

SHORT COMMUNICATION

Atomic contacts in protein structures. A detailed analysis of atomic radii, packing, and overlaps

Daniel Seeliger and Bert L. de Groot*

Computational Biomolecular Dynamics Group, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

ABSTRACT

A rigorous quantitative assessment of atomic contacts and packing in native protein structures is presented. The analysis is based on optimized atomic radii derived from a set of high-resolution protein structures and reveals that the distribution of atomic contacts and overlaps is a structural constraint in proteins, irrespective of structural or functional classification and size. Furthermore, a newly developed method for calculating packing properties is introduced and applied to sets of protein structures at different levels of resolution. The results show that limited resolution yields decreasing packing quality, which underscores the relevance of packing considerations for structure prediction, design, dynamics, and docking.

Proteins 2007; 68:595–601.
© 2007 Wiley-Liss, Inc.

Key words: packing quality; van-der-Waals-radii; protein design; structure validation.

INTRODUCTION

Billions of years of evolution optimized proteins to fulfill their functions efficiently. Regardless whether the protein functions as enzyme, molecular motor, transport protein, or receptor, a prerequisite for optimal function is a fine-tuned structural and dynamical framework, either directly or indirectly provided by the native structure of the protein. An important, but as yet unresolved question is which functional constraints exactly are imposed on a protein structure. Sequence and structure conservation patterns provide valuable hints in this respect, like the conservation of the structure in the catalytic site of an enzyme. However, such information is typically local and restricted to a specific class of proteins. The same holds for other localized structural constraints like disulphide bridges or specific salt bridges. Hence, the role of global structural determinants underlying or supporting function remains to be determined.

Protein design and engineering studies suggest a crucial role for packing in protein stability and function,^{1–4} including exact complementarity of neighbouring side chains.^{5–7} Even conservative mutations of single amino acids can lead to destabilizations.^{8,9} Additionally, the inclusion of an explicit packing term in protein design algorithms has significantly improved the accuracy of designed predictions,⁶ indicating that optimal packing is a crucial factor in protein structures. Packing densities in protein cores have been described as high and comparable to solid crystals.^{10,11} However, beyond average densities and free volume considerations,¹² the exact packing extent, in terms of atomic contacts, remains unknown.

Here, we have developed an approach to quantitatively determine the packing efficiency of a large set of protein structures at different levels of resolution. A “packing score” is introduced that allows a robust assessment of the degree of packing effi-

The Supplementary Material referred to in this article can be found at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/>

Grant sponsor: CSIR; Grant number: CMM0017

*Correspondence to: Bert L. de Groot, Computational Biomolecular Dynamics Group, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. E-mail: bgroot@gwdg.de

Received 15 September 2006; Revised 10 January 2007; Accepted 31 January 2007

Published online 17 May 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21447

ciency, resting on a set of atomic radii derived from a set of high resolution protein structures. We show that the distribution of close contacts and overlaps in protein structures is invariant and highly conserved in high-resolution X-ray structures, regardless of function, size, and secondary structure.

The implications for protein structure validation, protein dynamics, structure prediction, and design are discussed.

MATERIALS AND METHODS

Optimal packing in molecular systems is characterized by a maximum number of interatomic contacts. In proteins, the maximally attainable packing efficiency is primarily limited by the distribution of unequally sized atoms (C,H,N,O,S), by topological restraints imposed by the connectivity between atoms, and by secondary/tertiary structure restraints. To assess the relative degree of packing in native protein structures, we therefore quantified the packing efficiency, evaluated this packing score for a large number of proteins, and compared the results with a synthetic reference. The reference was constructed from a set of 1000 freely rotatable amino acids in solution, distributed according to the frequency as observed in natural proteins (see Supplementary Materials for details). This system was subjected to 20 ns of molecular dynamics simulation. Snapshots from this simulation were cooled down to 100 K with simulated annealing. As this reference shares the restrictions of native protein structures of unequally sized atoms and connectivities, but has no restrictions due to secondary and tertiary structure and also has no surface or active site that display poorer packing properties, this reference can be considered as upper limit of the packing efficiency for natively folded proteins, and hence may serve to estimate the relative packing efficiency of protein structures.

In contrast to previous studies,^{10,11,13–16} we do not consider packing in terms of occupied volume fractions. Rather, we focus on the thermodynamically determined distribution of favorable atomic contacts and unfavorable overlaps. A set of atomic radii was determined from a set of 106 high resolution protein structures (resolution < 1.2 Å). Contacts were counted for closely interacting ($d_{ij} \leq r_{ij} \leq d_{ij}^*1.3$), but nonoverlapping atoms. The requirement of maximizing the number of contacts while minimizing the number of overlaps ($r_{ij} < d_{ij}$) ensures counting of true contacts in favor of any secondary maxima. A full set of protein atomic radii was obtained by iteratively adapting the atom radii for the different atom types²⁹ (see Supplementary Material for a detailed description of the method).

The obtained radii were used to evaluate a packing score for a large set of protein structures at different levels of resolution. Nonprotein residues like water and ions

were neglected. Note that this does not affect the obtained radii. The packing score was defined as the average volume-weighted number of contacts per atom minus the average volume-weighted number of overlaps (Supplementary Materials).

For the synthetic reference ensemble of structures, built from the final configurations of the simulated annealing simulations, the same procedure for optimizing atomic radii was employed. Using these radii (data not shown), packing scores were calculated for the synthetic reference ensemble. The average value of these scores was scaled to 1.0 and serves as reference for the packing scores calculated from the experimental structures. The statistical error as estimated from the standard deviation in the ensemble is about 0.01, represented in Figure 1 by the thickness of the red line.

Hydrogen atoms

Instead of taking the hydrogen positions that are available for a number of high-resolution X-ray structures, we generated hydrogen positions using the HB2NET module of WHAT IF.¹⁷ We chose this strategy as only few data sets are complete, and because the bond lengths for C—H, N—H, and O—H are systematically underestimated in X-ray diffraction.¹⁸ A further advantage of the employed hydrogen placement algorithm is that it evaluates different protonation states and optimizes the hydrogen bond network within the structure, including side-chain flips of histidine, glutamine, and asparagine residues.

RESULTS AND DISCUSSION

Packing quality in protein structures

Packing scores were calculated for sets of protein structures determined by X-ray crystallography and NMR. X-ray structures were evaluated at different levels of resolution (see Supplementary Materials). NMR structures were compared with refined ensembles from the DRESS¹⁹ database (always the first model was taken from an NMR-ensemble, for this usually represents the lowest energy configuration). The results relative to the synthetic reference are shown in Figure 1.

Remarkably, packing scores of up to 88% of the synthetic reference (in red) were observed, indicating a high packing density for natively folded protein structures resolved at high-resolution. With decreasing resolution the packing efficiency is observed to decrease. While the packing scores for X-ray structures are located in a rather narrow range, values for NMR-structures (blue marks) show much more spread. This behavior is further exemplified for two structures of staphylococcal nuclease, of which one (PDB 1ey4) has been resolved by X-ray crystallography (resolution, 1.6 Å) and the other one by NMR (PDB 1jor). The right panel of Figure 1 shows the

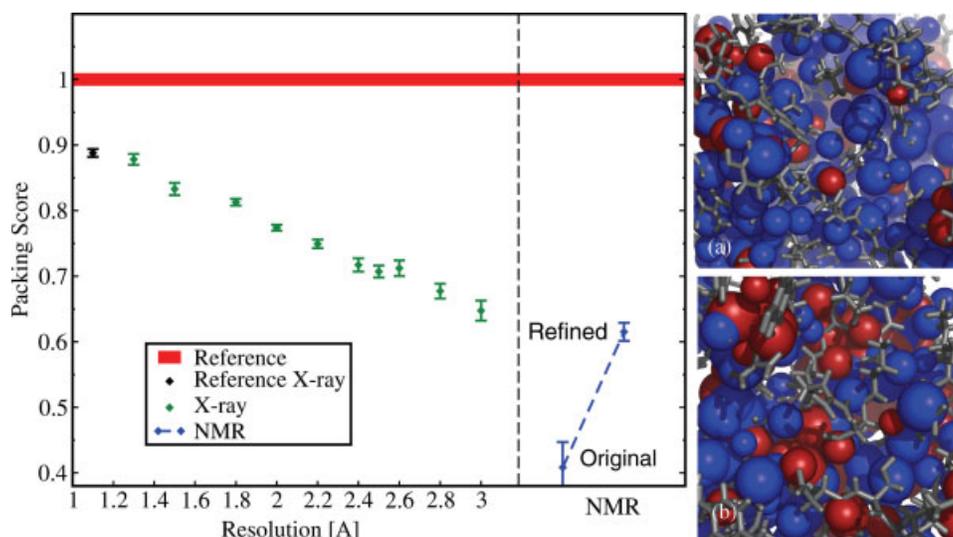


Figure 1

Left panel: Packing scores. Red line (reference): line thickness represents the standard deviation; black: the ensemble of high resolution structures that were used to derive the atomic radii; green: X-ray structures at different levels of resolution; blue: Ensemble of NMR-structures original from the PDB and the same structures from the DRESS database. Right panel: Two structures of staphylococcal nuclease. (a) PDB 1ey4, resolved by X-ray crystallography (Resolution, 1.6 Å). (b) PDB 1jor, resolved by NMR. The blue colored atoms are well packed and embedded in their local environment. Red colored atoms cause overlaps with their neighbors.

difference in atomic packing for fragments of the two structures. In the X-ray structure, apart from surface exposed groups, all atoms are well-packed by nearly ideal contacts (overall packing score, 0.76). In the NMR structure of the same protein, the packing is found to be less ideal because of more overlaps and fewer contacts (overall packing score, 0.45).

The distribution of atomic contacts can be illustrated by a reduced radial distribution function (RRDF), which

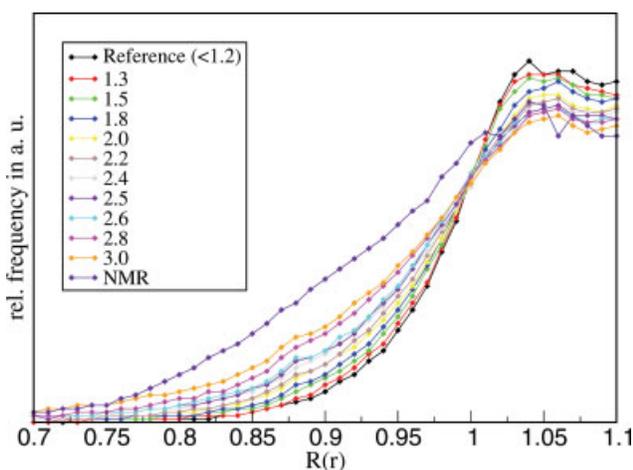


Figure 2

Reduced radial distribution functions (RRDF).

is a standard radial distribution function normalized to the ideal contacts distance (see Supplementary Material). This function displays all close contacts within a protein structure or an ensemble of structures. Values lower than 1.0 represent energetically unfavourable overlaps that should occur infrequently according to the Boltzmann distribution. Figure 2 shows the RRDFs of the same structure ensembles as used in Figure 1.

At high resolution the curves are steeper, representing a favorable ratio of contacts and overlaps. Furthermore, the plot shows that the distribution of atomic contacts in NMR-structures differs significantly from those in X-ray structures. While the curves for the different levels of resolution basically differ in steepness, the curve corresponding to the NMR-structures shows a systematic devi-

Table I

Comparison of Packing Scores for Identical Proteins

PDB ID	Resolution (Å)	Packing score	PDB ID	Resolution (Å)	Packing score
1act	2.8	0.46	2act	1.7	0.89
2ape	2.5	0.29	4ape	2.1	0.73
1lzm	2.4	0.37	2lzm	1.7	0.82
2tln	2.3	0.55	8tln	1.6	0.84
1alp	2.8	0.39	2alp	1.7	0.90
2fxb	2.3	0.59	1iqz	0.92	0.90

Left: obsolete, lower resolution structure; right: current higher resolution PDB entry.

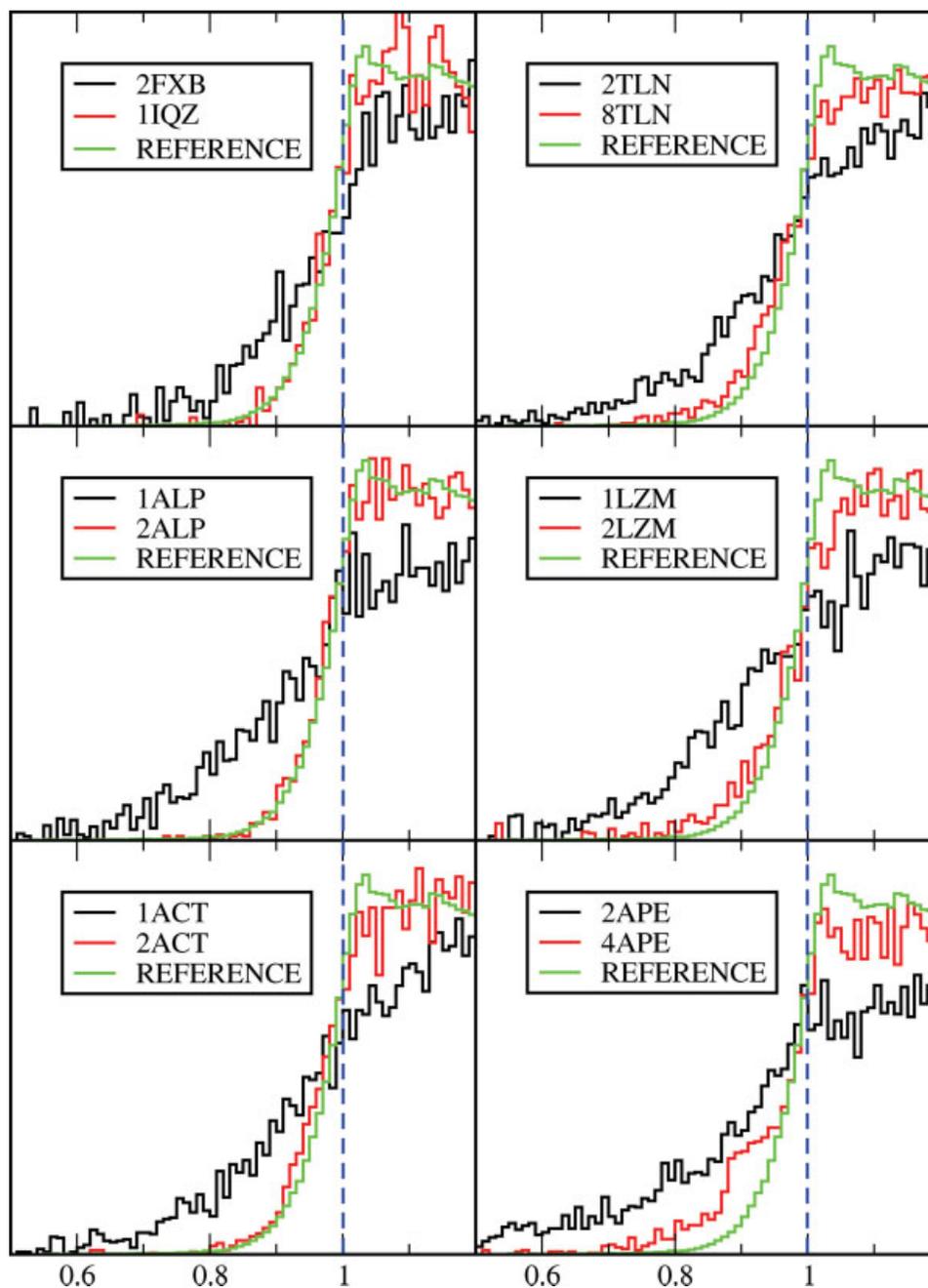


Figure 3

Comparison of identical protein structures at different levels of resolution. The black curves represent the reduced radial distribution functions of obsolete protein structures. The red curves represent the same function of the current PDB entries of these proteins. The green curve shows the RRDf of the reference set of 106 high-resolution X-ray structures.

ation. The amount of overlaps is much higher, which can be interpreted as systematic overpacking, in agreement with previous findings.^{20–25}

The question arises whether the observed resolution dependence reflects protein flexibility or, rather, a resolution-imposed coordinate uncertainty. In other words, could inherent flexibility or disorder that results in lim-

ited resolution cause a nonoptimal packing (“packing limits resolution”) or does limited resolution prevent building of an accurate well-packed model (“resolution limits packing”)? To address this question, we investigated several cases of the same protein structure solved at different levels of resolution. Comparison of packing scores of these identical protein structures shows that

Table II

Atomic Radii Derived from a Set of 106 High-Resolution X-Ray Structures

Atom type	Radius (Å)	Description	Atom type	Radius (Å)	Description
H0	1.19	Nonpolar hydrogens	CH2P	1.47	C _{β,γ,δ} in P
HAR	1.14	Aromatic hydrogen	CY	1.87	C _γ in Y
HA	1.03	H _α	CY2	1.63	C _η in Y
H	1.05	Polar hydrogen	CF	1.83	C _γ in F
HC	0.58	Hydrogen in charged groups (R,K)	CDR	1.69	C _δ in R
HDR	0.67	H _δ in arginine	CR1H	1.75	C _{δ2} in H
C	1.43	Carbon in C=O	CRHH	1.63	C _{ε1} in H
CA	1.48	C _α	O	1.41	Oxygen in C=O
CH1E	1.92	Aliphatic carbon with 1 hydrogen	OC	1.33	Oxygen in COO ⁻
CH2E	1.89	Aliphatic carbon with 2 hydrogens	OH1	1.31	Oxygen in C—O—H
CH3E	1.81	Aliphatic carbon with 3 hydrogens	NH1	1.37	Nitrogen with 1 hydrogen
CR1E	1.81	Aromatic carbon	NH2	1.45	Nitrogen with 2 hydrogens
CR1W	1.76	C _{ε2} , C _{η2} in W	NH3	1.35	Nitrogen with 3 hydrogens
C5	1.76	C _γ in H	NC2	1.45	N _ε in R
C5W	1.86	C _γ in W	NHS	1.40	Unprotonated N in H
CW	1.74	C _ε in W	SM	1.79	S in M
CH2G	1.76	C _α in G	S	1.83	S in C

packing scores significantly increase at higher resolution (Table I). The distribution of overlaps, represented by the left branch of the RRDF ($\mathcal{R}(r) < 1$), is a structural invariant for all protein structures. Figure 3 shows RRDFs of identical proteins at different levels of resolution. The curves of the higher resolution versions of these protein structures are remarkably close to the reference curve, strongly supporting the “resolution limits packing” scenario and not the “packing limits resolution” scenario. Hence, overlap distributions and packing considerations could be used as quality check for protein structures. Additionally, these results suggest that a rigorous packing term may aid structure refinement.

Atomic radii

A closer look at the derived contact radii listed in Table II reveals that most carbon, nitrogen, and oxygen radii are smaller compared with those from previous work.^{10,11,13–16} This is mostly caused by the use of explicit hydrogen atoms in this work. Also in comparison to Lennard-Jones parameters from force-fields, our atomic radii are generally smaller. This is due to the fact that, in force-fields, the local geometry of atoms is simultaneously determined together with interactions, particularly electrostatic interactions. Our approach aims at a geometrical description that reflects the optimal contact distance between atom pairs as a combined effect of all interactions.

A number of systematic deviations of atom radii became evident during the optimization that are found to be caused by the original classification of the atom types. Hence, a number of additional atom types were introduced. For instance an additional atom type was introduced for C_α atoms (atom type CA) since we found

that C_α atoms form much closer contacts than other aliphatic carbon atoms. Likewise hydrogen atoms connected to C_α atoms (atom type HA) form closer contacts than other nonpolar hydrogens making them more similar in size to polar hydrogens (atom type H). This example of a systematic protein-specific deviation indicates the significance of a protein-specific set of atomic radii derived from atomic-resolution protein structures.

Additionally, a set of specific combinations of atom types was defined to realistically account for electrostatics like small hydrogen-bond distances. The combinations are listed in Table III. The very small radius for charged hydrogens (atom type HC) is remarkable but may in part be due to the small number of contacts that these atoms form. Hence, the statistics for this atom type is limited. Note, however, that these atoms usually reside on the protein surface and are only infrequently involved

Table III

Lower Bounds for Distances of Specific Atom Type Combinations

Atom	Types	<i>D</i> (Å)	Atom	Types	<i>D</i> (Å)
O	O	3.3	HC	NHS	1.84
O	H	1.86	H	NHB	2.00
O	OC	2.84	HC	NHB	1.95
OH1	O	2.64	O	NC2	2.82
O	HC	1.70	O	NH2	2.84
OH1	H	1.62	O	NH3	2.60
OC	HC	1.74	O	NHS	2.66
OH1	HC	1.70	NH1	NHS	2.88
OC	H	1.60	NH2	NHS	3.00
O	NH1	2.82	NH3	NHS	2.84
H	NHS	2.0	O	CA	3.18
HA	O	2.30			

in intraprotein contacts. Therefore, these parameters do not significantly affect the packing scores.

Implications for ligand docking and protein dynamics

The high degree of packing observed in high-resolution protein structures renders it likely that packing density also plays a major role in protein–ligand interactions. Indeed, surface complementarity is frequently used as a criterion for the assessment of potential ligands in molecular docking programs.^{26,27} Accurate atomic radii together with a robust evaluation of packing can therefore be expected to significantly enhance the virtual screening of protein–ligand complexes.

In addition to structure, dynamics is often crucial for protein function, like in allosteric transitions or molecular recognition. Protein dynamics can be regarded complementary to protein structure: Of all possible atomic degrees of freedom, only those that are not restricted by interatomic interactions within the structure can contribute to conformational dynamics. Indeed, it has been found from experimental and simulated structures that only a limited number of “essential” collective degrees of freedom usually dictate the functional dynamics of a protein, indicating that all other degrees of freedom are effectively dampened or constrained by the structure.²⁸ Again, apart from topological constraints and specific interactions like hydrogen bonds, the underlying structural determinants were largely unknown so far. The packing properties of protein structures reported here shows that a major determinant underlying the restriction of protein dynamics to a small number of relevant collective modes are due to packing restraints.

CONCLUSIONS

Our results show that high resolution natively folded protein structures display a packing efficiency close to that of a condensed phase of free amino acids, regardless of the protein’s size and structural and functional origin. Efficient packing therefore represents a universal feature of protein structure. Additionally, efficient packing likely facilitates the restriction of protein dynamics to a limited number of modes essential for function. The calculated packing scores suggest that atomic packing is a structural constraint on protein architecture, offering novel opportunities for the interpretation of sequence alignments and genome data. The fact that packing efficiency shows a marked resolution dependence indicates that rigorous inclusion of an accurate packing term can be expected to enhance structure refinement at low and intermediate resolution levels. Furthermore, it underscores the significance of packing considerations for protein structure prediction, design, and docking.

REFERENCES

- Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. *Science* 1997;278:82–87.
- Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, Serrano L. Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat Struct Biol* 2002;9:485–493.
- Kuhlmann B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–1368.
- Walsh STR, Cheng H, Bryson JW, Roder H, DeGrado WF. Solution structure and dynamics of a de novo designed three-helix bundle protein. *PNAS* 1999;96:5486–5491.
- Desjarlais JR, Handel TM. De novo design of hydrophobic cores of proteins. *Prot Sci* 1995;4:2006–2018.
- Dahiyat BI, Mayo SL. Probing the role of packing specificity in protein design. *PNAS* 1997;94:10172–10177.
- Kellis JT, Nyberg K, Fersht AR. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 1989;28:4914–4922.
- Godoy-Ruiz R, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz J. A stability pattern of protein hydrophobic mutations that reflects evolutionary structural optimization. *Biophys J* 2005;89:3320–3331.
- Chen J, Stites WE. Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* 2001;40:15280–15289.
- Richards FM. Areas, volumes, packing, and protein structure. *Ann Rev Biophys Bioeng* 1977;6:151–176.
- Chothia C. Structural invariants in protein folding. *Nature* 1975;254:304–308.
- Liang J, Dill KA. Are proteins well-packed? *Biophys J* 2001;81:751–766.
- Li AJ, Nussinov R. A set of van der waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Prot: Struct, Funct Genet* 1998;32:111–127.
- Tsai J, Taylor R, Chotia C, Gerstein M. The packing density in proteins: standard radii and volumes. *J Mol Biol* 1999;290:253–266.
- Bondi A. Van der waals volumes and radii. *J Phys Chem* 1964;68:441–451.
- Iijima H, Jr JBD, Marshall GR. Calibration of effective van der waals atomic contact radii for proteins and peptides. *Prot: Struct, Funct Genet* 1987;2:330–339.
- Hoof RW, Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Prot: Struct Funct Genet* 1996;26:363–376.
- Rowland RS, Taylor R. Intermolecular nonbonded contact distances in organic crystal structures: comparison with distances expected from van der Waals radii. *J Phys Chem* 1996;100:7384–7391.
- Nabuurs SB, Nederveen AJ, Vranken W, Doreleijers JF, Bonvin AM, Vuister GW, Vriend G, Spronk CA. Dress: a database of refined solution NMR structures. *Prot: Struct, Funct Bioinform* 2004;55:483–486.
- Word JM, Lovell SC, LaBean TH, Tayler HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* 1999;285:1711–1733.
- Ratnaparkhi GS, Ramachandran S, Udgaonkar JB, Varadarajan R. Discrepancies between NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR are unreliable. *Biochemistry* 1998;37:6958–6966.
- Doreleijers J, Rullmann JAC, Kaptein R. Quality assessment of NMR structures: a statistical survey. *J Mol Biol* 1998;281:149–164.

23. Doreleijers JE, Vriend G, Raves ML, Kaptein R. Validation of nuclear magnetic resonance structures of proteins and nucleic acids: hydrogen geometry and nomenclature. *Prot: Struct, Funct and Genet* 1999;37:404–416.
24. Spronk CAEM, Ringe JP, Hilbers CW, Vuister GW. Improving the quality of protein structures derived by NMR spectroscopy. *J Biol NMR* 2002;22:281–289.
25. Garbuzynskiy SO, Melnik BS, Lobanov MY, Finkelstein AV, Galzit-skaya OV. Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? *Prot: Struct, Funct Genet* 2005;60:139–147.
26. Shoichet BK, Kuntz ID. Protein docking and complementarity. *J Mol Biol* 1991;221:327–346.
27. Fischer D, Lin SL, Wolfson HL, Nussinov R. A geometry-based suite of molecular docking processes. *J Mol Biol* 1995;248:459–477.
28. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. *Prot: Struct Funct Genet* 1993;17:412–425.
29. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 1996;118:11225–11236.