Atomic-level accuracy in simulations of large protein crystals

DARRIN M. YORK*†, ALEXANDER WLODAWER‡, LEE G. PEDERSEN§, AND TOM A. DARDEN¶‖

*Department of Chemistry, Duke University, Durham, NC 27706; †North Carolina Supercomputing Center, Microelectronics Center of North Carolina,
Research Triangle Park, NC 27709-12889; ‡Macromolecular Structure Laboratory, National Cancer Institute, Frederick, MD 21702-1201;
§Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290; and ¶National Institute
of Environmental Health Sciences, Research Triangle Park, NC 27709

ABSTRACT   Proper treatment of long-range Coulombic
forces presents a major obstacle to providing realistic molec-
ular dynamics simulations of macromolecules. Traditional ap-
proximations made to lessen computational cost ultimately lead
to unrealistic behavior. The particle mesh Ewald method
accommodates long-range Coulombic forces accurately and
efficiently by use of fast Fourier transform techniques. We
report a 1-ns simulation of bovine pancreatic trypsin inhibitor
in a crystal unit cell using the particle mesh Ewald methodol-
ogy. We find an rms backbone deviation from the x-ray
structure (0.33 Å) that is lower than that observed between
bovine pancreatic trypsin inhibitor in different crystal forms
and much lower than those of previous simulations. These
results bridge the gap between structures obtained from mo-
lecular simulation and those from experiment.

Molecular simulations of macromolecules frequently pro-
duce models with considerable deviation (usually exceeding
1 Å) from the experimentally determined crystal structure. A
major obstacle to accurate simulations of macromolecular
systems concerns the treatment of long-range ionic forces.
Often these forces are approximated by truncated or modified
Coulombic potentials that ignore interactions beyond a fixed
cutoff. Use of cutoff methods, however, can lead to signifi-
cantly reduced accuracy and artificial behavior (1–3).

The particle mesh Ewald method (4) employs charge
gridding and fast Fourier transforms to compute the Ewald
sum in order $N \log N$ steps, allowing its use in macromolec-
ular simulations. The method has been shown to be a
significant improvement over single or "twin-range" cutoff
methods in crystal simulations of large proteins (1). A fun-
damental question arises as to the reliability of molecular
simulation when rigorous treatment of long-range forces is
taken into account. To explore this question, it is instructive
to compare molecular simulation results of well-studied sys-
tems directly to accurate experimental data.

Bovine pancreatic trypsin inhibitor (BPTI) has served for
more than a decade as a benchmark system for testing
theoretical simulation techniques (5). High-resolution x-ray
(6, 7) and x-ray/neutron diffraction data (8) are available for
BPTI in several crystalline forms, in addition to a recent
solution structure determined by two-dimensional NMR (9).
Herein we report a 1-ns molecular dynamics (MD) simulation
of BPTI in a crystalline environment (form 1) using full Ewald
sums evaluated with the particle mesh Ewald method. The
protein backbone of the simulation average structure has a
rms deviation of 0.33 Å from the crystallographic structure.
This value is less than the deviations observed between
different crystalline forms of BPTI (7, 10). Hence, this work
represents a major advance in the atomic-level accuracy
attainable in simulations of biological macromolecules.

METHODS

Molecular mechanics and MD calculations were performed
using a modified version of the AMBER software package (11),
with Ewald sums implemented using the particle mesh Ewald
method as described (1, 4). The all-atom force field (12) was
employed for protein molecules in conjunction with explicit
TIP3P water (13) and chloride counterions (14). The starting
configuration of the asymmetric unit containing one protein
monomer and 60 water molecules was obtained from the 4PTI
structure in the Brookhaven Protein Data Bank (6). The unit
cell containing four asymmetric units was generated by space
group operations. After this, additional waters were placed,
bringing the total to 576 solvent molecules [corresponding to
36% solvent by volume (10)] and 24 of these extra waters
were replaced by chloride counterions, yielding a neutral unit
cell. Preparation and equilibration of the unit cell was anal-
ogous to previous crystal simulations (1, 15). Covalent bonds
involving hydrogen were constrained using a modified ver-
sion of the SHAKE algorithm (16). Simulations were performed
in the NVT ensemble (298 K) with a 1-fs integration time step
and carried out to 1 ns saving coordinates every 0.5 ps. The
1000-ps simulation required $\approx$370 h of Cray-YMP central
processing unit time (single processor).

RESULTS AND DISCUSSION

Simulated vs. Experimental Structures. Deviation of the
simulation structures from the crystallographic structure can
be monitored by the rms positional deviation (rmsPD) as a
function of time (Fig. 1). The monomer rmsPD values reach
equilibrium by 250 ps and remain stable for the duration of the
1-ns simulation with asymptotic values of 0.52 Å (backbone
atoms) and 0.63 Å (heavy atoms with crystallographic B
values < 20 Å^2). These values are considerably lower than
those of previous MD crystal simulations of BPTI (17–19) and
other large proteins (2, 15, 20). To assess the effect of the
long-range electrostatics vs. that of the crystalline environ-
ment on these results, a control crystal simulation using a 9-Å
residue-based nonbond cutoff was run. The results were
similar to those of previous simulations, with an average
monomer rmsPD of 1.8 Å after 250 ps.

Fig. 2 shows a superposition of the simulation cell-average
structure and the crystallographic structure. The rmsPD for
the fit is 0.33 Å for backbone atoms and 0.63 Å for all heavy
(nonhydrogen) atoms. This value is less than the deviations
observed between different crystalline forms of BPTI (7, 10)
(Table 1). The largest C_α deviations occur at residues 1 (0.93
Å), 25 (0.64 Å), 36 (0.63 Å), and 40 (0.63 Å). The latter three
residues appear in flexible loops or turns. In the simulation,
the protein termini were charged (pH 7); however, x-ray
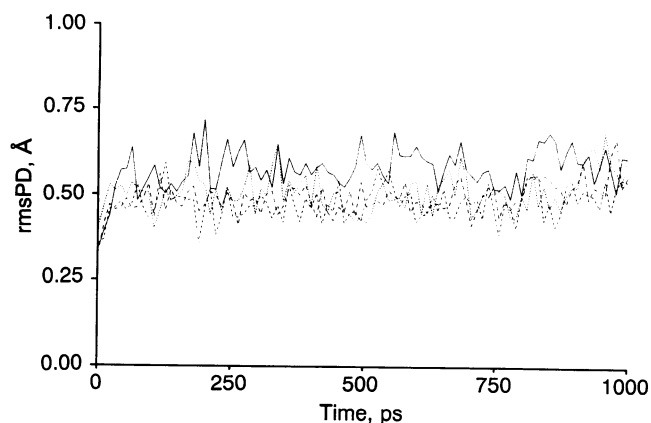diffraction data were collected at pH 9.5 (6). Since the

FIG. 1.   Time evolution of the backbone (N–C_α–C) rmsPD of the four simulation monomers from the BPTI form I crystallographic structure (6). The average rmsPD (250–1000 ps) for backbone atoms, heavy atoms, and heavy atoms with crystallographic temperature factors <20 Å² are 0.52 Å, 0.85 Å, and 0.63 Å, respectively.

measured pK_a of the BPTI N terminus is 8.2 (21), it is probable that in the crystallographic structure this group is neutral. One might therefore expect a difference between the simulation and crystallographic structures in this region.

Atomic Fluctuations.   Atomic fluctuations can be estimated from the crystallographic temperature factors (B values) through the relation $\langle \Delta r_i^2 \rangle^{1/2} = (3B_i/8\pi^2)^{1/2}$, where $\langle \Delta r_i^2 \rangle^{1/2}$ is the

Table 1.   Comparison matrix of the rmsPD of the simulated cell-average structure (MD), crystallographic structures of BPTI (I–III), and the NMR solution structure

	Atoms, Å				
	MD	I	II	III	NMR
MD	—	0.33	0.52	0.58	0.82
I	0.63	—	0.39	0.41	0.82
II	1.27	1.10	—	0.39	0.76
III	1.32	1.21	0.95	—	0.85
NMR	1.43	1.36	1.28	1.31	—

Experimentally determined structures of native BPTI include crystallographic structures of form I (6), form II (8), and form III (7) crystals (these three have different crystaline environments) and an average solution structure determined by two-dimensional NMR (9). The rmsPD values were determined using backbone atoms (above the diagonal) and heavy atoms (below the diagonal) in the rms fit. Atoms of residues 1–56 were used in all fits (as in refs. 7 and 10) except those involving the average NMR structure for which residues 2–56 were used (as in ref. 9).

rms positional fluctuation (rmsPF) of atom i, and B_i is the corresponding B value. Fig. 3 compares the simulated atomic fluctuations with those estimated from the crystallographic data. The rmsPF values calculated from time variance have significant correlation with the experimentally derived values (correlation coefficient 0.62/0.69 for main-chain/side-chain atoms, respectively); however, the former are smaller in magnitude. The experimentally derived fluctuations may be
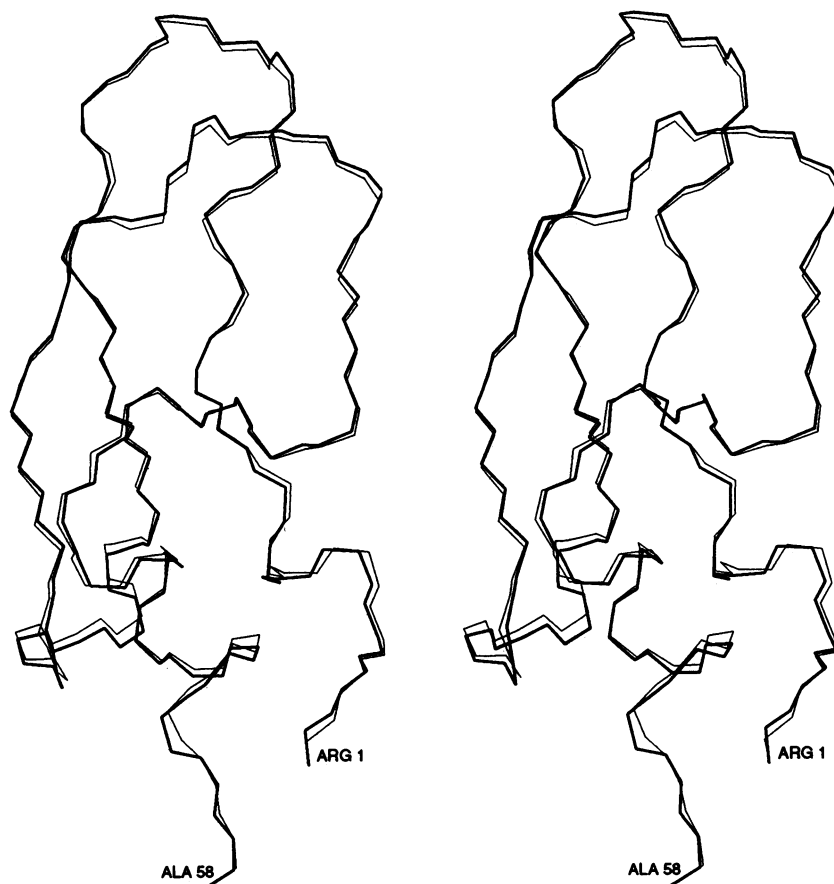

FIG. 2.   Stereo superposition of the backbone atoms of the crystallographic structure (thick lines) and the simulated cell-average structure (thin lines). The overall rms for the fit was 0.33 Å for backbone (N–C_α–C) atoms and 0.63 Å for heavy (nonhydrogen) atoms. The corresponding rmsPD values of the four monomer time-average structures were slightly greater (0.38 ± 0.05 Å, backbone; 0.69 ± 0.05 Å, heavy). The instantaneous cell-average structures were obtained by applying the reverse $P2_12_12_1$ symmetry operations to the protein monomers and averaging them, at each time point in the interval from 250 to 1000 ps. The cell-average structure is the time average of these instantaneous structures. Unlike fitting procedures that use the crystallographic structure *a priori*, this method requires only knowledge about the unit-cell symmetry and dimensions. Alternately, the monomer time-average structures were obtained from simple time averaging (250–1000 ps).

Biophysics: York *et al.*
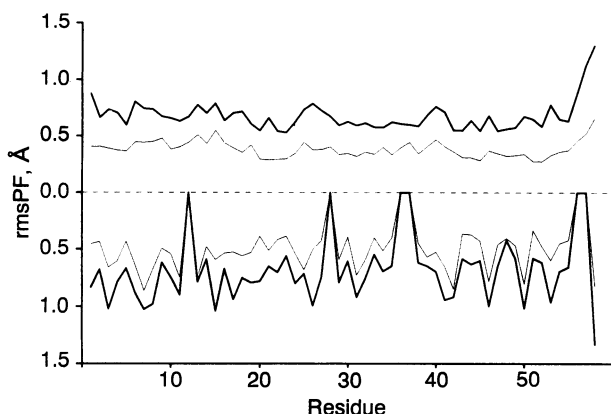
Proc. Natl. Acad. Sci. USA 91 (1994)   8717

FIG. 3. Comparison of atomic fluctuations (rmsPF) calculated from the average monomer time variances (250–1000 ps) (thin lines) and estimated from the crystallographic B values (thick lines). Residue averages for main-chain (N–C_α–CO) and side-chain heavy atoms are shown above and below the x axis, respectively. The average values for main-chain/side-chain atoms of the simulated and experimentally derived fluctuations are (0.39 Å/0.55 Å) and (0.68 Å/0.77 Å), respectively. The corresponding average values calculated from the cell variance are 0.55 Å/0.72 Å. The time variance is the average of the four monomeric time-variance results, each involving 1500 coordinate sets (two frames per ps). The cell variance is the total variance of the 6000 symmetry transformed coordinate sets from the cell-average structure described in Fig. 2.

considered an upper bound to fluctuations calculated from simulation because additional contributions to the thermal parameters such as lattice disorder are present (22). Atomic fluctuations calculated from the cell variance (time and inter-monomer variance) are similar in magnitude to the experimentally derived fluctuations and are still bounded by them. This result is in striking contrast with other crystal simulations in which fluctuations calculated from the cell variance severely overestimate the atomic fluctuations (15, 18–20). The artificially large fluctuations in these simulations result from large rmsPD between the individual time-average monomers, probably resulting from use of an electrostatic cutoff (18).

Secondary Structure. Analysis of secondary structure gives an overall perspective of the nature of intramolecular hydrogen bonding in the protein. The secondary structural assignments (23) for the cell-average structure are identical to those of the crystallographic structure. The length of the current simulation permits detailed study of the fluctuations in secondary structure in the nanosecond time domain. Table 2 shows the dynamic secondary structure definitions from the simulation. The antiparallel β-sheet (residues 18–35) shows breathing near the ends where the β structure transiently melts and reforms. In addition, the stable 3_{10}-helix (residues 3–6) shows transient α-helical character $\approx 3\%$ of the time. Conversely, the α-helix region (residues 48–55) remains invariant.

Intermolecular Hydrogen Bonds. Intermolecular hydrogen bonds stabilize the protein molecules in the crystal lattice. The crystal packing environment in the simulation closely resembles that of the crystallographic structure (Table 3). Only three hydrogen bond contacts in the cell-average structure differ by >1 Å from the x-ray structure in the donor–acceptor heavy atom distance, and these involve side-chain atoms with high thermal factors (residues 17, 26, and 42). An extensive discussion of intermolecular hydrogen bonding in BPTI has been given elsewhere (7, 10).

Aromatic Hydrogen Bonds. The high-resolution x-ray crystal structure of BPTI (8) revealed an unusual interaction between hydrogen-bond donors (backbone NH of Gly[37] and side chain NH$_2$ of Asn[44]) and the π-cloud of a buried aromatic side chain (Tyr[35]). These interactions are critical to maintaining the

Residue	X-ray*	B	E	G	H	S	T
3	G			72.2	2.7	24.5	0.6
4	G			95.2	2.7		2.2
5	G			95.2	2.7		2.2
6	G			81.5	2.7	9.1	6.7
7	S					99.6	
13	S					100	
18	E	6.8	91.8				
19	E		92.0				
20	E		93.1				
21	E		100				
22	E		99.9				
23	E		86.8				
24	E		73.0				
25	T			0.8			99.2
26	T			0.8			99.2
27	T			0.8			99.2
28	T					21.6	73.1
29	E		73.0		27.0		
30	E		86.8				
31	E		99.9				
32	E		100				
33	E		93.1				
34	E		92.0				
35	E		91.8				
37	S		0.7			99.3	
38	S					100	
39	S					80.5	
42	S					100	
43	S					98.7	
45	B	100					
46	S					100	
47	S					92.1	
48	H				100		
49	H				100		
50	H				100		
51	H				100		
52	H				100		
53	H				100		
54	H				100		
55	H				97.8	0.5	1.7
56	S				0.1	86.8	12.1

Table 2. Dynamical secondary structure assignments (percentages) from the simulation (250–1000 ps)

Secondary structural definitions were given by Kabsch and Sander (23). B, β-bridge; E, extended strand, participates in β-ladder; G, 3_{10}-helix; H, α-helix; S, bend; T, hydrogen-bonded turn. Only residues having assigned secondary structure >15% of the time are shown. A detailed account of secondary structure and intramolecular hydrogen bonding can be found elsewhere (9, 10).
*Secondary structural assignments for the cell-average and crystallographic structures were identical.

native structure since the crystal structure of the Y35G mutant is very different in this region (24). Similar aromatic hydrogen bonds are the subject of intense experimental and theoretical studies (25–27). The donor fragments are located above and below the plane of the tyrosine ring with the donor hydrogens ≈ 3.7 Å from the ring carbons. Since these amino/aromatic hydrogen bonds are thought to provide stabilization of 1–2 kcal/mol (1 cal = 4.184 J) (27), it was of interest whether these delicate interactions were maintained in the simulation. In fact, the integrity of both donor–π interactions is well preserved, the absolute deviation of the 12 donor-nitrogen–π-carbon distances between the crystallographic and cell-average structures being only 0.11 Å.

CONCLUSION

Rigorous modeling of long-range Coulombic forces presents a major obstacle to providing realistic molecular simulations.

Table 3. Intermolecular hydrogen bonds in BPTI: A comparison of crystal form I (6) and the simulation cell-average structure

Donor	Acceptor		Distance, Å			
			X–Y		Y–H	
			X-ray	MD	X-ray	MD
Asp[3] NH	H	Glu[49] OE1	3.07	3.38	2.12	2.44
Arg[17] NE	HE	Lys[26] O	2.97	2.76	2.05	1.79
Arg[17] NH-2	HH22	Ala[58] O	2.95	3.32	1.99	2.33
Arg[17] NE	HE	Ala[27] O	3.72	3.19	1.97	1.79
Arg[17] NH-1	HH12	Ala[58] O1	3.58	2.76	2.82	1.82
Arg[17] NH-1	HH12	Ala[58] O2	4.48	3.13	4.10	2.34
Arg[17] NH-2	H22	Ala[58] O2	2.67	2.73	1.90	1.79
Lys[26] NZ	HNZ2	Gly[57] O	4.91	2.72	4.37	2.29
Arg[39] NE	HE	Glu[49] OE-1	3.04	2.69	2.07	1.72
Arg[39] NE	HE	Glu[49] OE-2	3.32	3.54	2.48	2.86
Arg[39] NH-1	HH12	Try[21] OH	2.70	3.30	1.78	2.70
Arg[39] NH-2	HH21	Glu[49] OE-2	2.80	2.69	1.86	1.71
Arg[42] NH-1	HH11	Tyr[10] OH	3.26	3.88	2.31	3.01
Arg[42] NH-1	HH12	Arg[39] O	3.28	4.70	2.31	4.07
Ala[48] N	H	Arg[39] NH-2	3.34	3.67	2.45	2.81

X–H–Y–H-bond interaction, where X is the heavy atom of the donor, Y is the acceptor, and H is hydrogen. An extensive discussion of intermolecular hydrogen bonding for several crystal forms of BPTI can be found elsewhere (7, 10).

The results of our 1-ns simulation of BPTI in a unit cell using Ewald sums with the standard AMBER force field (11, 12) are in excellent agreement with the experimental crystal structure (0.33 Å backbone rms). These results represent a major advance in the atomic-level accuracy attainable in macromolecular simulations and underscore the importance of proper treatment of long-range forces. In the development of future generation force fields, accurate modeling of electrostatic interactions will undoubtedly play an essential role in bridging the gap between theory and experiment.

1. York, D., Darden, T. & Pedersen, L. (1993) *J. Chem. Phys.* **99**, 8345–8348.

2. Kitson, D. H., Avbelj, F., Moult, J., Nguyen, D. T., Mertz, J. E., Hadzi, D. & Hagler, A. T. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8920–8924.
3. Schreiber, H. & Steinhauser, O. (1992) *Biochemistry* **31**, 5856–5860.
4. Darden, T., York, D. & Pedersen, L. (1993) *J. Chem. Phys.* **98**, 10089–10092.
5. Daggett, V. & Levitt, M. (1993) *Annu. Rev. Biophys. Biomol. Struct.* **22**, 353–380.
6. Deisenhofer, J. & Steigemann, W. (1975) *Acta Crystallogr. B* **31**, 238–250.
7. Wlodawer, A., Nachman, J., Gilliland, G., Gallagher, W. & Woodward, C. (1987) *J. Mol. Biol.* **198**, 469–480.
8. Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984) *J. Mol. Biol.* **180**, 301–329.
9. Berndt, K., Güntert, P., Orbon, L. & Wüthrich, K. (1992) *J. Mol. Biol.* **227**, 757–775.
10. Wlodawer, A., Deisenhofer, J. & Huber, R. (1987) *J. Mol. Biol.* **193**, 145–156.
11. Weiner, S., Kollman, P., Case, D., Singh, U., Chio, C., Algona, G., Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
12. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comput. Chem.* **7**, 230–252.
13. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. (1983) *J. Chem. Phys.* **79**, 926–935.
14. Lybrand, T., McCammon, J. & Wipff, G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 833–835.
15. York, D., Darden, T., Pedersen, L. & Anderson, M. (1993) *Biochemistry* **32**, 1443–1453.
16. Ryckaert, J., Ciccotti, G. & Berendsen, H. (1977) *J. Comput. Phys.* **23**, 327–341.
17. van Gunsteren, W. F. & Karplus, M. (1982) *Biochemistry* **21**, 2259–2274.
18. van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hol, W. G. J. & Postma, J. P. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4315–4319.
19. Berendsen, H. J. C., van Gunsteren, W. F., Zwinderman, H. R. J. & Geurtsen, R. G. (1986) *Ann. N.Y. Acad. Sci.* **482**, 269–286.
20. Heiner, A. P., Berendsen, H. J. C. & van Gunsteren, W. F. (1992) *Proteins* **14**, 451–464.
21. Yang, A.-S., Gunner, M. R., Sampogna, R., Sharp, K. & Honig, B. (1993) *Proteins* **15**, 252–265.
22. Petsko, G. & Ringe, D. (1984) *Annu. Rev. Biophys. Bioeng.* **13**, 331–371.
23. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
24. Housset, D., Kim, K.-S., Fuchs, J., Woodward, C. & Wlodawer, A. (1991) *J. Mol. Biol.* **220**, 757–770.
25. Levitt, M. & Perutz, M. (1988) *J. Mol. Biol.* **201**, 751–754.
26. Perutz, M. (1993) in *The Chemical Bond*, ed. Zewail, A. (Academic, Boston), p. 17.
27. Rodhan, D., Suzuki, S., Suenram, R., Lovas, F., Dasgupta, S., Goddard, W., III, & Blake, G. (1993) *Nature (London)* **362**, 735–737.