Journal of Biomolecular NMR, **22:** 317–331, 2002.
KLUWER/ESCOM

Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water

Bin Xia, Vickie Tsui, David A. Case, H. Jane Dyson* & Peter E. Wright*
Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92122, U.S.A.

Received 5 October 2001; Accepted 18 January 2002

Key words: hydrogen bond, solvent refinement, structure determination method, structure refinement

Abstract

The inclusion of explicit solvent water in molecular dynamics refinement of NMR structures ought to provide the most physically meaningful accounting for the effects of solvent on structure, but is computationally expensive. In order to evaluate the validity of commonly used vacuum refinements and of recently developed continuum solvent model methods, we have used three different methods to refine a set of NMR solution structures of a medium sized protein, *Escherichia coli* glutaredoxin 2, from starting structures calculated using the program DYANA. The three different refinement protocols used molecular dynamics simulated annealing with the program AMBER in vacuum (VAC), including a generalized Born (GB) solvent model, and a full calculation including explicit solvent water (WAT). The structures obtained using the three methods of refinements were very similar, a reflection of their generally well-determined nature. However, the structures refined with the generalized Born model were more similar to those from explicit water refinement than those refined in vacuum. Significant improvement was seen in the percentage of backbone dihedral angles in the most favored regions of ϕ,ψ space and in hydrogen bond pattern for structures refined with the GB and WAT models, compared with the structures refined in vacuum. The explicit water calculation took an average of 200 h of CPU time per structure on an SGI cluster, compared to 15–90 h for the GB calculation (depending on the parameters used) and 2 h for the vacuum calculation. The generalized Born solvent model proved to be an excellent compromise between the vacuum and explicit water refinements, giving results comparable to those of the explicit water calculation. Some improvement for ϕ and ψ angle distribution and hydrogen bond pattern can also be achieved by energy minimizing the vacuum structures with the GB model, which takes a much shorter time than MD simulations with the GB model.

Introduction

Water plays an important role in the structure, stability, and function of biological macromolecules (Nicholls, 2000). The role of water in biological function has generally been described by the observation of well-ordered water molecules in X-ray crystal structures (Dauter et al., 1997). Long-lived water molecules can be localized by NMR spectroscopy (Otting and Wüthrich, 1989). However, the role of water has traditionally been largely ignored in the calculation of NMR structures: although the constraints are derived from experimental data collected from samples in solution, the refinement of solution structures through molecular dynamics (MD) simulated annealing is generally carried out in vacuum with reduced molecular charge, with no explicit accounting for the effect of solvent on the structure.

To simulate solvation effects on biological macromolecules, both explicit and implicit solvent models have been developed. For the explicit solvent model, MD simulation is carried out in the presence of a box

*To whom correspondence should be addressed. E-mail: wright@scripps.edu; dyson@scripps.edu

of discrete water molecules. The box needs to be sufficiently large to keep at least 10 Å between the edges of the box and the closest solute atoms (Cheatham and Kollman, 1996). This will require thousands or even tens of thousands of water molecules in the system. Such an MD calculation is normally very time consuming and computationally expensive. The implicit solvent model, or continuum solvation model, treats the solvent (usually water) as a high dielectric continuous medium surrounding the solute near its van der Waals surface, interacting with charges on the low dielectric solute molecule (Cramer and Truhlar, 1999). In principle, this model should account for the effects of solvation with significantly lower computation time, since it includes only the atoms from the solute.

Different continuum solvation models have been developed over the years. Among them, the generalized Born (GB) model has received considerable attention (Still et al., 1990; Qiu et al., 1997; Onufriev et al., 2000). The GB model is an approximation to the Poisson-Boltzman equation, and provides a more effective simulation to the solvation effect. It has been recently implemented into several molecular dynamic modeling programs, such as CHARMM and AMBER, as an effective force field for MD simulation of macromolecule structures (Bashford and Case, 2000). Some applications of GB MD simulation on peptides, proteins, and nucleic acids structures have appeared and show very encouraging results (Dominy and Brooks, 1999; Schaefer et al., 1998; Williams and Hall, 1999; Tsui and Case, 2000, 2001; Cornell et al., 2001; Rapp and Friesner, 1999; Calimet et al., 2001), but have not previously been applied in the context of NMR refinements. Here, we report a detailed comparison of *E. coli* glutaredoxin 2 (Grx2) structures refined in vacuum, in GB mode, and in explicit water using AMBER (version 6) (Case et al., 1999).

Experimental methods

NMR constraints

A total of 200 structures were generated from DYANA (Güntert et al., 1997) using 501 dihedral angle constraints and 3757 unambiguously assigned distance constraints, which include 475 intra-residue, 816 sequential, 1277 medium range, and 1189 long range constraints (Xia et al., 2001). The 100 structures with lowest DYANA target function values were used for further MD refinement in AMBER. A total of 501 dihedral angle constraints and 7874 distance constraints were used for AMBER calculation. The distance constraints include: 2312 intra-residue, 1494 sequential, 1408 medium range, 1492 long range, and 1168 ambiguously assigned constraints. These constraints were used for all simulations except for the first 300 ps of each explicit water simulation.

MD simulations

All MD simulations were performed on a cluster of SGI Origin 2400 and 3800 servers with a total 256 500 MHz R12000 CPUs and 128 GBytes of memory.

For the MD refinement in vacuum, the 100 DYANA structures were first energy minimized in vacuum for 1000 steps without any constraints, and then two cycles of 30 ps MD simulated annealing were performed in vacuum, from 1000 K to 0 K. For the GB refinement, 1 ps energy minimization with GB model was first performed on the DYANA structures, followed by two cycles of 30 ps simulated annealing, from 800 K to 0 K. A 2000 step energy minimization with NMR constraints in vacuum was applied after the MD simulated annealing for the top 20 VAC, and a 2000 step energy minimization using GB model was applied to the top 20 GB structures.

For the explicit water refinement, either the top 20 GB structures or the top 20 VAC structures were used as initial structures. A solvent box whose edges are 10 Å from the closest protein atoms was added to each structure. The number of water molecules added for each structure varied from 8088 to 10732. The initial structures were subjected to 3 cycles of 40 ps of MD with constant volume, gradually heating up from 0K to 50K, 50K to 100 K, and 100 K to 300 K, with 5 kcal/(mol Å^2) harmonic constraints on the protein to their starting structures. The time constant for the heating bath coupling is 1 ps, and the pressure relaxation time is 0.2 ps. This was followed by 2 cycles of 40 ps constant pressure MD with the pressure relaxation time of 0.2 ps and 1.0 ps, with 5 kcal/(mol Å^2) harmonic constraints on the protein to their starting structures, at 300 K. Then, another 40 ps constant volume MD with 5 kcal/(mol Å^2) harmonic constraints on the protein was applied. This was followed by a 60 ps constant volume MD with decreased force constant of harmonic constraints on the protein (0.5 kcal/(mol Å^2). These two steps are at 300 K, and the time constant for the heating bath coupling is 1.0 ps, and the pressure relaxation time is 1.0 ps, with time-step 2 fs. Finally, a 100 ps MD simulation without harmonic restraints to the initial

Table 1. Structural statistics for Grx2 structures refined with AMBER under various conditions

MD refinement	VAC	GB	WATV	WATG
Backbone heavy atom RMSD from mean structure (Å)				
All residues	0.61	0.72	0.60	0.65
Regular secondary structure	0.49	0.49	0.47	0.46
All heavy atom RMSD from mean structure (Å)				
All residues	1.02	1.05	0.98	0.99
Regular secondary structure	0.93	0.91	0.89	0.87
ϕ and ψ angle distribution (%)				
Residues in most favorable region	86.6	90.8	91.0	91.4
Residues in additional allowed region	12.2	8.3	8.1	7.7
Residues in generously allowed region	0.7	0.4	0.3	0.3
Residues in disallowed region	0.5	0.5	0.5	0.6
AMBER energy				
Average AMBER energy (kcal/mol)	−3278	−8386	−8347	−8368
Average constraint energy (kcal/mol)	13.4	21.0	20.1	20.3
Violation statistics				
Maximum NOE violation (Å)	0.24	0.31	0.30	0.31
Number of NOE violations > 0.2 Å	1	6	2	3
Number of torsion angle violations > 5°	12	6	9	8
Hydrogen bonds				
Total number of H-bonds (20 structures)	4344	3838	3775	3773
Average number of H-bonds per structure	217	192	189	189

VAC: Best 20 of 100 DYANA structures refined in vacuum (VAC), energy minimized in vacuum.
GB: Best 20 of 100 DYANA structures refined with generalized Born model, energy minimized with GB model.
WATV: Refined with explicit water, using the 20 vacuum structures as initial structures, energy minimized with GB model.
WATG: Refined with explicit water, using the 20 GB structures as initial structures; energy minimized with GB model.

structure was carried out with temperature coupling constant of 1.0 at 300 K, with time-step 1 fs. The total simulation time for each structure is thus 400 ps. After the final MD cycle, water molecules were removed from the PDB file, and the protein molecules were subjected to a 2000 step energy minimization in GB mode. This minimization moved the backbone heavy atoms by less that 0.2 Å RMSD (all atoms by less than 0.3 Å RMSD); hence the structures are essentially those from the water simulation, but the energies can be compared in a straightforward fashion to the VAC and GB refinement results.

Structure analysis

Hydrogen bonds and dihedral angles were measured using MOLMOL (Koradi et al., 1996). RMSD was calculated using the in-house program SUPPOSE. PROCHECK_NMR (Laskowski et al., 1996) was used for generating the Ramachandran plots. Structural quality was also evaluated using the PROCHECK and WHATIF tests at http://biotech.embl-heidelberg.de:8400.

Results

Structure calculation

The initial structures for the AMBER refinements in vacuum and with the GB model were generated using DYANA (Güntert et al., 1997), and consisted of the 100 structures with lowest target function from a family of 200 DYANA structures. The top 20 GB

Table 2. Summary of the frequency of occurrence of HN_i-to-CO_{i-2} hydrogen bonds in the best 20 structures from 100 DYANA structures refined in vacuum (VAC), best 20 structures from 100 DYANA structures refined with generalized Born model (GB), refined with explicit water, using the 20 vacuum structures as initial structures (WATV), refined with explicit water, using the 20 GB structures as initial structures (WATG)

Hydrogen bond		Occurrence			
Donor (HN)	Acceptor (O)	VAC	GB	WATV	WATG
36	34	19	2	0	1
56	54	20	0	0	0
110	108	20	1	1	0
127	125	9	0	3	0
128	126	6	0	3	0
162	160	17	20	20	20
167	165	20	0	0	0

VAC: Best 20 structures from 100 DYANA structures refined in vacuum.
GB: Best 20 structures from 100 DYANA structures refined with generalized Born model.
WATV: Refined with explicit water, using the 20 vacuum structures as initial structures.
WATG: Refined with explicit water, using the 20 GB structures as initial structures.

structures and top 20 vacuum structures (i.e., those with lowest AMBER energy) were selected from the 50 structures with lowest constraint violation energy. The selected structures were further energy minimized and used for detailed structure comparison and analysis. Each of these sets of structures was then used for further explicit water refinements. This arrangement of the experiments was used, rather than an explicit water refinement from the DYANA structures, in order to maximize the efficient use of computer time. As will be explained later, the explicit water refinements are extremely computer-intensive, and direct refinement of the complete family of DYANA structures was therefore impractical.

Figure 1 shows the superimposed backbone conformation of four ensembles of Grx2 structures, the top 20 structures refined in vacuum (VAC) (Figure 1A), the top 20 structures refined with generalized Born model (GB) (Figure 1B), the top 20 structures refined with explicit water starting from the VAC structures (WATV) (Figure 1C), and the top 20 structures refined with explicit water starting from the GB structures (WATG) (Figure 1D). The overall folds of four sets of structures are identical. It seems that the vacuum structures are a little more compact than the GB and explicit water structures. The average absolute lengths of the principal axes for the vacuum structures are: 26.1 Å × 25.2 Å × 13.3 Å, and those for the GB structures are: 26.9 Å × 25.6 Å × 13.3 Å. The average absolute lengths of the principal axes for the explicit water structures are quite similar to those of the GB structures: 27.1 Å × 25.6 Å × 13.3 Å (WATG) and 26.9 Å × 25.7 Å × 13.3 Å (WATV).

The structure statistics for all four sets of structures are compared in Table 1. The heavy atom RMSDs for regions of regular secondary structure are very similar. The GB and explicit water structures have a significantly better distribution of ϕ and ψ angles: about 4% more of the ϕ and ψ angles are located in the most favorable region of the Ramachandran plot compared to those of the vacuum structures. The percentages of residues in the disallowed region are the same for all four sets of structures. The number of NOE violations > 0.2 Å is extremely low for all 4 sets of structures, with the value obtained for the VAC structures slightly lower than for the other three sets. Consistent with this, the average AMBER constraint violation energy for the VAC structures is smaller than that for other three sets of structures, possibly because the NMR constraints, which were identical to those used for the published structure calculation (Xia et al., 2001), had been optimized for a vacuum MD calculation, through several previous rounds of calculation. The per residue backbone heavy atom RMS fluctuations are compared in Figure 2. Most of the residues show a similar pattern for all four sets of structures.

Local geometries

The average packing environments for the VAC and GB structures were evaluated using the WHATIF 'qualcheck' and 'qual2check' analysis of directional atomic contacts (Rodriguez et al., 1998). For the 'qualcheck' test, all structures were in the normal range for highly-refined protein structures (Z-scores > -0.90); 1–4 residues per structure showed abnormal packing environments, with a minority of structures showing residues 45–47 in an abnormal environment. The 'second generation' of this test (qual2check) noted only 0–1 abnormal residues, with Z-scores greater than -0.5. There were generally no examples of bad bond-lengths, or bumpchecks, or of unlikely sidechain rotamer positions.

Hydrogen bonds

The average number of hydrogen bonds observed in the vacuum structures is significantly higher, about

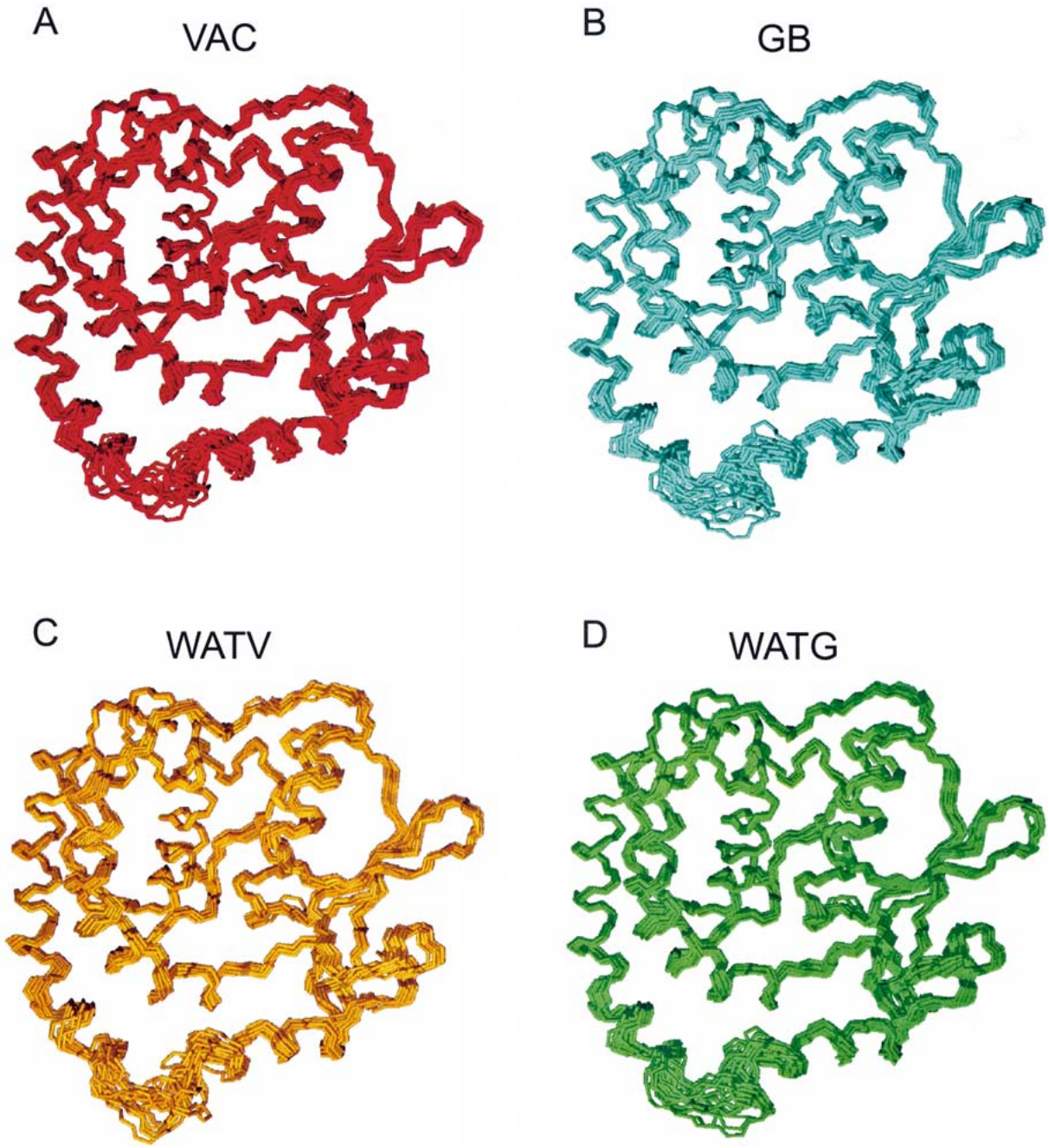

Figure 1. Superposition of ensembles of reduced Grx2 structures. (A) Best 20 structures from 100 DYANA structures refined in vacuum (VAC); (B) best 20 structures from 100 DYANA structures refined with generalized Born model (GB); (C) refined with explicit water, using the 20 vacuum structures as initial structures (WATV); (D) refined with explicit water, using the 20 GB structures as initial structures (WATG).

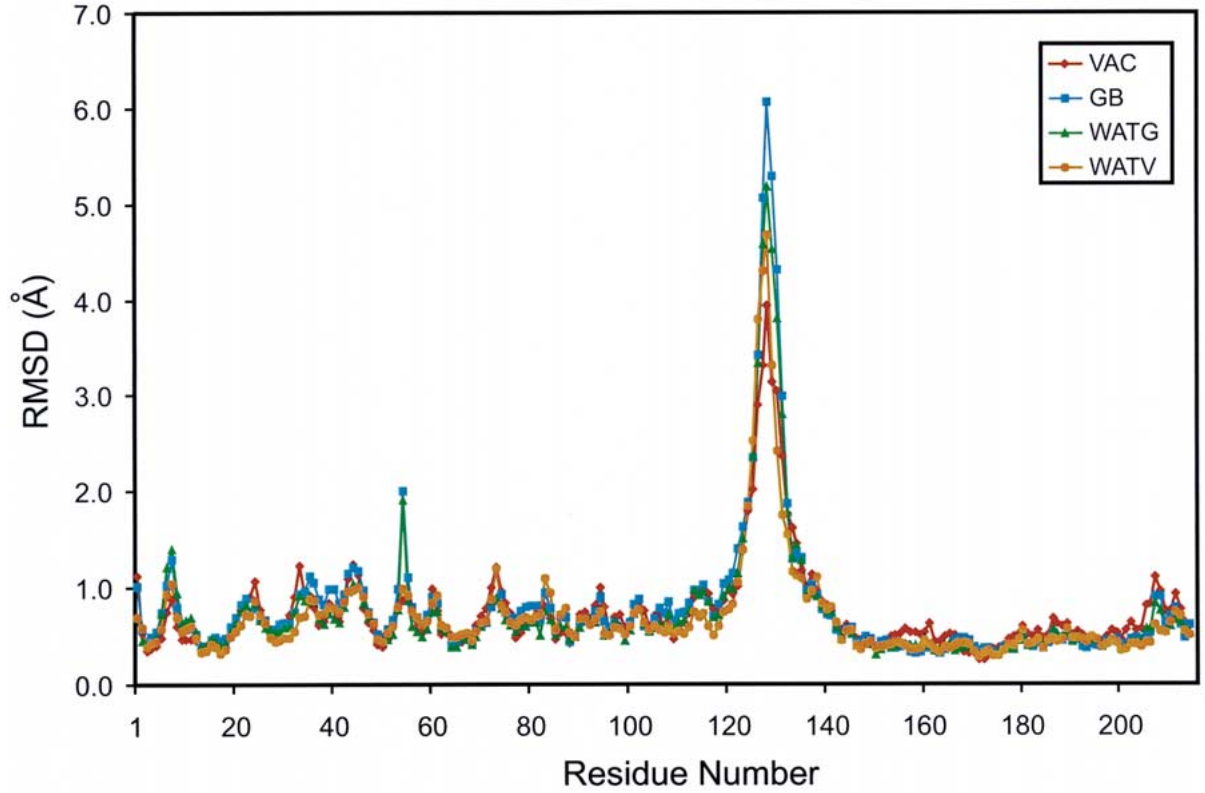

Figure 2. Per residue backbone heavy atom RMSD for the four ensembles of Grx2 structures. VAC in red, GB in blue, WATV in orange, WATG in green.

15%, than the averages for the GB structures and explicit water structures (Table 1). In general, this can be explained by the overall preference of backbone and side chain N and O atoms for hydrogen bonding in a 'vacuum' environment, even when the overall charge potential is turned down, as it was in the present calculations. This increased tendency for hydrogen bonding can give rise to the formation of incorrect hydrogen bonds in loop regions or regions that are not well defined due to lack of NMR constraints. For example, a common feature of structures refined in vacuum is a tendency to form incorrect HN_i-to-CO_{i-2} hydrogen bonds. Table 2 shows that the refinement of the structures using either the GB model or with explicit water virtually eliminates these hydrogen bonds: there are HN_i-to-CO_{i-2} hydrogen bonds observed between 7 pairs of residues in more than 5 of the vacuum structures; with the sole exception of that between K160 and N162, all of these hydrogen bonds are absent or are much less populated in the GB structures or explicit water structures.

Torsion angles

Figure 3 displays the Ramachandran plots for all four sets of structures. The distribution of ϕ and ψ angles for the vacuum structures (Figure 3A) shows a clear difference from those of the GB and explicit water structures (Figures 3B and 3D). The dihedral angles for the GB and explicit water structures are more evenly distributed around the 'most favored' regions of the Ramachandran plot (darkest gray areas in Figure 3), and there are fewer of them in the 'additionally allowed' and 'generously allowed' regions. In particular, the population of dihedral angles in the 'bridge' region between the α_R and β regions is decreased compared to those from the vacuum refinement. It is noticeable that further refinement of the vacuum structures with explicit water changed their distribution pattern to resemble the GB-refined structures (Figure 3C).

The differences between the average per-residue dihedral angles between the various structures is shown in Figure 4. The differences between the dihedral angles of the vacuum-refined structures and the

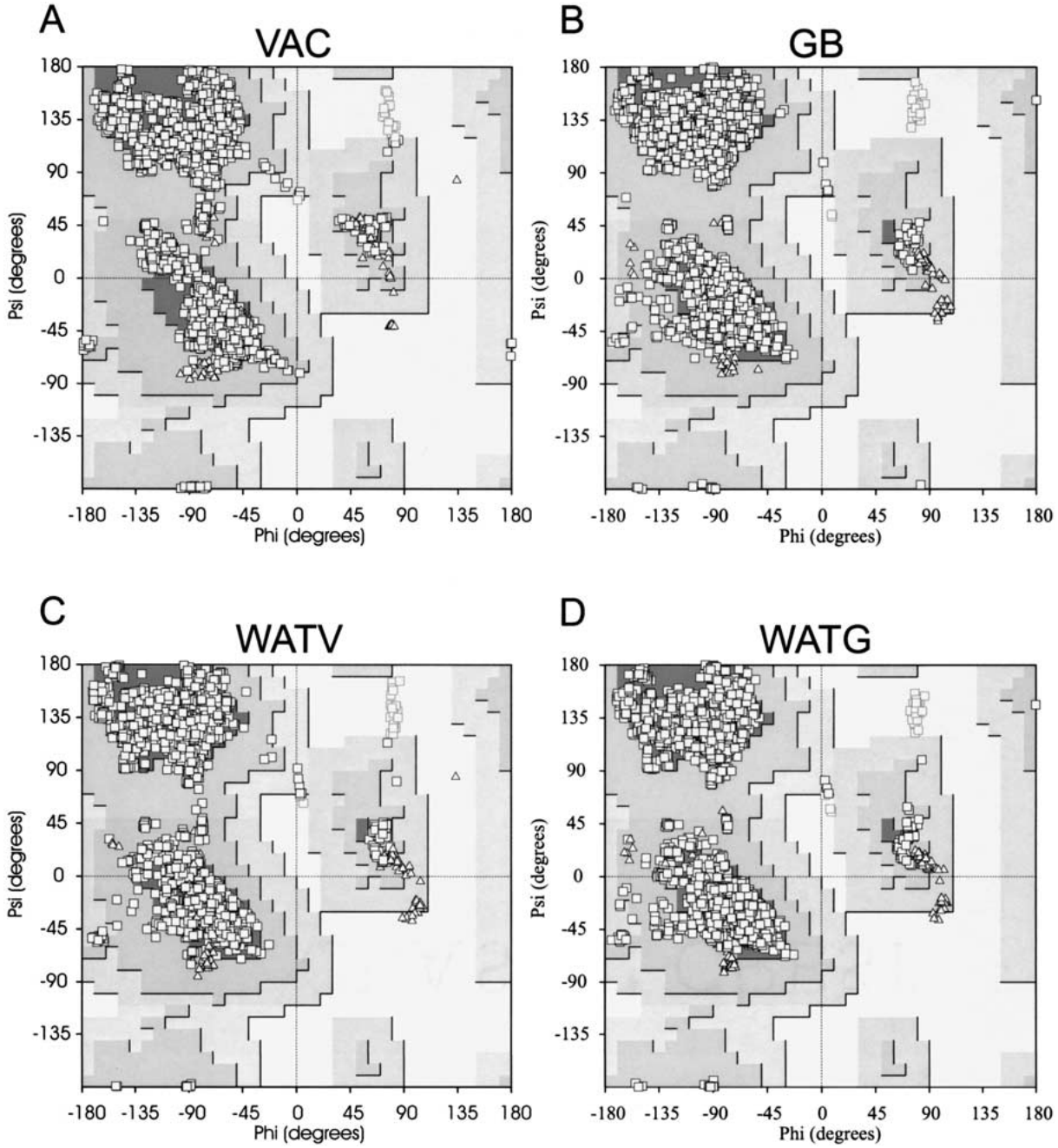

Figure 3. Ramachandran plots for the ensembles of (A) VAC, (B) GB, (C) WATV, and (D) WATG structures, obtained using PROCHECK_NMR (Laskowski et al., 1996). Darkest gray areas: Most favored, medium gray areas: Additionally allowed, light gray areas: Generously allowed, white areas: Disallowed. Squares: Non-glycine residues, triangles: Glycine residues, gray squares: Non-glycine residues in disallowed regions.

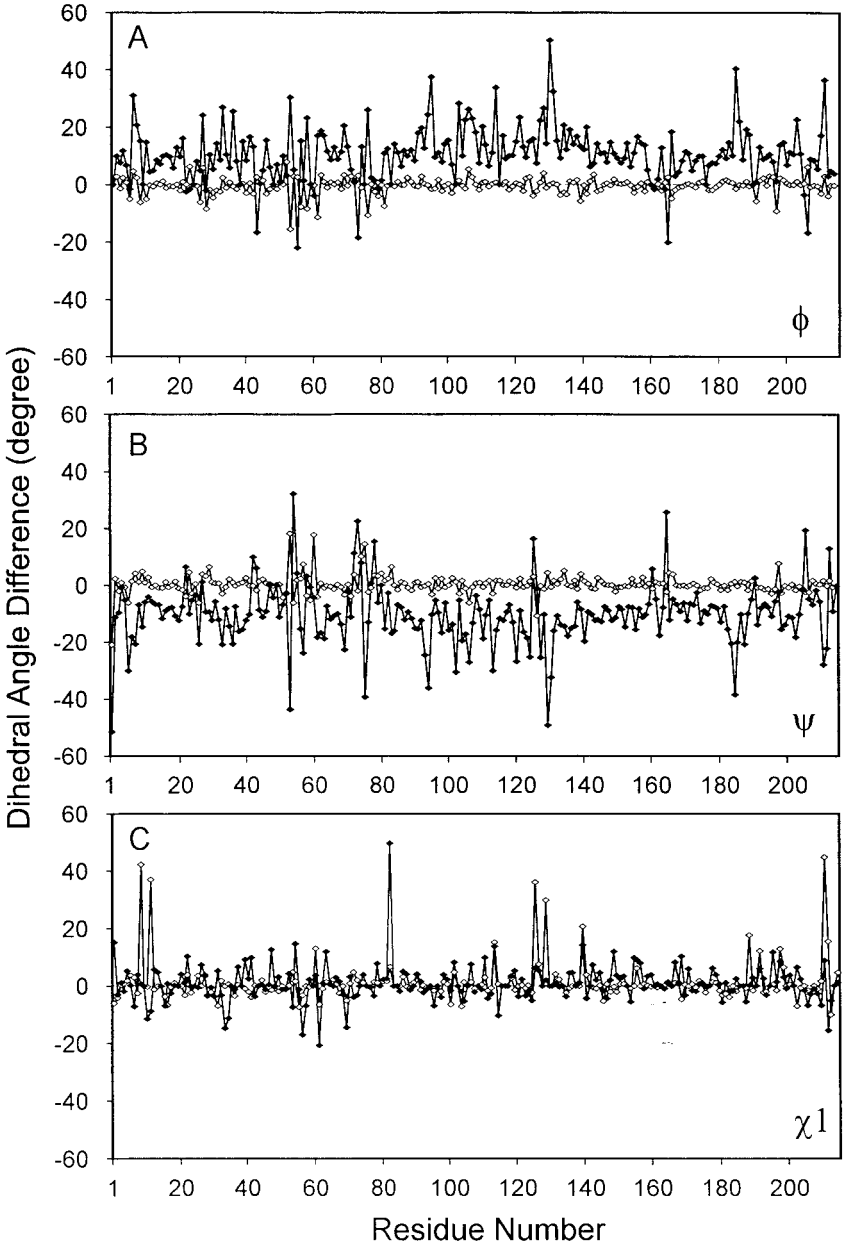

Figure 4. Differences of average dihedral angles between VAC and GB structures refined with explicit water (A) ϕ angles (B) ψ angles (C) χ_1 angles. Filled squares: VAC-WATV, open squares: GB-WATG.

water-refined vacuum structures (VAC-WATV) (solid symbols in Figure 4A and 4B) are large and generally positive for ϕ and negative for ψ. On the other hand, the equivalent calculation comparing the GB-refined structures and the GB-refined, water-refined structures (GB-WATG) (open symbols in Figures 4A and B) gives extremely small differences. This translates to a systematic difference of both ϕ and ψ, of about 13° on average, between the GB structures and the vacuum structures, whereas the average ϕ and ψ angle differences for each residue between the GB structures and any of the sets of explicit water structures are generally very small and average to zero. The reasons for these differences are not clear. It is of interest to note (Table 1) that the number of torsion angle violations > 5° is the largest for the VAC structures, while the NOE

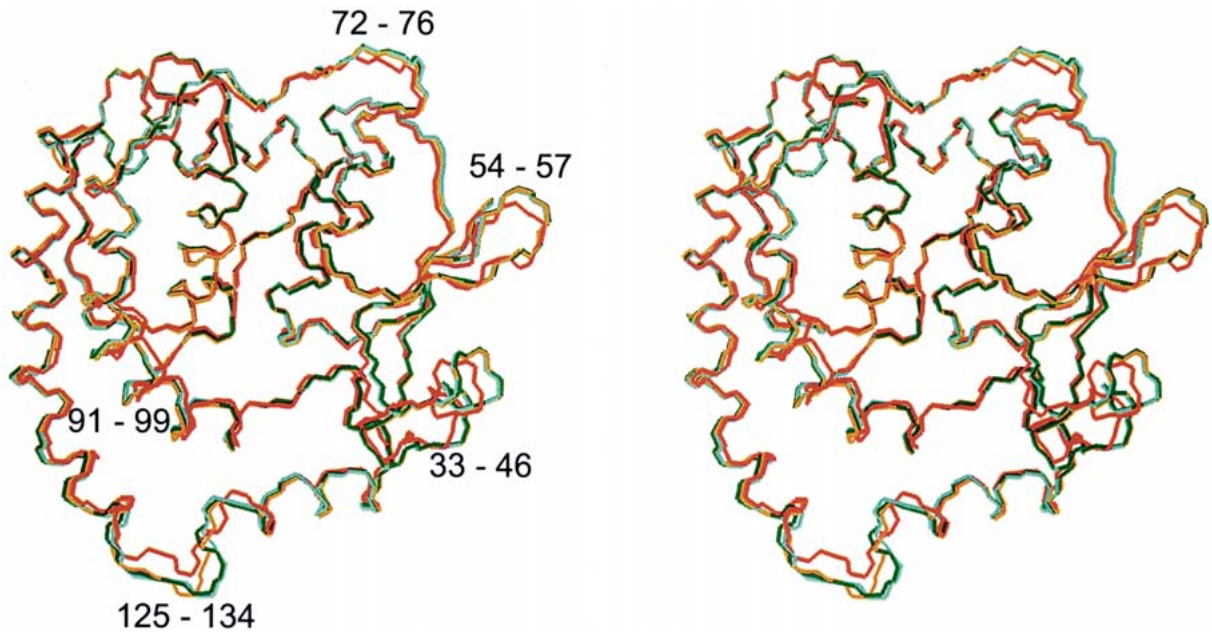

Figure 5. Stereo view of superimposition of mean structures of ensembles of VAC (red), GB (cyan), WATV (orange), and WATG (green) structures.

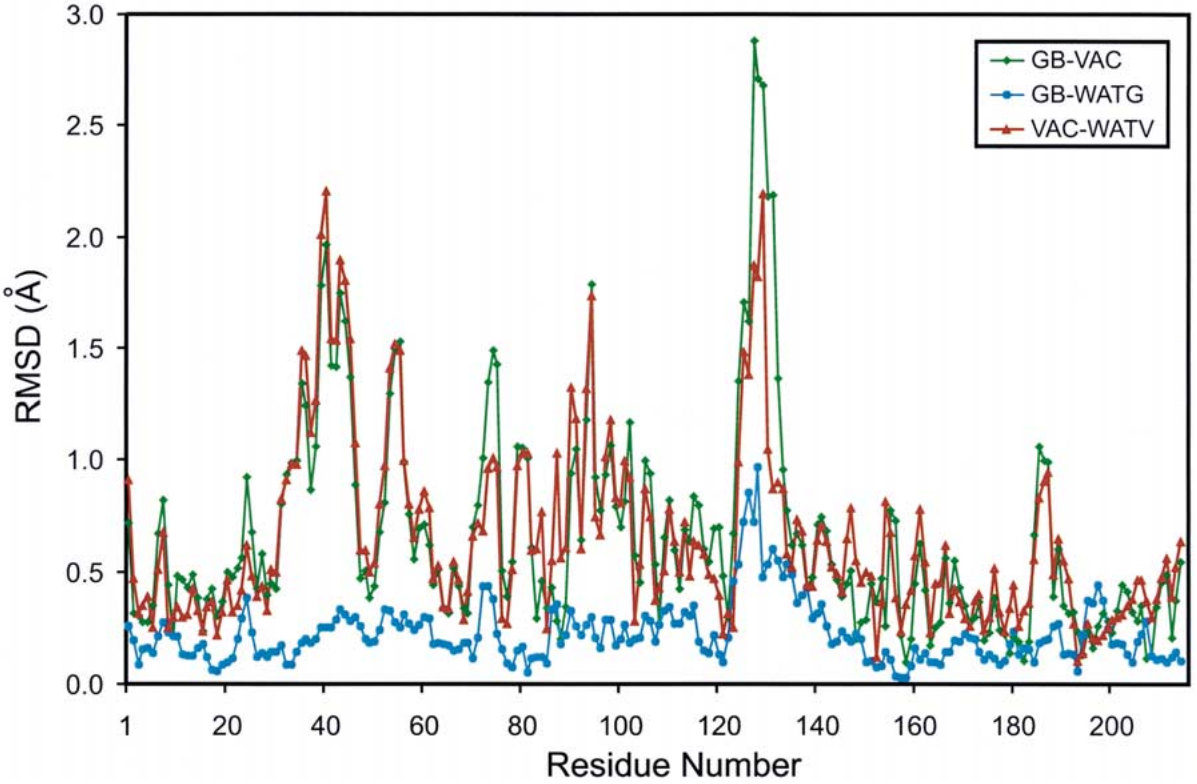

Figure 6. Pairwise per residue backbone heavy atom RMSD between mean structures of ensembles of Grx2 structures generated from various refinements. Red line: VAC/WATV, blue line: GB/WATG, green line: GB/VAC.

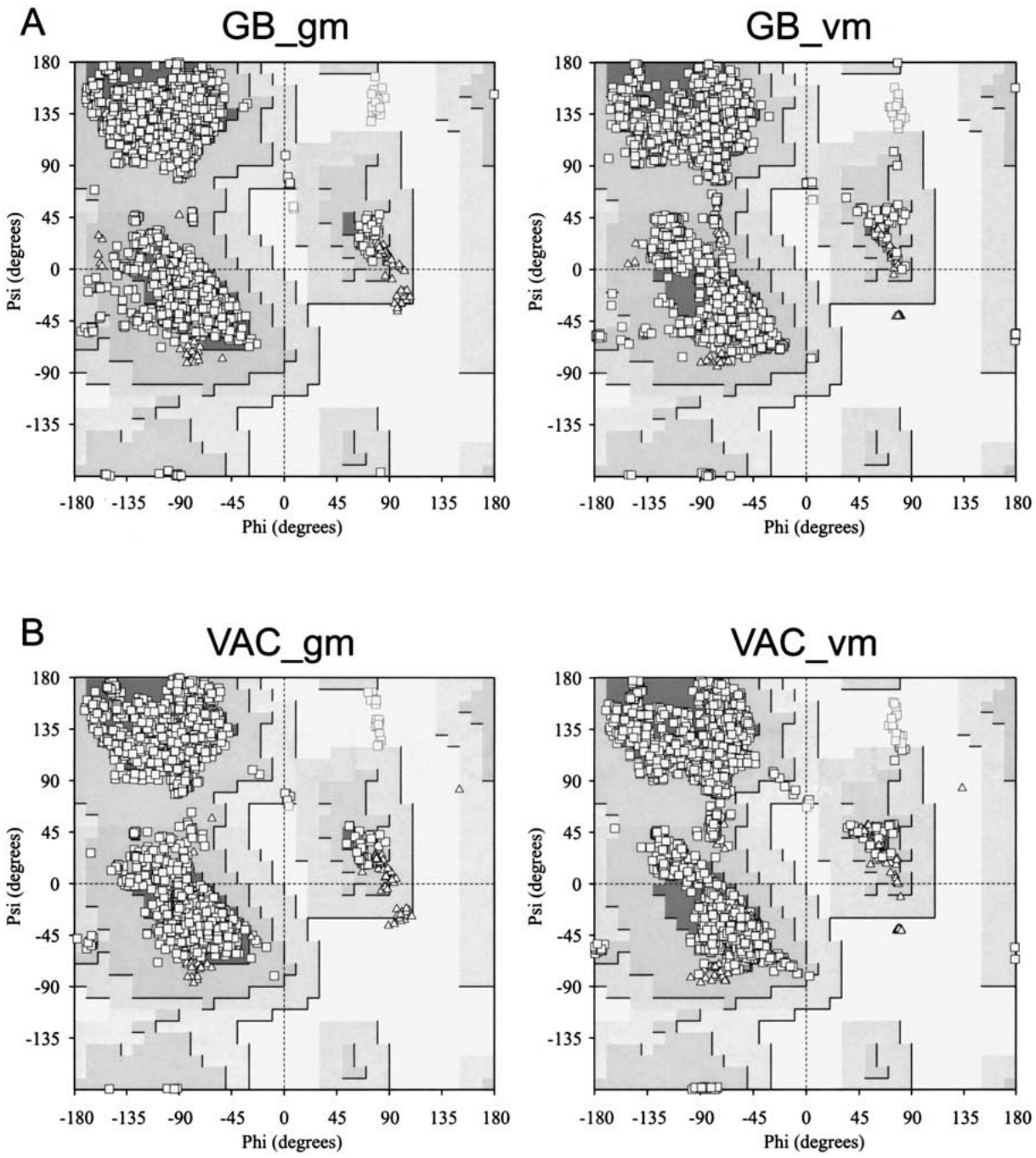

Figure 7. Ramachandran plots for the ensembles of (A) GB and (B) VAC structures energy minimized in vacuum and with GB model respectively. 'GB_gm' stands for structures refined and energy minimized with GB model, 'GB_vm' for structures refined with GB model, but energy minimized in vacuum, 'VAC_gm' for structures refined in vacuum, but energy minimized with GB model, 'VAC_vm' for structures refined and energy minimized in vacuum.

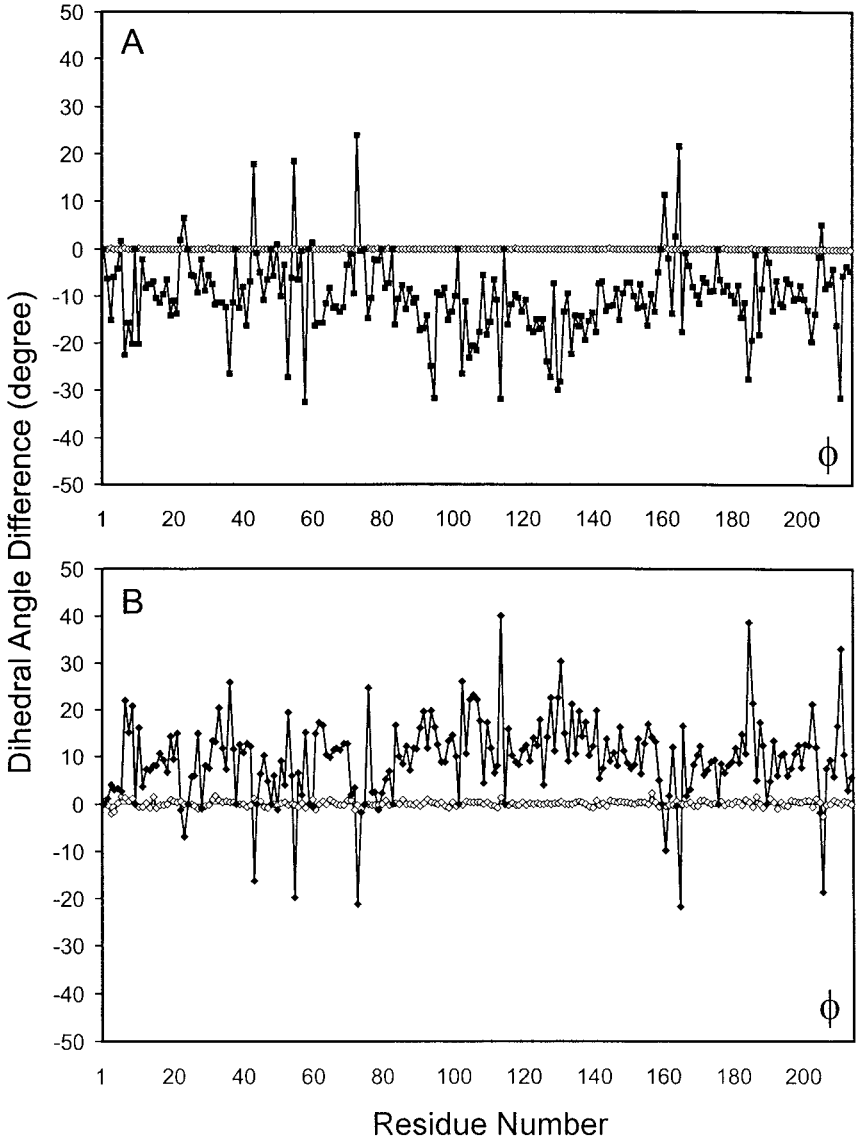

Figure 8. Difference for average ϕ angle between structures without energy minimization and structures energy minimized in vacuum and with GB model, respectively. (A) GB structures; filled squares: Differences obtained by subtraction of ϕ angle values for GB structures with minimization in vacuum from the GB values with no minimization, open squares: Differences obtained by subtraction of ϕ angle values for GB structures with minimization in GB mode from the GB values with no minimization. (B) VAC structures; filled squares: Differences obtained by subtraction of ϕ angle values for VAC structures with minimization in GB mode from the VAC values with no minimization, open squares: Differences obtained by subtraction of ϕ angle values for VAC structures with minimization in VAC mode from the VAC values with no minimization.

violations are lowest. This may be an indication that at this very fine level of resolution the force fields used to apply the NOE and dihedral angle constraints may not be quite equally balanced, giving greater weight, for example, to the NOEs in a vacuum calculation. The only difference between the conditions for the VAC refinement and for the other refinements is the reduced charge that is applied to the VAC calculation. Apart from this, it is hard to rationalize why the differences between the VAC structures and all of the others should be systematic as they are. The average χ_1 angles for all four sets of structures are similar for most of the residues (Figure 4C). The χ_1 angles for residues Cys9 and Cys12 show very large variations, probably due to the lack of NMR constraints on the sulfur atom of the cysteine side-chain.

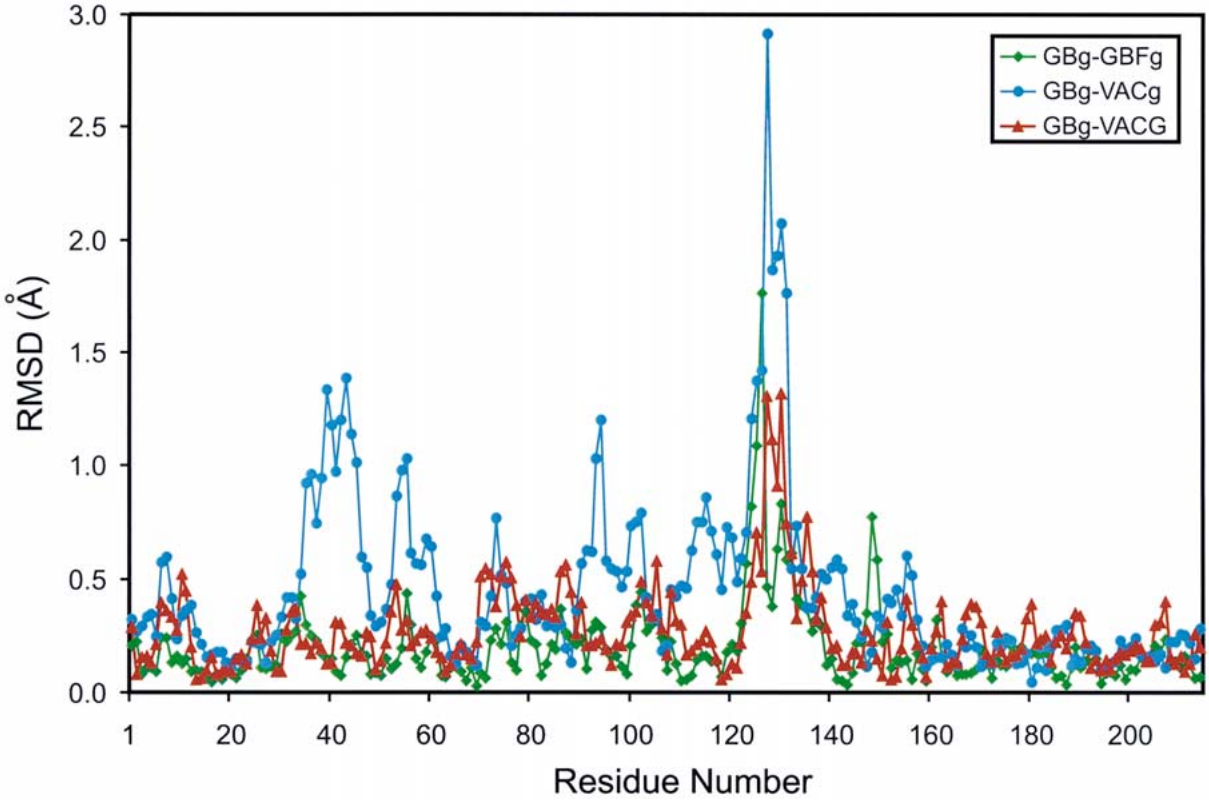

Figure 9. Pairwise per residue backbone heavy atom RMSD between GBg and GBFg mean structures, between GBg and VACg mean structures and between GBg and VACG mean structures. 'GBg' represents structures refined by MD simulation and energy minimized with the GB model; 'GBFg' represents structures refined with a faster version of the GB simulation and energy minimized with the GB model; 'VACg' refers to structures refined with MD simulation in vacuum and energy minimized in GB mode; 'VACG' refers to structures refined with MD simulation in vacuum, followed by one round of MD simulation in GB mode.

Mean structures

Figure 5 shows the four superimposed mean structures from the four sets of structures. Pairwise per residue backbone heavy atom RMSDs between mean structures of GB, VAC, WATG, and WATV are compared in Figure 6. The region for residues 125–134 is not well defined due to the lack of NMR constraints and has large RMSD in all structure sets. It is clear that the GB and WATG mean structures are very similar, with RMSD for most of the residues less than 0.5 Å, while the comparison of the VAC and WATV mean structures is generally larger and shows significant differences in several regions (33–46, 54–57, 72–76, 80–82, 91–99, 186–188). The pairwise RMSD between the VAC and GB mean structures is similar to the VAC/WATV comparison, an indication that the variant mean structure is the VAC structure. Some regions have RMSDs of more than 1 Å. These structural differences can be visualized in Figure 5.

Energy minimization

The effect of energy minimization in vacuum and with the GB model was tested on the four sets of structures. An energy minimization (4000 steps) with NMR constraints in vacuum state was applied to the GB, WATG, and WATV structures after MD simulation. An energy minimization (2000 steps) in GB mode was applied to the VAC structures. As shown in Figure 7, a GB energy minimization on the ensemble of VAC structures caused the distribution pattern of the ϕ and ψ angles in the Ramachandran plot to be more similar to that of the GB structures than the VAC structures before energy minimization. The percentage of ϕ and ψ angles in the most favored region is also improved (Table 3). On the other hand, vacuum energy minimization of the GB structures made the ϕ and ψ angle distribution pattern in the Ramachandran plot more similar to that of the VAC structures. The distribution of the backbone dihedral angles in the favored regions is also similar

Table 3. Comparison of structural statistics for the Grx2 solution structures refined with MD simulated annealing in vacuum and with the generalized Born model, with various types of energy minimization

MD simulation	VAC			GB		
Energy minimization	nm	vm	gm	nm	gm	vm
Backbone RMSD from mean structure (Å)						
All residues	0.61	0.61	0.60	0.72	0.66	0.69
Regular secondary structure	0.49	0.49	0.47	0.51	0.47	0.49
All heavy atom RMSD from mean structure (Å)						
All residues	1.02	1.02	0.91	1.06	1.00	1.04
Regular secondary structure	0.93	0.93	0.80	0.92	0.87	0.91
ϕ *and* ψ *angle distribution (%)*						
Residues in most favorable region	86.6	86.6	90.8	90.7	90.7	87.2
Residues in additional allowed region	12.1	12.2	8.2	8.4	8.4	11.5
Residues in generously allowed region	0.8	0.7	0.4	0.4	0.3	0.9
Residues in disallowed region	0.4	0.5	0.5	0.5	0.5	0.5
AMBER energy and constraint energy						
Average AMBER energy (kcal/mol)	−3139	−3278	−8307	−8386	−8375	−3212
Average constraint energy (kcal/mol)	13.9	13.4	21.0	21.0	21.0	13.4

VAC: Best 20 structures from 100 DYANA structures refined in vacuum.
GB: Best 20 structures from 100 DYANA structures refined with generalized Born model.
Each followed by energy minimization in vacuum (vm), or with the GB model (gm); or with no energy minimization (nm).

to that of the VAC structures (Table 3). Applying a GB energy minimization to the top 20 GB structures and applying a vacuum energy minimization to the top 20 VAC structures basically makes no difference to the structures. The differences for average ϕ and ψ angles observed between VAC and GB structures, can be reproduced between structures that are energy minimized in vacuum and GB mode for either the VAC or GB structures (Figure 8). Therefore, this difference is probably a reflection of the differences in the force field between the vacuum and GB modes in AMBER.

The H-bond pattern also changes with different energy minimization mode. Energy minimization of VAC structures in GB mode gives a similar H-bond pattern to that of the GB structures. *Vice versa,* energy minimization of the GB structures in vacuum mode produced an H-bond pattern similar to that of the VAC structures (data not shown). The effects of energy minimization on the explicit water structures are similar to those of the GB structures (data not shown).

Faster GB simulation

The GB MD simulation process can be sped up by changing the frequency for evaluating the GB slowly-varying forces and the frequency for updating effective radii in AMBER calculation. A faster GB simulation was performed where the effective radii were updated every two steps, and derivatives of the energy with respect to the effective GB radii were updated every four steps. Details of the procedure are described elsewhere (Tsui and Case, 2001). This protocol has been shown to conserve energy nearly as well as the standard scheme in which all terms in the potential and forces are updated at each step. The Ramachandran plot for the top 20 structures from faster GB refinement (GBF) is very similar to that of the GB structures (data not shown). The structural statistics for the 20 GBF ensemble are compared with those of the 20 GB structures in Table 4, and they are almost identical. The averaged ϕ and ψ angle values and the hydrogen bond pattern are also very similar between the GBF structures and the GB structures.

The per residue backbone heavy atom RMSDs between the GBF mean structure and the GB mean structure are generally very small, less than 0.4 Å for most of the residues (Figure 9). Although energy minimization of the VAC structures in GB mode can produce structures with similar ϕ and ψ angle distribution pattern and similar hydrogen bonding pattern to the GB structures, their mean structure and the GB mean structure have relative larger per residue RMSD com-

Table 4. Summary of structural statistics for the GBF and GB structures

GB protocol	GBF	GB
Backbone heavy atom RMSD from mean structure (Å)		
All residues	0.67	0.72
Regular secondary structure	0.49	0.50
All heavy atom RMSD from mean structure (Å)		
All Residues	1.00	1.06
Regular secondary structure	0.87	0.92
ϕ and ψ angle distribution (%)		
Residues in most favorable region	90.7	90.7
Residues in additional allowed region	8.2	8.4
Residues in generously allowed region	0.5	0.3
Residues in disallowed region	0.5	0.5
AMBER energy and constraint energy		
Average AMBER energy (kcal/mol)	−8376	−8375
Average constraint energy (kcal/mol)	20.5	21.0

pared with the RMSD between the mean structures of GBF structure and GB structures (Figure 9).

Computation time

Timing comparisons between the three sets of structures described here reflect not only the intrinsic speed of simulation using different potentials, but also the different protocols that were used. In particular, the VAC and GB simulated annealing refinements are based on 60 ps of simulation (two annealing runs of 30 ps each), whereas the WAT simulations were run for 400 ps of simulation. This difference is appropriate, since conformational sampling is slower in explicit water than in either VAC or GB, mainly due to the viscous damping forces of the water. It may well be possible to carry out shorter explicit water refinements, which would save computation time. Comparisons between the VAC and GB simulations also reflect the use of a much longer non-bonded cutoff (20 Å) for the GB runs, compared to 8 Å for the VAC refinements. Nevertheless, some useful information can be gleaned from comparisons of computing times.

It took about 200 h of CPU time per structure for the explicit water refinement, about 90 h for the GB refinement, and about 2 hours for the vacuum refinement. The time for the GB refinement would be reduced by about a factor of 3 if an 8 Å cutoff were used instead of 20 Å. The 'faster' GB simulation outlined above reduces the timings by another factor of 2, so that each structure would require about 15 h of CPU time; this is about 8 times that required for the VAC simulation.

Applying energy minimization with the GB model to the top 20 VAC structures can improve the distribution of ϕ and ψ angles and eliminate the differences observed between VAC and GB structures for the average ϕ and ψ angles and hydrogen bond pattern. This takes only modest additional computational time, but has the potential drawback that the mean structure has a larger RMSD to the mean of the GB structures (Figure 9).

Yet another alternative method is to apply one cycle of MD simulation with the GB model to the top 20 VAC structures. This generally produces structures with good agreement to the GB structures (Figure 9).

Conclusions

The structures calculated by using the generalized Born method are very similar to those obtained using the much more computationally expensive explicit water method. The GB method provides a good model for solvent effects in the refinement of macromolecule solution structures. Compared to the MD refinement in vacuum, MD refinement with the GB model can eliminate incorrect HN_i-to-CO_{i-2} hydrogen bonds in the vacuum structure and improve the ϕ and ψ angle distribution pattern. We also found that the problems of incorrect hydrogen bonds and non-optimal ϕ and ψ angle distribution in the VAC structures can be solved by performing energy minimization in GB mode after the vacuum MD simulation. This decreases the computation time significantly. Since the regular GB MD refinement strategy is still quite computationally expensive, we have tested different combinations of MD simulation refinement and energy minimization using GB model, and the computation costs for them are significantly lower. They generally produce structures in good agreement with GB structures. The method to be used for refinement of a particular solution structure can therefore be chosen based on available computational power and the desired local accuracy.

Acknowledgements

We thank Linda Tennant for technical assistance, John Chung and Gerard Kroon for NMR maintenance and members of the Wright/Dyson group for helpful discussions. This work was supported by grants GM43238 (HJD), GM36643 (PEW) and GM45811 (DAC) from the National Institutes of Health.

References

Bashford, D. and Case, D.A. (2000) *Annu. Rev. Phys. Chem.*, **51**, 129–152.

Calimet, N., Schaefer, M. and Simonson, T. (2001) *Proteins*, **45**, 144–158.

Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, III, T.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1999). *AMBER 6*, University of California, San Francisco, CA.

Cheatham, III, T.E. and Kollman, P.A. (1996) *J. Mol. Biol.*, **259**, 434–444.

Cornell, W., Abseher, R., Nilges, M. and Case, D.A. (2001) *J. Mol. Graph. Model.*, **19**, 136–145.

Cramer, C.J. and Truhlar, D.G. (1999) *Chem. Rev.*, **99**, 2161–2200.

Dauter, Z., Lamzin, V.S. and Wilson, K.S. (1997) *Curr. Opin. Struct. Biol.*, **7**, 681–688.

Dominy, B.N. and Brooks, III, C.L. (1999) *J. Phys. Chem.*, **103**, 3765–3773.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graphics*, **14**, 51–55.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Nicholls, P. (2000) *Cell Mol. Life Sci.*, **57**, 987–992.

Onufriev, A., Bashford, D. and Case, D.A. (2000) *J. Phys. Chem.*, **B104**, 3712–3720.

Otting, G. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 1871–1875.

Qiu, D., Shenkin, P.S., Hollinger, F.P. and Still, W.C. (1997) *J. Phys. Chem. A*, **101**, 3005–3014.

Rapp, C.S. and Friesner, R.A. (1999) *Proteins*, **35**, 173–183.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T. and Vriend, G. (1998) *Bioinformatics.* **14**, 523–528.

Schaefer, M., Bartels, C. and Karplus, M. (1998) *J. Mol. Biol.*, **284**, 835–848.

Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T. (1990) *J. Am. Chem. Soc.*, **112**, 6127–6129.

Tsui, V. and Case, D.A. (2000) *J. Am. Chem. Soc.*, **122**, 2489–2498.

Tsui, V. and Case, D.A. (2001) *Biopolymers (Nucl. Acid Sci.)*, **56**, 275–291.

Williams, D.J. and Hall, K.B. (1999) *Biophys. J.*, **76**, 3192–3205.

Xia, B., Vlamis-Gardikas, A., Holmgren, A., Wright, P.E. and Dyson, H.J. (2001) *J. Mol. Biol.*, **310**, 907–918.