic and residue SASAs errors with respect to NACS ones for POPS-*A* (**A**) and POPS-*R* (**B**). In (B) the NACS areas are calculated at atomic level and summed up to residues.

used to simulate the whole amino acid or nucleotide was also optimized in the fitting procedure.

To assess the predictive power of POPS-*A* and POPS-*R* we followed a cross-validation-like resampling procedure (26–29). The size of the datasets (120 000 atoms and 12 000 residues areas for POPS-*A* and POPS-*R*, respectively) prevented the leave-one-out cross-validation resampling. The natural choice was to perform a *k*-fold cross-validation resampling with $k = 89$, which is the number of molecules in the database. The SASAs in both the atomic and residue datasets were partitioned into 89 subsets, each of them containing the SASAs of the atoms (or residues) of a specific molecule. Then, the p_i and p_{ij} parameters were fitted to a training set composed by 88 subsets, and were used to predict the SASAs of the omitted subset that represents the test molecule, TM. The predicted atomic (or residues) SASAs of the omitted test molecule, A_i^{TM}, were compared with the corresponding NACS SASAs, A_i^{NACS}. The cross-validation prediction error e_i^{NACS} for the test molecule i was evaluated as:

$$e_i^{NACS} = \frac{1}{N}\sum_{j=1}^{N} |A_j^{TM} - A_j^{NACS}|,$$

where N corresponds to the number of atoms (or residues) in the test molecule i. This procedure was repeated 89 times, each time leaving out one of the test molecules (subsets), and the single cross-validation prediction errors e_i^{NACS} were averaged to obtain the overall cross-validation prediction error e_{CV}^{NACS}:

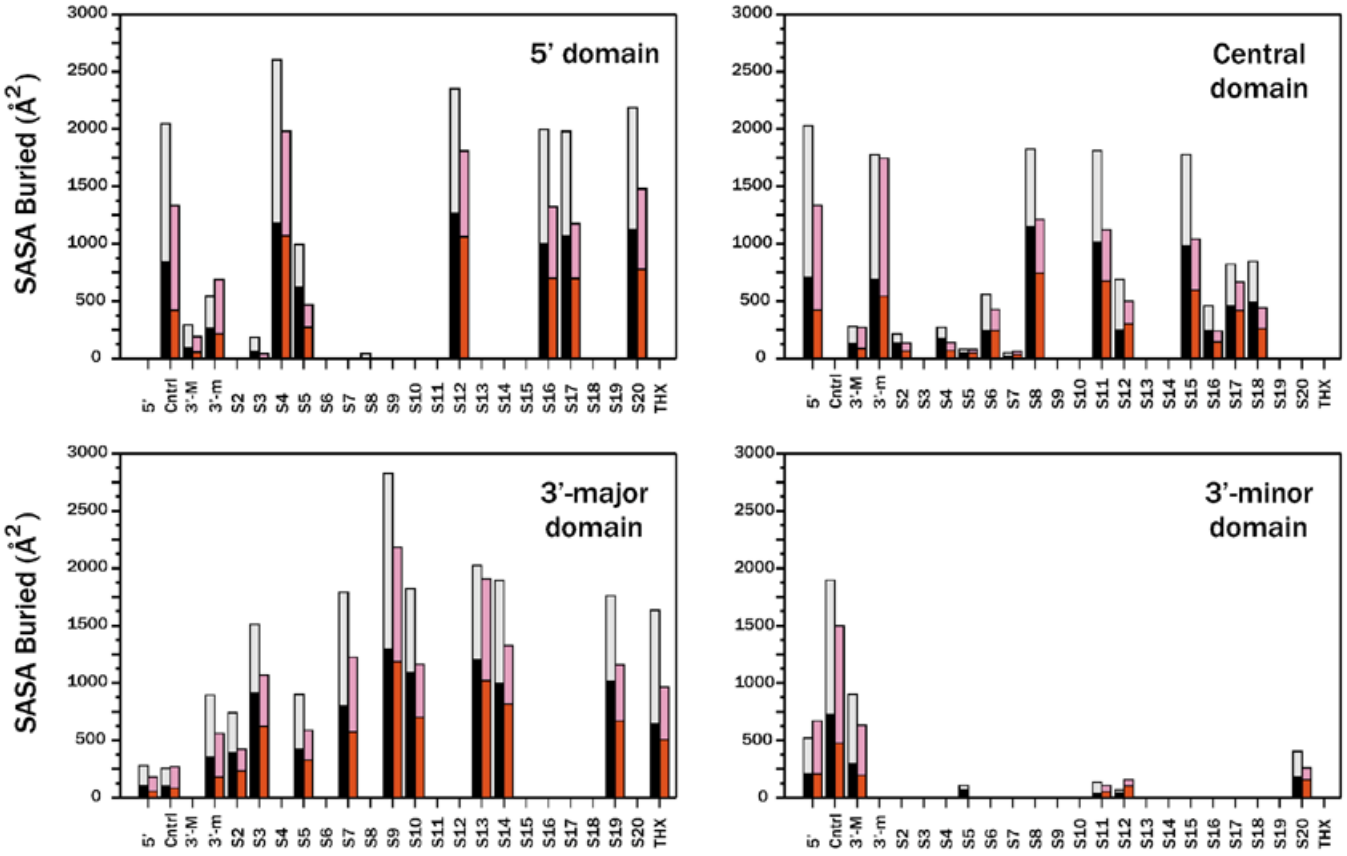

Figure 2. Naccess and POPS-*R* SASAs buried by the overlap between different components of 30S from the high resolution structure (10). Black and gray bars correspond to the Naccess hydrophobic and hydrophilic SASAs, respectively, while red and pink bars are the corresponding POPS-*R* SASAs. Cntrl, 3′-M and 3′-m indicate the central, 3′-major and 3′-minor domains, respectively.

$$e_{CV}^{NACS} = \frac{1}{89}\sum_{j=1}^{89} e_i^{NACS} \quad .$$

To establish whether the p_i and p_{ij} parameters were converged, we evaluated the e_{CV}^{FD} error, which corresponds to the e_{CV}^{NACS} error, but the reference areas are those obtained through a fitting to the full dataset, A_i^{FD}, i.e. including all the 89 molecules.

Finally two parameters will often be mentioned as measures of the errors obtained from the fitting: (i) the average absolute error *aae* and (ii) the average absolute precentual error *aape*.

NACS was also used to calculate the average fraction of hydrophobic and hydrophilic contributions to the total SASA for each residue in the database. According to these average fractions the POPS-*R* SASA can be partitioned into hydrophobic and hydrophilic contributions even witha residue-based method.

Finally, the free energy of solvation loss upon complex formation was estimated as $\Delta G_{solv}^{loss} = -\Delta SASA^{Phob}\sigma^{Phob} - \Delta SASA^{Phil}\sigma^{Phil}$, where $\Delta SASA^{Phob}$ and $\Delta SASA^{Phil}$ are the hydrophobic and hydrophilic SASAs buried upon interaction, and σ^{Phob} and σ^{Phil} are the hydrophobic and hydrophilic solvation parameters set equal to 12 and –60 cal(mol Å^2)$^{-1}$, respectively (18). These values are consistent with experimentally fitted ones (30,31), and for the sake of simplicity they have been extended to nucleic acids.

RESULTS

POPS-*A* and POPS-*R* models validation

The total NACS SASAs and the % errors of POPS-*A* and POPS-*R* of the molecules used to fit the empirical p_i and p_{ij} parameters are reported in Table 1. As a comparison, the % errors obtained with the original parameters of Hasel *et al.* (22) (to be compared with the POPS-*A* SASAs) are also reported. With these original parameters the *aape* on the total SASAs relative to the NACS ones is 26%, with SASAs of proteins generally underestimated and the nucleic acids ones generally overestimated. After optimization of the POPS-*A* parameters the *aape* relative to the NACS values is only 7%. All the molecules with relatively higher errors have an even larger error for the Hasel *et al.* results (e.g. 2bnh, 4jdw, 1eur, 1gof, etc.), indicating that these targets may represent a difficult test for an approximate formulation. Our results reproduce nucleic acids areas significantly better than the Hasel *et al.* results (22). The poorer performances of the original parameters is probably a consequence of the size of the molecules used in the Hasel *et al.* fitting, which was not originally designed as ad hoc for proteins and nucleic acids.

For POPS-*R* the *aape* is 6%. The total areas are surprisingly well reproduced even by a coarse-grained method. However, the POPS-*A* and POPS-*R* cross-validation prediction errors e_{CV}^{NACS}, which are equal to 2.6 and 23.4 Å^2, respectively, indicate that for POPS-*R* the particularly good prediction of the total

Table 2A. Solvent accessible surface area buried upon interaction of various components of 30S as calculated from the 70S crystallographic structure

	Isolated[a] SASA (Å^2)	16S SASA (Å^2)	5' SASA (Å^2)	central SASA (Å^2)	3'-major SASA (Å^2)	3'-minor SASA (Å^2)	16S ΔG_{solv}^{loss} (kcal·mol^{-1})
16S	217213	-	-	-	-	-	-
5'	78986	-	-	1376	190	1022	-
central	53818	-	1372	-	299	1662	-
3'-major	70584	-	195	299	-	1056	-
3'-minor	25156	-	1051	1950	942	-	-
S2	12640	599	-	142	456	-	61
S3	11957	1002	68	-	944	-	85
S4	12937	2134	2047	102	-	-	171
S5	7752	1034	408	99	549	26	90
S6	6992	525	-	525	-	-	49
S7	10173	1211	-	107	1104	-	93
S8	8758	1372	9	1362	1	-	109
S9	8336	2317	-	-	2317	-	173
S10	6567	995	-	-	995	-	76
S11	6886	846	-	797	-	56	73
S12	9501	2140	1382	680	-	132	153
S13	10451	1953	-	-	1953	-	138
S14	5542	1457	-	-	1457	-	89
S15	6331	1035	-	1035	-	-	98
S16	5730	1491	1281	239	-	-	134
S17	7719	1834	1119	782	-	-	143
S18	5701	561	-	466	-	103	53
S19	6221	1157	-	-	1157	-	77
S20	6605	1521	1312	-	-	221	133
THX	2060	1021	-	-	1021	0	81

The free energy of solvation loss upon complex formation, ΔG_{solv}^{loss} is also reported (see Materials and Methods). Nomenclature follows that in Yusupov *et al.* (12).
[a]SASA of the isolated component.

SASAs is due to some errors compensation. The almost negligible e_{CV}^{FD} errors, 0.1 and 0.5 Å^2 for POPS-*A* and POPS-*R*, respectively, suggest that both sets of empirical parameters are converged. The small POPS-*A* prediction error (2.6 Å^2) we found is comparable to the 2.2 Å^2 *aae* found by Weiser *et al.* (32) in their analytical method based on tetrahedrally directed neighbor densities (TDND). This method has 21 parameter types derived from a set of 19 compounds of different size (11–4346 atoms). It proved to compute first and second derivatives very fast, but its performance in combination with MD packages is not clear. We had already confidence in the effectiveness of the original POPS formula within MD simulations (18,20,21) and therefore we decided to improve the performance of the same methodology.

To compare the computer performance of POPS-*A* with the TDND approach, we calculated the POPS-*A* SASA of the bovine chymotrypsin complex (PDB entry 1ca0, 4346 atoms) on a SGI R10000/195 MHz processor. The SASA of this molecule was calculated in 0.234 s with the TDND model (32), while only 0.089 s were needed with POPS-*A*. In both cases the reported CPU times do not include pre-processing steps, like the neighboring atoms list calculation, since these extra steps are already part of a MD calculation scheme. We conclude that the POPS-*A* model produces a slightly larger error than the TDND model, but it speeds up the calculation by a factor of ~2.5. In a MD simulation with inclusion of implicit solvent, areas are calculated at every time step. Thus, a key factor in selecting the POPS-*A* model is its required computational time.

In Figure 1 we report the atomic and residue errors distributions from POPS-*A* and POPS-*R*. For POPS-*A*, ~33% of the atomic SASAs are within 1 Å^2 from the NACS values, and that ~90% of the atomic SASAs are within 5 Å^2 from the NACS values. The fraction of errors >10 Å^2 is substantially negligible. For POPS-*R*, the distribution has a maximum around 15%, and the error becomes negligible only at around 40 Å^2. However, it must be considered that the average SASA per residue in the dataset is 55 Å^2 for the amino acids and 177 Å^2 for the nucleotides. The limitation of POPS-*R* is, of course, on the use of a single sphere to model the whole residue, and we refer our errors to the NACS areas obtained at atomic level and summed up to residue level. Nevertheless, we will show that our coarse-grained model is able to detect key interactions with a sensitivity not far from all-atoms models.

To further test the POPS-*R* method and show its predictive ability for a large system like the 70S ribosome, we used the high resolution structure of the 30S ribosome subunit (10) for calculating atomic areas with NACS, and then compared these with residue areas with POPS-*R*. The *aape* relative to the total

Table 2B. Same as Table 2A, but for the 50S ribosome

	Isolated[a] SASA (Å^2)	23S SASA (Å^2)	I SASA (Å^2)	II SASA (Å^2)	III SASA (Å^2)	IV SASA (Å^2)	V SASA (Å^2)	VI SASA (Å^2)	5S SASA (Å^2)	23S ΔG^{loss}_{solv} (kcal·mol^{-1})	5S ΔG^{loss}_{solv} (kcal·mol^{-1})
23S	385688	-	-	-	-	-	-	-	-	-	-
I	83021	-	-	3219	948	390	1338	395	-	-	-
II	110796	-	3491	-	1935	2081	6608	240	1022	-	-
III	54497	-	892	1819	-	1445	39	815	-	-	-
IV	53115	-	437	2045	1731	-	2682	1707	-	-	-
V	90992	-	1429	6376	43	2955	-	1723	229	-	-
VI	43572	-	408	286	846	1642	1576	-	-	-	-
5S	21278	1276	-	1069	-	-	212	-	-	-	-
L1	12662	1454	-	-	-	69	1385	-	-	107	-
L2	10348	2242	-	430	410	1225	336	-	-	162	-
L3	13299	3502	5	110	-	829	1126	1639	-	269	-
L4	11773	3114	1449	1573	-	-	302	-	-	249	-
L5	8716	901	-	-	-	-	901	-	694	79	55
L6	9433	664	-	144	-	-	68	459	-	58	-
L7/L12	12776	666	-	666	-	-	-	-	-	50	-
L9	9868	342	100	-	-	-	265	-	-	32	-
L11	7618	652	-	652	-	-	-	-	-	60	-
L13	7383	1685	529	674	-	-	292	283	-	138	-
L14	6880	879	-	-	-	380	220	341	-	76	-
L15	6322	829	261	423	-	-	186	-	-	59	-
L16	8373	1633	-	991	-	-	725	-	74	134	7
L18	6792	641	-	-	-	-	641	-	935	45	82
L19	4008	354	-	-	-	176	-	197	-	25	-
L22	6603	1254	363	303	307	326	40	5	-	112	-
L23	5487	1047	426	-	657	-	-	-	-	82	-
L24	7753	1412	1412	-	-	-	-	-	-	117	-
L25	6308	50	-	50	-	-	-	-	348	5	30
L29	4524	500	500	-	-	-	-	-	-	45	-
L30	3937	851	-	851	-	-	-	-	98	66	10

[a]SASA of the isolated component.

POPS-*R* SASA of the isolated 16S (subdivided into domains 5′, central, 3′-major and 3′-minor), these four domains and of the 19 resolved proteins is 5%, and this number also holds for the hydrophobic and hydrophilic contributions to the total POPS-*R* SASAs. The NACS and POPS-*R* SASAs buried by the overlap between different parts of 30S are reported in Figure 2, and POPS-*R* clearly reproduces the NACS patterns. Even small interactions, such as those of S2, S5 and S7 with the central domain and of S11 and S12 with the 3′-minor domain, are detected by POPS-*R* which only fails to detect the interactions of S8 with domain 5′, and S5 with the 3′-minor domain. The average absolute error on the relative amount of total SASA buried upon interaction, $aaeb_{\mathrm{Tot}}$ (defined below), is only 4%.

$$aaeb_{\mathrm{Tot}} = \frac{1}{N^{\mathrm{overlaps}}} \sum_{i}^{N^{\mathrm{overlaps}}} 100 \left| \left(\frac{A_i^{\mathrm{Buried}}}{A_i^{\mathrm{Isolated}}} \right)^{\mathrm{POPS}\text{-}R} - \left(\frac{A_i^{\mathrm{Buried}}}{A_i^{\mathrm{Isolated}}} \right)^{\mathrm{NACS}} \right|$$

The absolute error on the relative amount of POPS-*R* hydrophobic (and hydrophilic) SASA buried upon interaction and relative to the NACS values, $aaeb_{\mathrm{Phob}}$ (defined below), is 9%.

$$aaeb_{\mathrm{Phob}} = \frac{1}{N^{\mathrm{overlaps}}} \sum_{i}^{N^{\mathrm{overlaps}}} 100 \left| \left(\frac{A_i^{\mathrm{Phob,Bur}}}{A_i^{\mathrm{Tot,Bur}}} \right)^{\mathrm{POPS}\text{-}R} - \left(\frac{A_i^{\mathrm{Phob,Bur}}}{A_i^{\mathrm{Tot,Bur}}} \right)^{\mathrm{NACS}} \right|$$

where $A^{\mathrm{Phob,Bur}}$ and $A^{\mathrm{Tot,Bur}}$ are the hydrophobic and total SASAs buried upon interaction. We can conclude that POPS-*R* is able to give almost quantitative estimation of: (i) isolated SASAs; (ii) burial at surface due to interaction between molecules/fragments; and (iii) partitioning into hydrophobic and hydrophilic contributions.

The POPS-*A* and POPS-*R* versions and their optimized parameters will be available at the web site http://mathbio.nimr.mrc.ac.uk/~ffranca/POPS. The CPU time required to calculate the SASAs of 30S, 50S and 70S on a Pentium-III 800 MHz processor is equal to 30, 63 and 188 s, respectively. These CPU times include the preprocessing steps.

Key interactions in the ribosome

The SASAs buried upon interaction of the different components (proteins and RNA domains) within 30S and 50S from the 70S structure (12) are reported in Table 2A. In addition, the free energy of solvation loss upon complex formation ($\Delta G^{\mathrm{loss}}_{\mathrm{solv}}$) is reported for each protein.

Table 3. SASA buried upon interaction of the proteins of the 30S and 50S subunits as calculated from the 70S crystallographic structure

Mol.1	Mol.2	ΔSASA1 (Å^2)	ΔSASA2 (Å^2)
S2	S8	180	290
S3	S10	242	211
S3	S14	494	549
S3	S5	5	5
S4	S5	130	119
S5	S8	271	284
S6	S15	8	13
S6	S18	852	543
S7	S9	116	131
S7	S11	182	143
S8	S12	79	69
S8	S17	156	151
S9	S10	91	87
S9	S13	104	97
S9	S14	75	151
S10	S14	614	788
S11	S18	205	423
S12	S17	46	60
S13	S19	139	154
S13	THX	5	5
S15	S17	38	39
L3	L13	2	1
L3	L14	168	170
L5	L18	63	48
L7/L12	L11	420	450
L14	L19	385	570
L16	L25	109	204
L23	L29	204	148

Proteins in the 30S subunit interact with 809 nt of 16S, indicating that almost 50% of the nucleotides are in contact with an amino acid, while for the 50S there are 956 of these contacts (only 32% of nucleotides interact with an amino acid). The crystallographic study of the 50S subunit from *Haloarcula marismortui* reported 1157 protein/RNA van der Waals contacts (8).

Figure 2 and Table 2A and 2B indicate that only the 3′-minor domain of 16S is not involved in any significant interaction with proteins. Only three proteins show significant surface burial with more than one domain. These are S5 with the 5′- and 3′-major domains, and S12 and S17 with the 5′ and central domains. The 30S subunit also presents the largest number of protein–protein interactions since ~8300 Å^2 of SASA are buried by overlaps between S-proteins, while only 2900 Å^2 of SASA are buried by L-protein overlaps in 50S. Solvent accessible buried areas upon interaction between 30S proteins, as calculated from the 70S structure, are reported in Table 3. It is interesting to note that the largest buried areas are observed for proteins S6 and S18, which have been demonstrated to form a heterodimer with a key role in the cooperative binding to the S15–rRNA complex during the ordered assembly of the ribosome. Large buried areas are also observed for proteins S3–S14 and S10–S14; these are all tertiary binding proteins (33) and their mechanism of binding could be similar to the one proposed for S6 and S18.

In the case of the large subunit, POPS-*R* is able to detect all the main interactions described for the high resolution structure of the *H.marismortui* 50S subunit (8). For this reason, POPS-*R* can be used as a helpful tool in the refinement of low resolution structures, since it is able to identify key interactions starting from less-resolved structures.

In Figure 3A and B we report the relative contribution of each amino acid of the S- and L-proteins to: (i) the total SASA of the isolated amino acids of the S- and L-proteins (calculated with NACS as Ala-Xaa-Ala sequences); (ii) the total POPS-*R* SASA of the isolated S- and L-proteins as folded in the 70S structure; (iii) the total POPS-*R* SASA buried upon interaction of the S- and L-proteins with the RNAs of their own subunit; and (iv) the total POPS-*R* SASA buried upon interaction of S- and L-proteins with the RNAs of the other subunit. Passing from the isolated amino acid to the folded one in both S- and L-proteins, hydrophilic residues like Arg and Lys increase their solvent accessibility, while hydrophobic ones such as Ala, Leu, Ile and Val become considerably buried.

Arginine-rich motifs are known to have high affinity and specificity for RNA (34,35). Indeed, we found that Arg and Lys residues play a key role in these interactions. In fact, for both the S- and L-proteins, Arg residues contribute mostly to interactions with the RNA of the other subunit, while Lys residues contribute mostly to interactions with the RNA of the same subunit.

The poly-functionality of the guanidinium group makes this residue an optimal moiety for protein–RNA interactions, and the long arm of the Arg sidechain can effectively direct the guanidinium group towards acceptors sites of the other subunit. Glycines in the plots show a trend similar to Arg residues, and this could be ascribed to a necessity of flexibility for optimizing interactions with the RNA of the other subunit.

As for the analysis of RNA–RNA interactions, the plots of Figure 2 show that interactions between RNA–RNA domains bury, generally, more hydrophilic surface (61 and 68% according to NACS and POPS-*R*, respectively) than protein–RNA interactions. Domain 3′-minor interacts substantially with all the other domains, but a greater fraction of its buried area is in common with the central domain. The only other strong domain–domain interaction occurs between the 5′ and the central domains. The most frequently occurring nucleotides in ribosomal RNAs are adenines, followed by cytosines, guanines and uridines. Adenine nucleotides are also mostly involved in domain–domain interactions since 25% of the SASA buried in the inter-domains interactions of both 16S and 23S belongs to A (to be compared with only 20% of adenines in the isolated SASAs of 16S and 23S). It has been observed that adenine is a key residue for helix packing in RNA (36), showing a surface complementarity with the minor groove that optimizes a combination of van der Waals, electrostatic and hydrogen bonding contacts. In addition, adenines are abundant in the recurrent loop E motif present within the three-way junction loop that binds protein S7 to 16S and is involved in RNA–RNA interactions (37).

The overall SASA buried at the interface of the 30S and 50S subunits amounts to 8500 Å^2, and it is shown in Figure 4. The POPS-*R* surface at the interface of the 30S and 50S subunits forms a triangular patch (Fig. 4, fitting in the white triangle)

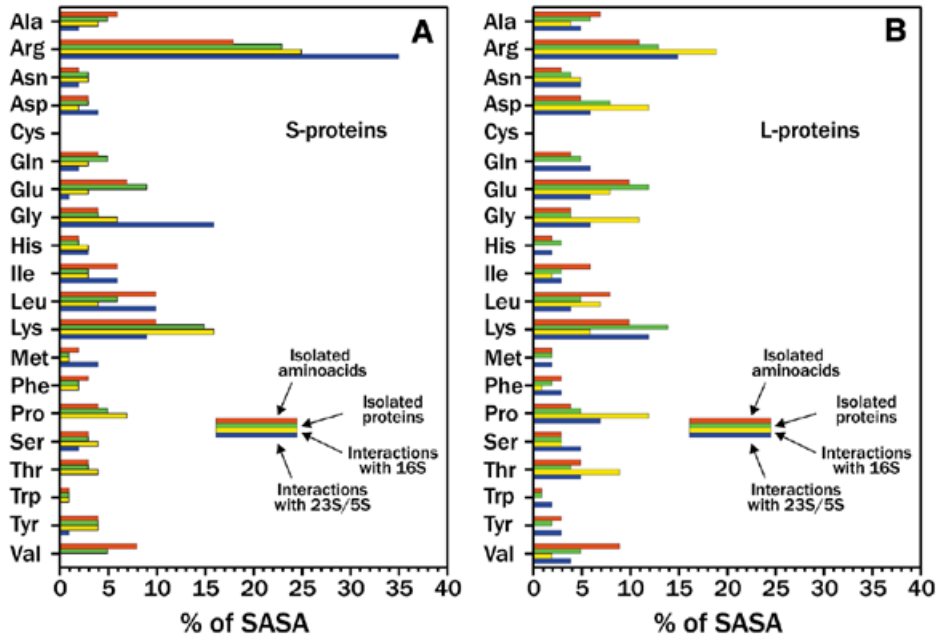

Figure 3. Relative contribution of each amino acid of (**A**) S- and (**B**) L-proteins to: red bars, total SASA of the isolated amino acids of S- and L-proteins; green bars, total SASA of isolated S- and L-proteins; yellow bars, total SASA buried upon interaction of S- and L-proteins with 16S; blue bars, total SASA buried upon interaction of S- and L-proteins with 23S/5S.

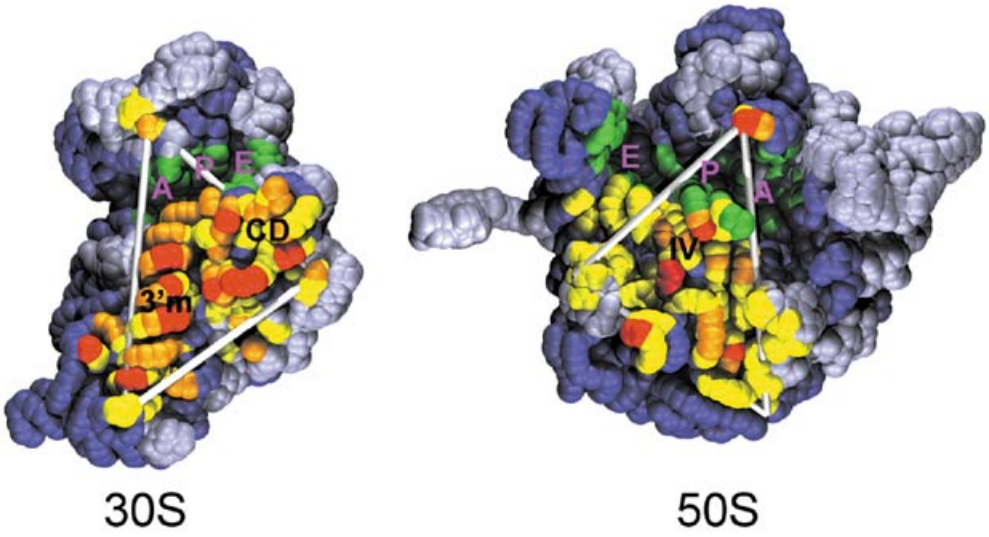

Figure 4. Interface view of the 30S and 50S subunits from the 70S crystallographic structure (12). Yellow, orange and red spheres correspond to residues which, upon 30S and 50S association, bury <37.5 Å^2, between 37.5 and 75 Å^2, and >75 Å^2 of SASA, respectively. Green spheres correspond to residues that interact with the A-, E- and P-tRNAs.

(12) where the most exposed residues (in red) belong mainly to the 3′-minor domain. This allows us to speculate that the 3′-minor domain role is mainly to interact with the other subunit. Together with observed electrostatic complementarity of the interface residues (38), exposed surface complementarity of RNA domains is striking. The 70S residues interacting with the tRNAs are marked in green (Fig. 4), and form a deep groove which, in the case of the 30S, separates the head from the base. Almost all the 30S–50S interactions occur in the lower part of the triangle, below the tRNA binding sites. The presence of only one 30S–50S anchoring point above the tRNAs binding groove could allow for relatively easy movements of these regions.

The nucleotide–nucleotide contacts at the interface involve mostly G–A bases (48 contacts), followed by G–G bases (40 contacts), and U–G and C–A bases (28 contacts in both cases). It is well known that non-canonical Watson–Crick base pairs participate in a large number of edge-to-edge interactions with one or more bases (39).

Ribosome–tRNA interactions

When interactions of 30S and 50S with tRNAs were considered, we used POPS-*R* for the two ribosome subunits, while we used the POPS-*A* approach for the tRNAs, since the tRNAs were resolved at atomic level.

Table 4. SASA buried upon interaction of the tRNAs with the various components of the 30S and 50S subunits as calculated from the 70S crystallographic structure

Mol.1	Mol.2	ΔSASA1 (Å^2)	ΔSASA2 (Å^2)
A-tRNA	16S	197	458
A-tRNA	5'	57	102
A-tRNA	3'-major	107	223
A-tRNA	3'-minor	37	138
A-tRNA	23S	729	1727
A-tRNA	II	178	462
A-tRNA	IV	207	548
A-tRNA	V	347	727
A-tRNA	S13	13	16
A-tRNA	L16	13	12
A-tRNA	PtRNA	21	34
A-tRNA	COD	88	136
P-tRNA	16S	325	823
P-tRNA	central	59	137
P-tRNA	3'-major	242	605
P-tRNA	3'-minor	28	89
P-tRNA	23S	964	1688
P-tRNA	IV	320	753
P-tRNA	V	646	948
P-tRNA	S9	88	241
P-tRNA	S13	195	416
P-tRNA	L5	46	88
P-tRNA	L16	136	154
P-tRNA	COD	86	135
E-tRNA	16S	182	454
E-tRNA	central	90	218
E-tRNA	3'-major	92	235
E-tRNA	23S	622	1193
E-tRNA	I	26	46
E-tRNA	IV	234	575
E-tRNA	V	377	641
E-tRNA	S7	461	594
E-tRNA	S11	11	18
E-tRNA	L1	153	269
E-tRNA	P-tRNA	72	100

POPS-*R* is able to clearly identify all previously reported key interactions between 70S and tRNAs (12). In addition, we observe a number of further interactions and we are able to identify some, as yet undetected, conformational changes with respect to the isolated structure of 30S. In fact, clear differences in the binding modes of A-, P- and E-tRNA to the ribosome are detected. The SASAs of interaction between 70S and the A-, P- and E-tRNAs are close to 3000, 5000 and 4000 Å^2, respectively. Most strikingly, proteins represent only 1% of the surface of interaction of the A-site. This fraction increases to 26% in the P-site, and reaches its maximum at 35% in the E-site. This gradient in the amount of proteic surface at the tRNAs binding sites could be needed in order to allow for the translocation mechanism to occur. In Table 4 the buried areas upon interaction of tRNAs and 30S and 50S components with tRNAs are reported. As already pointed out, both S- and L-proteins interact with tRNAs (12), and proteins S7, S9 and S13 are found to interact with the anticodon region of the tRNAs. The interaction between S7 and E-tRNA involves quite a high number of proteic residues (from Val75 to Gln86, and from Asp140 to Tyr151) which interact only with one stretch of E-tRNA (from G30 to G42). In addition to this, we observe the participation of proteins L1 and L16 to the interaction with E- and P-tRNAs. One of our most interesting findings concerns protein S9, which protrudes its terminal residue (Arg128) to interact with nucleotides OMC32-U33-OMG34-A35 in the anticodon region of P-tRNA (OMC, 2′-*O*-methylcytidine; OMG, 2′-*O*-methylguanosine). This arginine is universally conserved and it is found to form a hydrogen bond with a phosphate of A35 (12). The differences between the S9 terminal arm in the 70S structure compared with the 30S high resolution structure (10) are quite remarkable. By superimposing S9/30S on S9/70S we observe a decrease of ~200 Å^2 of the area buried in the interaction with P-t RNA (Fig. 5). This is mainly due to a movement of 3.3 Å of the terminal Arg128 of S9 in order to reach the negatively charged surface of P-tRNA. In the 30S subunit this arginine has its guanidinium group folded back in a cation-π interaction with Tyr125 of the same chain (Fig. 5). These intramolecular interactions have been found to be strongly recurrent in the PDB and they are recognized as important non-covalent binding interactions (40).

The interaction between S13 and P-tRNA involves the C-terminal region of S13 from Thr116 to Lys126. This region contains the Lys120-Lys121-Lys122 and Arg125-Lys126 motifs. A comparison similar to that carried out for S9 gives a decrease in the buried SASA due to the interaction of S13/30S with P-tRNA of 250 Å^2. The reduction is due to a movement of the C-terminal tail residues of S13 which swing away from the anticodon region of P-tRNA. The conformational changes that POPS-*R* reveal might be a sign of movements occuring during translation, similar to those of the lever in a watch.

In the 50S, contacts only occur between L1 and A-tRNA and L16 and P-tRNA. The Arg52-Arg53-Ser54 motif is mostly involved in the interaction of L1 with the elbow of A-tRNA around 5MU54 (5MU, 5-methyluridine), while Ser125 interacts (probably through a H-bond) with atom O4 of C56. The interaction of L16 with P-tRNA again involves a H-bond donor residue, Ser22, which overlaps with the ribose of A64, while Asp97 and Asp99 interact with the phosphates of G1 and C2 of P-tRNA.

Energetics of observed interactions

Only limited experimental thermodynamic data are available for the interaction of S- and L-proteins with ribosomal RNAs, and the ΔG values of these interactions are always close to –10 kcal mol^{-1} (33,41,42). Association requires a negative ΔG for the interaction and, therefore, an unfavorable energetic term due to the loss of solvation (ΔG_{solv}^{loss}) must be overcome. POPS-*R* can provide an estimate of the upper bound magnitude for this term, evaluated by assuming rigid body association. Conformational changes in the unassociated species can reduce this value (43). Considering that the crystal structure of the isolated 30S and 50S are quite similar to 70S (12) we have good reasons to consider this approximation valid in this context.

The magnitude of the ΔG_{solv}^{loss} values reported in Table 2A suggests that a significant amount of free energy of solvation is lost for any of the interactions considered, and indicate that

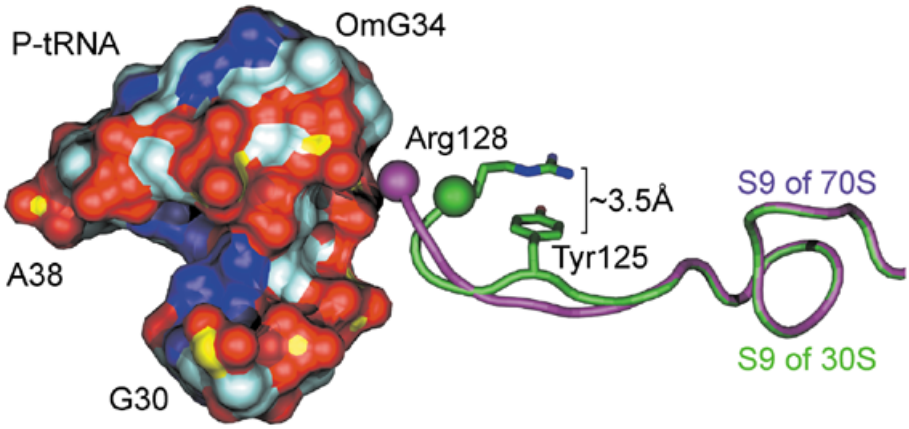

Figure 5. Detailed view of the S9 P-tRNA interaction. Only the surface of the nucleotide stretch between G30 and A38 is reported for P-tRNA. Red, dark blue, light blue and yellow surfaces correspond to O, N, C and P atoms, respectively. The purple ribbon corresponds to the tail of S9 of 70S, while the green ribbon corresponds to the tail of 30S's S9 after superposition with 70S's S9. The sidechains of Arg128 and of Tyr125 of S9 of 30S are also reported. The purple and green spheres are centered at the C^{α} of Arg128.

strong stabilizing interactions are required to obtain association. This conclusion offers an explanation for the high fraction of Arg and Lys residues at the protein–RNA interface of the ribosome already found in other protein–nucleic acids complexes. These positively charged residues promote association through strong interactions with the negatively charged phosphates (44). Hydrogen bonds between proteins and RNAs represent another key stabilizing factor for the complexes. Previous statistical analysis of protein–nucleic acid complexes indicated the occurrence of approximately one intermolecular H-bond per 125 Å^2 of interface area (35,44).

The value of ΔG_{solv}^{loss} term related to the interaction of the whole 30S and 50S to form 70S is quite high (298 kcal mol^{-1}) and the complessive experimental ΔG of interaction in the 70S of *Escherichia coli* is about –12 kcal mol^{-1} (45). The high ΔG_{solv}^{loss} we estimated is consistent with the high ionic strength required to prevent ribosome denaturation, since one of the major roles of cations present in solutions [often Mg^{2+}(45)] is to reduce repulsive interactions between phosphates.

CONCLUSIONS

We have presented and validated two novel approaches for the analytical calculation of the SASA of proteins and nucleic acids at atomic and residue levels, named POPS-*A* and POPS-*R*, respectively. The analytical formulation on which POPS-*A* is based is simple, easily derivable and fast to compute. It has already been proven to be well suited for practical use in MD simulations as an approximation to the first solvation shell. The two models have been trained to approximate the atomic and residue accurate NACS SASAs of a database of 89 biological molecules. The cross-validation resampling procedure we followed indicated that POPS-*A* predicts atomic SASAs with an *aae* of 2.6 Å^2. The POPS-*A* approach has been implemented in the GROMOS96 (46) package as part of the implicit solvation contribution (to be published).

The residue based approach POPS-*R* was validated through a comparison with accurate all-atoms approaches. We used the high resolution structure of the 30S ribosome subunit (10) for calculating atomic areas with NACS, and then compared these with residue areas with POPS-*R*. The *aape* is 5% for the 30S components. Our coarse-grained model is therefore able to detect key interactions with a sensitivity not far from all-atoms models.

POPS-*R* was used to examine, in detail, the structures of the 70S, 30S and 50S ribosomes. Most of the interaction within the subunits and at their interfaces were clearly identified. Some interesting differences between 30S alone and within the 70S were highlighted. Owing to the presence of the P-tRNA in the 70S ribosome, conformational rearrangements occur within the subunits, exposing Arg and Lys residues to negatively charged binding sites of P-tRNA.

In our opinion, POPS-*R* can be a valuable tool for the structural biology community in filtering key interactions of large macromolecular assemblages and in complementing their refinement process. On the other hand, POPS-*A* can be used for more detailed calculations and in combination with accurate computer simulations methods.

ACKNOWLEDGEMENTS

We are grateful to J. Kleinjung, A. Lane, S. Martin and A. Ramos for discussions and critical reading of the manuscript. We would also like to thank R. Ali for assembling the protein database. L.C. thanks the CIMCF of Universitá Federico II, Naples, for technical support.

REFERENCES

1. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
2. Brenner,S. and Levitt,M. (2000) Expectations from structural genomics. *Protein Sci.*, **9**, 197–200.
3. Doudna,J. (2000) Structural genomics of DNA. *Nature Struct. Biol.*, Suppl **7**, 954–956.
4. Marcotte,E., Pellegrini,M., Thompson,M., Yeates,T. and Eisenberg,D. (1999) A combined algorithm for genome-wide prediction of protein function. *Nature*, **402**, 83–86.
5. Thornton,J., Todd,A., Milburn,D., Borkakoti,N. and Orengo,C. (2000) From structure to function: approaches and limitations. *Nature Struct. Biol.*, Suppl **7**, 991–994.
6. Darst,D. (2001) Bacterial RNA polymerase. *Curr. Opin. Struct. Biol.*, **11**, 155–162.

7. Luger,K., Mader,A., Richmond,R., Sargent,D. and Richmond,T. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
8. Ban,N., Nissen,P., Hansen,J., Moore,P. and Steitz,T. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
9. Nissen,P., Hansen,J., Ban,N., Moore,P. and Steitz,T. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–930.
10. Wimberly,B., Brodersen,D., Clemons,W., Carter,A., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
11. Schluenzen,F., Tocilj,A., Zarivach,R., Harms,J., Gluehmann,M., Janell,D., Bashan,A., Bartels,H., Agmon,I., Franceschi,F. and Yonath,A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*, **102**, 615–623.
12. Yusupov,M., Yusupova,G., Baucom,A., Lieberman,K., Earnest,T., Cate,J. and Noller,H. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.
13. Puglisi,J., Blanchard,S. and Green,R. (2000) Approaching translation at atomic resolution. *Nature Struct. Biol.*, **7**, 855–861.
14. Kauzmann,W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, **14**, 1–64.
15. Jones,S. and Thornton,J. (1997) Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.*, **272**, 121–132.
16. Robertson,A.D. and Murphy,K.P. (1997) Protein structure and the energetics of protein stability. *Chem. Rev.*, **97**, 1251–1267.
17. Bowie,J., Lüthy,R. and Eisenberg,D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, **253**, 164–170.
18. Fraternali,F. and van Gunsteren,W. (1996) An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. *J. Mol. Biol.*, **256**, 939–948.
19. Kleinjung,J., Bayley,P. and Fraternali,F. (2000) Leap-dynamics: efficient sampling of conformational space of proteins and peptides in solution. *FEBS Lett.*, **470**, 257–262.
20. Ferrara,P. and Caflish,A. (2000) Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl Acad. Sci. USA*, **97**, 10780–10785.
21. Gsponer,J. and Caflish,A. (2001) Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.*, **309**, 285–298.
22. Hasel,W., Hendrikson,T. and Still,W. (1988) A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.*, **1**, 103–116.
23. Hubbard,S., Campbell,S. and Thornton,J. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.*, **220**, 507–530.
24. Still,W., Tempczyk,A., Hawley,R. and Hendrickson,T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, **112**, 6127–6129.
25. Wodak,S. and Janin,J. (1980) Analytical approximation to the accessible-surface area of proteins. *Proc. Natl Acad. Sci. USA*, **77**, 1736–1740.
26. Friedl,H. and Stampfer,E. (2002) Cross-validation. In El-Shaarawi,A. and Piegorsch,W.E. (eds), *Encyclopedia of Environmetrics*. Wiley, Chichester, UK, pp. 452–460.
27. Friedl,H. and Stampfer,E. (2002) Jackknife resampling. In El-Shaarawi,A. and Piegorsch,W.E. (eds), *Encyclopedia of Environmetrics*. Wiley, Chichester, UK, pp. 1089–1098.
28. Friedl,H. and Stampfer,E. (2002) Resampling methods. In El-Shaarawi,A. and Piegorsch,W.E. (eds), *Encyclopedia of Environmetrics*. Wiley, Chichester, UK, pp. 1768–1770.
29. Efron,B. and Gong,G. (1983) A leisurely look at the bootstrap, the jackknife and the cross-validation. *The American Statistician*, **37**, 36–48.
30. Wesson,L. and Eisenberg,D. (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.*, **1**, 227–235.
31. Nolde,E., Arseniev,A. and Efremov,R. (1997) Atomic solvation parameters for protein in a membrane environment. Application to transmembrane α-helices. *J. Biomol. Struct. Dyn.*, **15**, 1–18.
32. Weiser,J., Shenkin,P. and Still,W. (1999) Approximate solvent-accessible surface areas from tetraedrally directed neighbor densities. *Biopolymers*, **50**, 373–380.
33. Recht,M. and Williamson,J. (2001) Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J. Mol. Biol.*, **313**, 35–48.
34. Cheng,A., Calabro,V. and Frankel,A. (2001) Design of RNA-binding proteins and ligands. *Curr. Opin. Struct. Biol.*, **11**, 478–484.
35. Jones,S., Daley,D., Luscombe,N., Berman,H. and Thornton,J. (2001) Protein–RNA interactions: a structural analysis. *Nucleic Acids Res.*, **29**, 943–954.
36. Doherty,E., Batey,R., Masquida,B. and Doudna,J. (2001) A universal mode of helix packing in RNA. *Nature Struct. Biol.*, **8**, 339–343.
37. Leontis,N. and Westhof,E. (1998) A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.*, **283**, 571–583.
38. Baker,N., Sept,D., Joseph,S., Holst,M. and McCammon,J. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
39. Leontis,N. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
40. Gallivan,J. and Dougherty,D. (1999) Cation-π interactions in structural biology. *Proc. Natl Acad. Sci. USA*, **96**, 9459–9464.
41. Mougel,M., Ehresmann,B. and Ehresmann,C. (1986) Binding of *Escherichia coli* ribosomal protein S8 to 16S rRNA: kinetic and thermodynamic characterization. *Biochemistry*, **25**, 2756–2765.
42. Draper,D. (1990) Structure and function of ribosomal protein-RNA complexes: thermodynamic studies. In Hill,W., Dahlberg,A., Garrett,R., Moore,P., Schlessinger,D. and Editors,J.R. (eds), *The Ribosome Structure, Function and Evolution*. American Society of Microbiology, pp. 160–167.
43. Spolar,R. and Record,M. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 769–770.
44. Nadassy,K., Wodak,S. and Janin,J. (1999) Structural features of protein-nucleic acid recognition sites. *Biochemistry*, **38**, 1999–2017.
45. Wishnia,A. and Boussert,A. (1977) The non-specific role of Mg^{2+} in ribosomal subunit association: kinetics and equilibrium in the presence of other divalent metal ions. *J. Mol. Biol.*, **116**, 577–591.
46. van Gunsteren,W., Billeter,S., Eising,A., Hünenberger,P., Krüger,P., Mark,A., Scott,W. and Tironi,I. (1996) *Biomolecular Simulations: the GROMOS96 Manual and User Guide BIOMOS b.v.*, 1st Edn. Laboratory of Physical Chemistry, ETH Zentrum, Groningen, German.
47. Orengo,C., Michie,A., Jones,S., Jones,D., Swindells,M. and Thornton,J. (1997) Cath—a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.