SCREENED CHARGE ELECTROSTATIC MODEL IN PROTEIN-PROTEIN DOCKING SIMULATIONS

JUAN FERNANDEZ-RECIO[1], MAXIM TOTROV[2], RUBEN ABAGYAN[1]

[1]*Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037,USA, E-mail: jfrecio@scripps.edu, abagyan@scripps.edu*

[2]*Molsoft, LLC, 3366 Torrey Pines Court, La Jolla, CA 92037, USA, E-mail: totrov@molsoft.com*

A new method for considering solvation when calculating electrostatics for protein docking is proposed. The solvent-exposed charges are attenuated by induced solvent polarization charges. Modified charges are pre-calculated and the correction doesn't affect the speed of the actual simulation. The new Screened Charge Electrostatic Model (SChEM) results in an improved discrimination of near-native solutions from false positives in docking simulations as compared to conventional 'non-solvated' charge assignment. A series of protein-protein complexes were analyzed by running automated rigid-body Monte-Carlo docking simulations using the 3-D coordinates of the unbound components. In all but one case, the use of solvation screened charges for electrostatic calculations helped to improve the rank of the near-native solution after rigid-body simulations. The SChEM also drastically improved the results of the subsequent refinement of the interface side-chains. In all cases the final lowest energy solution was found within 3.0 Å r.m.s.d. of the crystal structure.

1 Introduction

In silico prediction of protein-protein interactions will undoubtedly play an essential role in the structural genomics era. Currently ongoing structural genomics projects will drastically increase the number of available 3-D protein structures[1], and a variety of computational tools will be needed to efficiently use this structural information, with the ultimate goal of understanding the complex network of protein-protein interactions in a living organism. In this context, a number of docking methods have been developed to predict the structure of a protein-protein complex given the 3-D coordinates of its individual components.[2] Although early rigid-body docking methods based on purely geometrical criteria were adequate when the approaching subunits artificially presented the same conformation as in the complex,[3] [4] the prediction results were clearly poorer when using the 3-D coordinates of the uncomplexed subunits.[5] It was soon evident that geometry-based approaches were not accurate enough to model the induced fit of the interacting surfaces upon binding. Treatment of interface flexibility as well as more realistic energy approximations had to be developed.

The inclusion of energy determinants, together with molecular flexibility, can result in more realistic simulations. A global minimization procedure with a

complete energy description has been reported and successfully applied in the prediction of a lysozyme-antibody complex[6] and in a blind prediction contest.[7] The method, based on ICM methodology,[8] used a pseudo-brownian Monte-Carlo minimization[9] and a subsequent biased probability Monte Carlo[10] optimization of the interface side-chains. The use of a soft interaction energy function pre-calculated on a grid[11] can drastically increase the speed of the docking simulations, as has been observed for protein-ligand docking.[12]

Electrostatic interactions play an important role in protein-protein docking. The accurate definition of electrostatics, including solvation considerations, is critical for the correct ranking of the near-native solutions. Here we present a method for calculating the solvation-corrected grid electrostatic energy and show that it substantially improves the results of rigid-body docking simulations and side-chain refinement.

2 Methods

2.1 Energy description

The interaction energy potentials were pre-calculated on a grid[11] within a 3-D box covering approximately half of the total receptor surface (including the known or hypothetical receptor binding site). The energy estimate used during docking simulations consisted of the following terms:[13]

$$E = E_{\mathrm{Hvw}} + E_{\mathrm{Cvw}} + E_{\mathrm{el}} + E_{\mathrm{hb}} + E_{\mathrm{hp}} \quad (1)$$

where E_{Hvw} is the van der Waals potential for a hydrogen atom probe, E_{Cvw} the van der Waals potential for a heavy atom probe (a generic carbon of 1.7 Å radius was used), E_{el} an electrostatic potential generated from the receptor with a distance dependant dielectric constant, E_{hb} the hydrogen-bonding potential calculated as spherical Gaussians centered at the ideal putative donor and/or acceptor sites, and E_{hp} a hydrophobicity potential roughly proportional to the buried hydrophobic surface area. Van der Waals potentials were initially truncated to a maximum energy value of 1.0 kcal mol^{-1} to avoid inter-molecule repulsive clashes arising from the rigidity of their side-chains.

2.2 Solvation correction for electrostatics

Previously we used a distance-dependent dielectric model _=4*r [14] to approximate the effects of solvent screening of electrostatic interaction. While this approach is fast and simple, it does not reflect the dependence of the solvation effect on the degree of the solvent exposure.

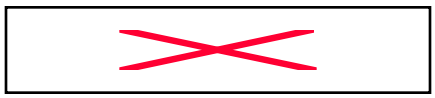
Charge scaling, i.e. the reduction of solvent-exposed charges, has previously been used to account for solvation effects in certain systems, such as DNA.

Polarization of the solvent near the charged solute atom results in the formation of an induced solvent charge of the opposite sign, attenuating the electrostatic interactions of the solute. This effect can be considered as an effective reduction of the solute charges. The optimal degree of charge reduction has to be determined from such factors as solvent accessibility or fit to experimental values. Here we propose a method for charge scaling based on continuum dielectric solvation model.

Continuum dielectric solvation model represents the solvent as a medium of high dielectric constant (ε_{out} =78.5 for water), while the interior of the solute has relatively low dielectric constant (we use ε_{in} =4 in this work). Poisson equation has to be solved to evaluate accurately the electric field in such a system:

$$\nabla(\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r}))=\rho(\mathbf{r}) \qquad (2)$$

where ε is the dielectric constant (permittivity), ϕ is the electric potential and ρ is the charge density. The boundary element (BE)[15] method is a popular approach to solving the Poisson equation in continuum dielectric electrostatics calculations. The BE method is based on the representation of electrostatic solvation effects by appropriate induced surface charge density $\boldsymbol{\sigma}_s$ on the solute/solvent boundary. The full electrostatic field at a point **r** is represented as



(3)

where the first term represents the standard Coulomb field of the atomic charges q_i and the second term accounts for the field of the induced surface charge. This representation provides a basis for quantitative evaluation of optimal scaled charges: the induced surface charge density can be projected onto the nearby atoms. The electric field of the resulting corrected atomic charges should approximate closely the exact solution, i.e. the combined field of the original atomic charges and of the induced surface charge distribution.

To obtain numeric solution in the BE method, the boundary is typically split into patches, or boundary elements, and surface charge densities σ_i are evaluated by solving a system of linear equations[15]. Serendipitously, the REBEL (rapid boundary element electrostatics)[16] implementation of the method uses per-atomic BE's, i.e. patches of the molecular surface assigned to the atoms of the solute and generates the surface charge values $\sigma_i S_i$ for these patches (S_i is the area of the *i*-th BE). That allowed us to use a very simple approach to scale the surface charges by adding to each partial atomic charge q_i the corresponding induced solute charge:

$$q_i' = q_i + \sigma_i S_i \qquad (4)$$

While the approach obviously simplifies the complex nature of the electrostatic solvation, it largely reproduces the expected effect of solvent on electrostatic

interactions in docking: partially buried (but involved in the interaction) charges contribute strongly to the binding, while exposed charges interact relatively weakly.

1.3 Docking simulations

The two-step docking procedure used in this work consisted of a rigid-body docking step followed by side-chain refinement (scheme in Figure 1). The resulting conformations from the first rigid body step were further optimized by an ICM[17] global optimization algorithm, with flexible interface ligand side-chains and a grid map representation of the receptor. In this refinement step, the internal energy for the ligand interface side-chains was also considered, including the van der Waals, hydrogen bonding and torsion energy calculated with ECEPP/3 parameters[18], and the Coulomb electrostatic energy (distance-dependent dielectric constant; $\varepsilon=4*r$).

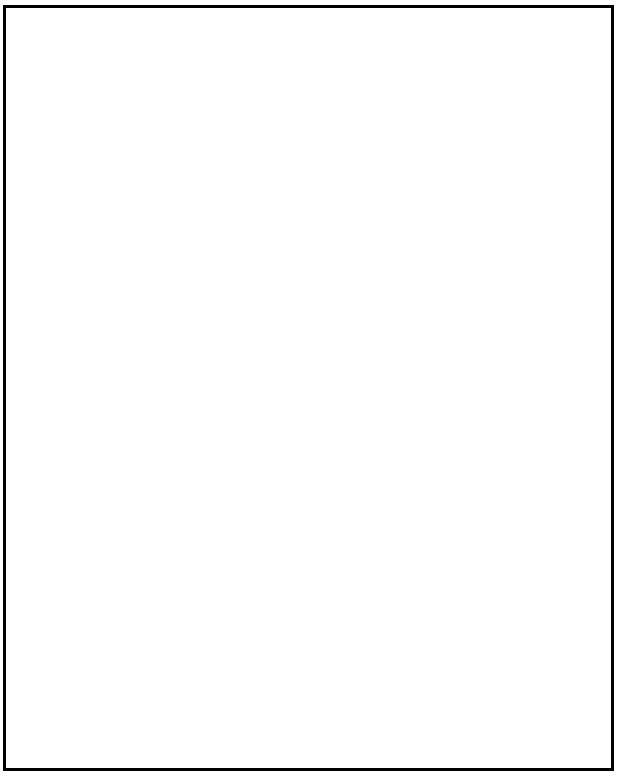


Figure 1. Scheme of the docking protocol used in this work. The ligand molecule was positioned in a random orientation inside the grid potential box and was systematically rotated to generate 120 different starting conformations. The six positional variables of the ligand were sampled by a pseudo-Brownian Monte Carlo optimization, in which each random step was followed by local minimization. New conformations were selected according to the Metropolis criterion[19] with a temperature of 5000K. Each simulation was terminated after 20,000 energy evaluations. Low energy conformations with pairwise r.m.s.d. for the ligand interface C^{α} atoms greater than 4 Å were retained in a conformational stack, and their interface side-chains were further optimized using a Biased Probability Monte Carlo procedure. The simulation temperature for this refinement step was set to 300K, and the total number of energy evaluations was 1,000 times the number of flexible interface torsion angles. The surface-based solvation energy[20] was included in the final energy to select the best refined solutions.

2 Results

2.1 Rigid-body docking simulations

Docking simulations have been applied to the selected protein-protein complexes listed in Table 1. For all of them, the 3-D structures of their unbound subunits are available.

Table 1. PDB codes of the protein-protein complexes and unbound subunits used in this work.

COMPLEX	RECEPTOR		LIGAND	
PDB	Name	PDB	Name	PDB
1ca0	Chymotrypsin	5cha	APPI	1aap
1cbw	Chymotrypsin	5cha	BPTI	1bpi
2sni	Subtilisin	2st1	CI-2	2ci2
1taw	Trypsin (bovine)	5ptp	APPI	1aap
3tgi	Trypsin (rat)	1ane	BPTI	1bpi
1brc	Trypsin mutant (rat)	1bra	APPI	1aap

For all test cases, automated rigid-body docking simulations starting from the unbound subunits were performed, using initially the uncorrected electrostatic energy. Amongst the low energy solutions stored for each complex, we always found at least one conformation within 4 Å r.m.s.d. (calculated for the ligand interface C^{α} atoms when only the receptor C^{α} atoms were superimposed onto the crystallographic structure) from the experimental structure. Rank (according to total energy) and r.m.s.d. values for the near native solutions found in all test cases are shown in Table 2 (column *corrected charges*).

Table 2. Rigid-Body docking results with uncorrected-and corrected charge electrostatic.

Complex	uncorrected charges		corrected charges	
	Rank	r.m.s.d. (Å)	Rank	r.m.s.d. (Å)
1ca0	16 of 233	1.4	6 of 228	1.4
1cbw	10 of 220	1.5	5 of 231	0.9
2sni	43 of 243	2.4	66 of 220	2.5
1taw	3 of 232	3.2	1 of 228	3.5
3tgi	210 of 238	0.3	23 of 243	0.6
1brc	20 of 218	3.5	1 of 237	3.7

In none of the complexes, the near native conformation was found as the lowest energy solution. The best scored near native solution (ranked 3rd) was found for

trypsin/APPI complex (PDB 1taw), with an r.m.s.d. of 3.2 Å from the real structure. Interestingly, for the trypsin/BPTI complex (PDB 3tgi), we found a solution very close to the real structure (0.3 Å r.m.s.d.), but unfortunately very poorly scored (ranked 210th).

To evaluate if our SChEM approach could help to remove false positives (e.g. conformations with large interaction surfaces and over-estimated interactions of solvated charges), we performed rigid-body docking simulations for all complexes using the new corrected electrostatics. The results (Table 2; column *corrected charges*) clearly improved by using solvation-corrected electrostatics. The near native solution now ranked first for two of the six test cases (1taw and 1brc), and was found within the 6 lowest energy solutions in more than half of the cases.

2.2 Interface refinement of docking solutions

The resulting solutions obtained after rigid-body docking with uncorrected electrostatics, were further refine by optimizing the interface side-chains (using uncorrected electrostatics during the refinement). The results (Table 3; column *uncorrected charges*) show that the refinement of interacting side-chains improved the rank of the near native solution in all cases except one (1brc). Moreover, in two of the complexes, the near native solution is now ranked in first place (2sni and 1taw).

Table 3. Re-evaluation of docking solutions after interface refinement

Complex	*uncorrected charges*		*corrected charges*	
	Rank	r.m.s.d. (Å)	Rank	r.m.s.d. (Å)
1ca0	2 of 233	1.1	1 of 228	1.2
1cbw	3 of 220	1.1	1 of 231	0.7
2sni	1 of 243	2.6	1 of 220	2.7
1taw	1 of 232	2.8	1 of 228	2.9
3tgi	62 of 238	0.5	1 of 243	0.6
1brc	33 of 218	3.1	1 of 237	1.8

When the docking solutions obtained using the corrected electrostatics are refined (also using the solvation-corrected electrostatics), the ranking is further improved (Table 3; column *corrected charges*). In all cases, the near native solution is now the lowest energy conformation.

3 Discussion

Electrostatic forces play an important role in protein-protein interactions. The solvation effects present a major difficulty in modeling electrostatic interactions. We propose and apply a simple screened charge model to account for the attenuation of electrostatic interactions by the solvent in docking simulations. For the six complexes in the test set, the proposed correction results in a dramatic improvement of the ranking of the near-native solution. Already at the rigid-body docking step, the near-native solution is ranked first for two complexes, versus none in the case of uncorrected electrostatics. The influence of the correction on the refinement step is relatively minor. As can be seen in Figure 2 (refinement with uncorrected electrostatics) and Figure 3 (refinement with corrected electrostatics), interface refinement of rigid-body solutions removes clashes from interacting side-chains in the wrong conformation. Refinement is able to mimic the induced fit of the association, and the final conformation of interacting ligand side-chains is very close to the native conformation. However, the effect of the solvation correction is essential for the accurate scoring of the near-native solution. The correction apparently helps to remove false positives, e.g. conformations with extended interacting surfaces and over-estimated interactions of solvated charges. The lowest energy conformation (21.8 Å r.m.s.d.) (Figure 2) found with uncorrected electrostatics presents numerous electrostatic interactions of highly exposed residues for which the attraction is over-estimated. On the contrary, the near-native solution is correctly ranked as the lowest energy conformation when the corrected electrostatic term is used.

As this work shows, failing to consider solvation effects when calculating electrostatics in protein-protein docking simulations often results in numerous misdocked configurations. We show here a way of including such solvation effects in electrostatic energy which greatly improves docking accuracy without significant computational overhead.

A broader docking test using this novel approach for electrostatics is reported in an upcoming publication[21].

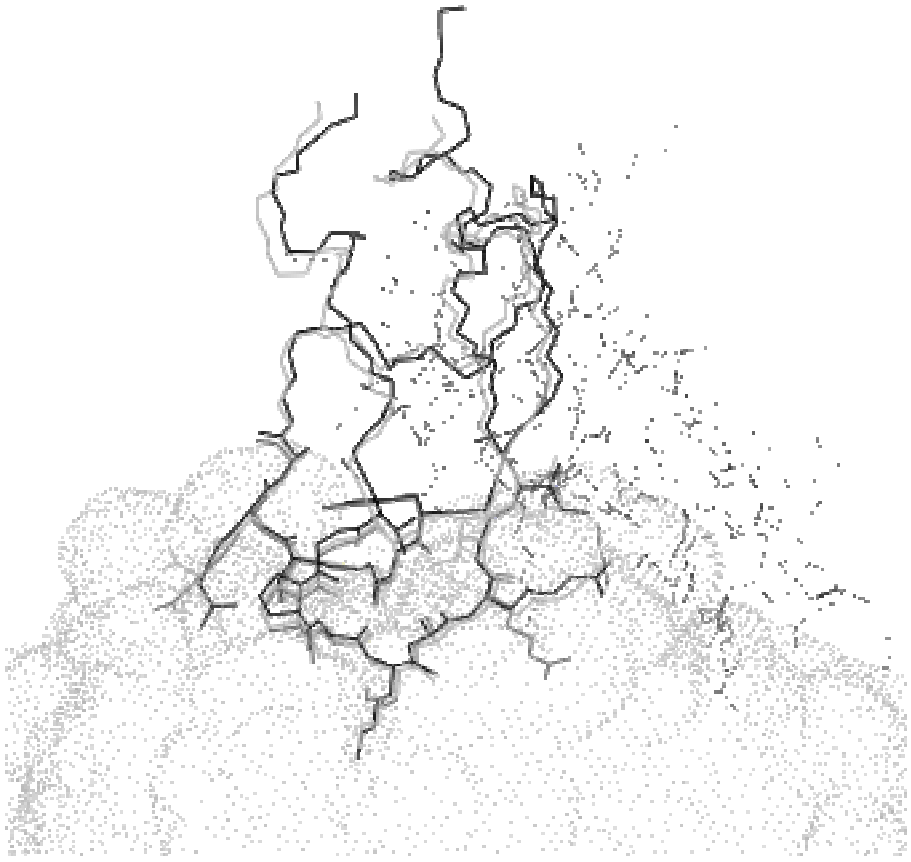


Figure 2. Docking results for trypsin/BPTI (PDB 3tgi) using uncorrected electrostatics. Near native solution obtained after rigid-body docking (dark gray) and after refinement (black) compared to the crystallographic structure (light gray). The lowest energy solution obtained after refinement is shown in thin lines. The trypsin surface is shown as light gray dots.

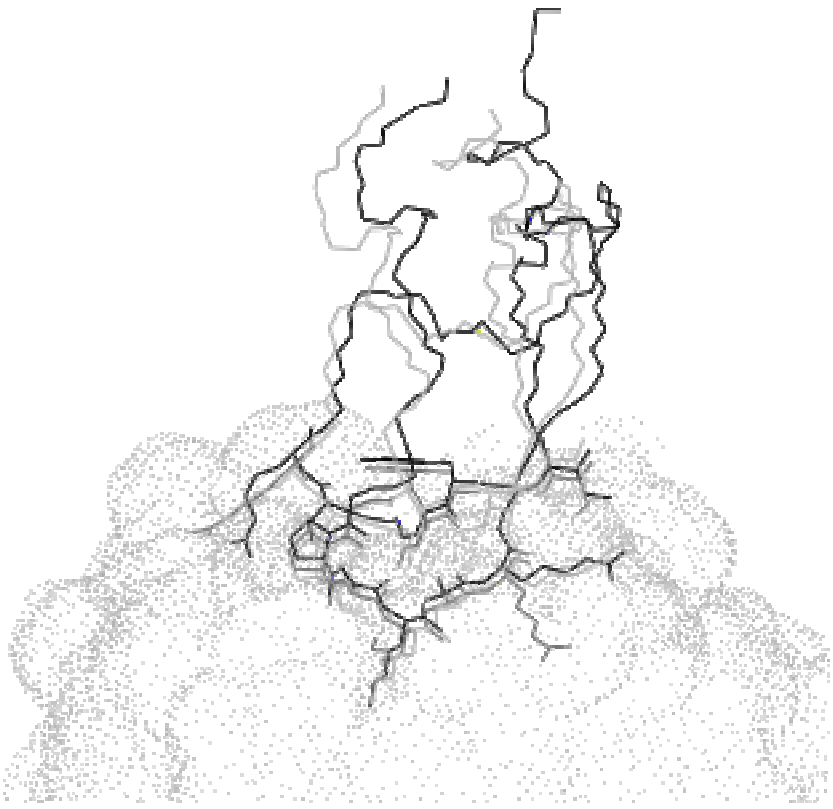


Figure 3. Docking results for trypsin/BPTI (PDB 3tgi) using corrected electrostatics. The near native solution obtained after rigid-body docking (dark gray) and after refinement (black) compared to the crystallographic structure (light gray). The near native solution obtained after refinement is the lowest energy solution. The trypsin surface is shown as light gray dots.

4 Acknowledgements

This work was supported by NIH Grant R01 GM55418. We wish to thank Brian Marsden for his continuous support of the LINUX clusters and Molsoft for making the ICM program available for the project.

References

1. Sanchez, R. & Sali, A. "Large-scale protein structure modeling of the Saccharomyces cerevisiae genome"*Proc. Natl. Acad. Sci. USA* **95,** 13597-602 (1998)
2. Sternberg, M. J., Gabb, H. A. & Jackson, R. M. "Predictive docking of protein-protein and protein-DNA complexes"*Curr. Opin. Struct. Biol.* **8,** 250-6 (1998)
3. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. & Vakser, I. A. "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques"*Proc. Natl. Acad. Sci. USA* **89,** 2195-9 (1992)
4. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. "A geometric approach to macromolecule-ligand interactions"*J. Mol. Biol.* **161,** 269-88 (1982)
5. Norel, R., Petrey, D., Wolfson, H. J. & Nussinov, R. "Examination of shape complementarity in docking of unbound proteins"*Proteins* **36,** 307-17 (1999)
6. Totrov, M. & Abagyan, R. "Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy"*Nat. Struct. Biol.* **1,** 259-63 (1994)
7. Strynadka, N. C., Eisenstein, M., Katchalski-Katzir, E., Shoichet, B. K., Kuntz, I. D., Abagyan, R., Totrov, M., Janin, J., Cherfils, J., Zimmerman, F., Olson, A., Duncan, B., Rao, M., Jackson, R., Sternberg, M. & James, M. N. "Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase"*Nat. Struct. Biol.* **3,** 233-9 (1996)
8. Abagyan, R. & Argos, P. "Optimal protocol and trajectory visualization for conformational searches of peptides and proteins"*J. Mol. Biol.* **225,** 519-32 (1992)
9. Abagyan, R., Totrov, M. & Kuznetsov, D. "ICM: a new method for structure modeling and design: Applications to docking and structure prediction from the distorted native conformation"*J. Comp. Chem.* **15,** 488-506 (1994)
10. Abagyan, R. & Totrov, M. "Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins"*J. Mol. Biol.* **235,** 983-1002 (1994)

11. Goodford, P. J. "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules"*J. Med. Chem.* **28,** 849-57 (1985)
12. Totrov, M. & Abagyan, R. "Flexible protein-ligand docking by global energy optimization in internal coordinates"*Proteins* **Suppl 1,** 215-20 (1997)
13. Totrov, M. & Abagyan, R. (2001) in *Drug-receptor thermodynamics: Introduction and applications*, ed. Raffa, R. B. (John Wiley & Sons, Ltd., pp. 603-624.
14. McCammon, J. A., Wolynes, P. G. & Karplus, M. "Picosecond dynamics of tyrosine side chains in proteins"*Biochemistry* **18,** 927-42 (1979)
15. Zauhar, R. J. & Morgan, R. S. "A new method for computing the macromolecular electric potential"*J. Mol. Biol.* **186,** 815-20 (1985)
16. Totrov, M., Abagyan, R. "Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide"*Biopolymers* **60,** 124-133 (2001)
17. MolSoft (2000) (MolSoft LLC, San Diego).
18. Nemethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S. & Scheraga, H. A. "Energy parameters in polypeptides. 10. Improved geometric parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides"*J. Phys. Chem.* **96,** 6472-84 (1992)
19. Metropolis, N. A., Rosenbluth, A. W., Rosenbluth, N. M., Teller, A. H. & Teller, E. "Equation of state calculations by fast computing machines"*J. Chem. Phys.* **21,** 1087-92 (1953)
20. Wesson, L. & Eisenberg, D. "Atomic solvation parameters applied to molecular dynamics of proteins in solution"*Protein Sci.* **1,** 227-35 (1992)
21. Fernandez-Recio, J., Totrov, M., Abagyan, R. "Soft protein-protein docking in internal coordinates" *Protein Sci.* (In press)