PROTEINS: Structure, Function, and Bioinformatics 55:568–581 (2004)

Investigation of Glucose Binding Sites on Insulin

Vincent Zoete,[1,2] Markus Meuwly,[1,2,*] and Martin Karplus[1,3,*]
[1]*Laboratoire de Chimie Biophysique, ISIS/Université Louis Pasteur, Strasbourg Cedex, France*
[2]*Chemistry Department, University of Basel, Basel, Switzerland*
[3]*Department of Chemistry and Biological Chemistry, Harvard University, Cambridge, Massachusetts*

ABSTRACT **Possible insulin binding sites for D-glucose have been investigated theoretically by docking and molecular dynamics (MD) simulations. Two different docking programs for small molecules were used; Multiple Copy Simultaneous Search (MCSS) and Solvation Energy for Exhaustive Docking (SEED) programs. The configurations resulting from the MCSS search were evaluated with a scoring function developed to estimate the binding free energy. SEED calculations were performed using various values for the dielectric constant of the solute. It is found that scores emphasizing non-polar interactions gave a preferential binding site in agreement with that inferred from recent fluorescence and NMR NOESY experiments. The calculated binding affinity of −1.4 to −3.5 kcal/mol is within the measured range of −2.0 ± 0.5 kcal/mol. The validity of the binding site is suggested by the dynamical stability of the bound glucose when examined with MD simulations with explicit solvent. Alternative binding sites were found in the simulations and their relative stabilities were estimated. The motions of the bound glucose during molecular dynamics simulations are correlated with the motions of the insulin side chains that are in contact with it and with larger scale insulin motions. These results raise the question of whether glucose binding to insulin could play a role in its activity. The results establish the complementarity of molecular dynamics simulations and normal mode analyses with the search for binding sites proposed with small molecule docking programs. Proteins 2004;55:568–581.**

Key words: insulin; α-D-glucose; molecular docking; SEED; MCSS; scoring; molecular dynamics; normal modes

INTRODUCTION

Insulin is a 51-amino acid hormone that regulates the physiological glucose level in blood. While synthesized and stored in the pancreas as a zinc-containing hexamer, actually composed of three dimers, the insulin hormone interacts with its cell surface receptor as a monomer. The insulin monomer contains two chains, the A chain consisting of 21 amino acids and the B chain with 30 amino acids. An understanding of the function of insulin at a molecular level of detail is not yet available.[1] Although there have been numerous experimental investigations of the activity of modified insulins[2] and of the structures of particular mutants,[3] such studies have so far provided limited information concerning the atomistic aspects of the function of this important hormone. A recent NMR NOESY and fluorescence quenching study of monomeric insulin in the presence of low concentration of D-glucose (2 μM for fluorescence spectroscopy and 2–4 mM in 20% deuterated acetic acid for NMR spectroscopy[4,5]) suggested that there exists a binding site with specific interactions between the bound glucose and certain insulin side chains.[6] The authors proposed that glucose binding might be involved in insulin activity, although no direct evidence for this was given. The structure of the complex was not determined, but the analysis of the experimental results showed that the proton resonances involved in this interaction correspond to methyl protons of insulin. It was suggested that the resonances can be attributed to surface residue(s), probably Val, Ile, or Leu, although no specific assignments were made. The insulin monomer has two isoleucine, A2 and A10, four valine, A3, B2, B12, and B18, and six leucine residues, A13, A16, B6, B11, B15, and B17 (Fig. 1). All are solvent-exposed, except Leu(B11) and Leu(B15), resulting in a large number of side chains that could possibly interact with glucose; in the dimer, Val(B12) is buried at the interface. By molecular dynamics simulation of the free insulin dimer (the experiments were done with the monomer), followed by manual docking of D-glucose on the insulin surface, the authors proposed that Val(B2) and Leu(B17) residues are good candidates for the observed NOEs with the D-glucose protons. Fluorescence studies of the system led to an estimate of the binding free energy of about −2 ± 0.5 kcal/mol.[6]

In this paper, we investigate the docking of D-glucose to insulin by a two-step procedure. We first perform a global search for glucose binding sites on the insulin monomer by an exhaustive search with the MCSS[7] and SEED[8,9] docking programs. MCSS has been found to provide results in good agreement with NMR experiments when used with a

*Correspondence to: Martin Karplus, Laboratoire de Chime Biophysique, ISIS/Universiteé Louis Pasteur, 8, alleé Gaspard Monge, BP 70028, 67083 Strasbourg Cedex, France and Department of Chemistry and Biological Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138. E-mail: marci@tammy.harvard.edu or Markus Meuwly, Chemistry Department, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. E-mail: m.meuwly@unibas.ch

Received 24 September 2004; Accepted 24 November 2004

Published online 1 April 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20071

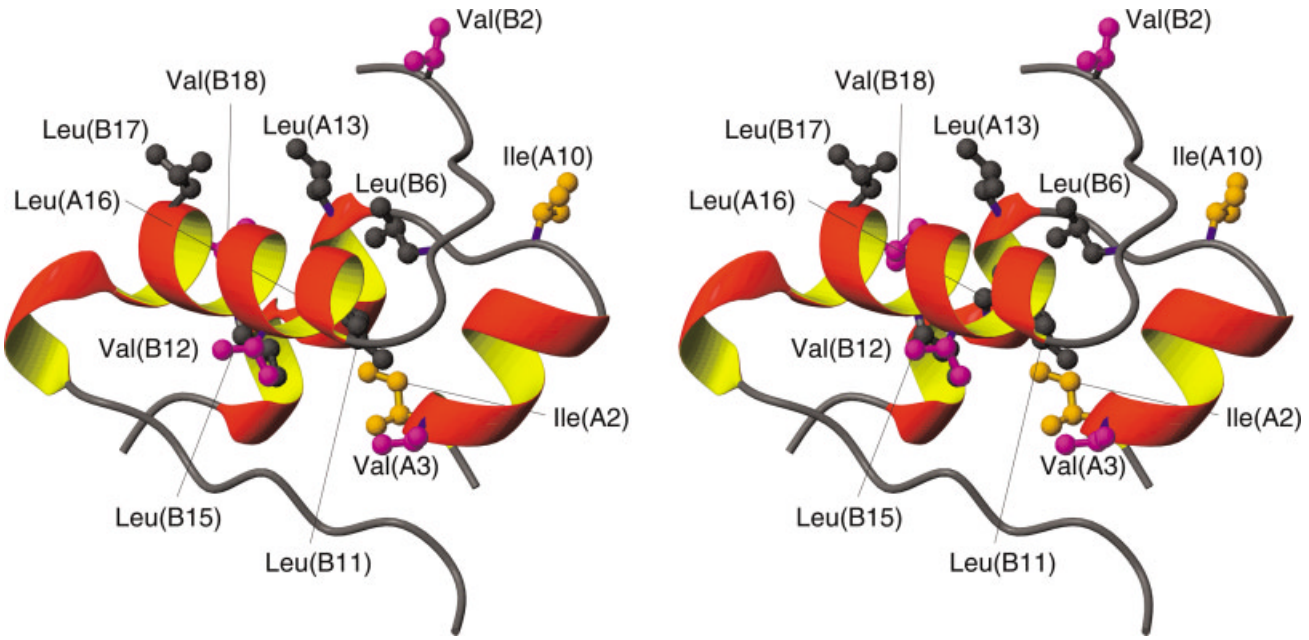

Figure 1

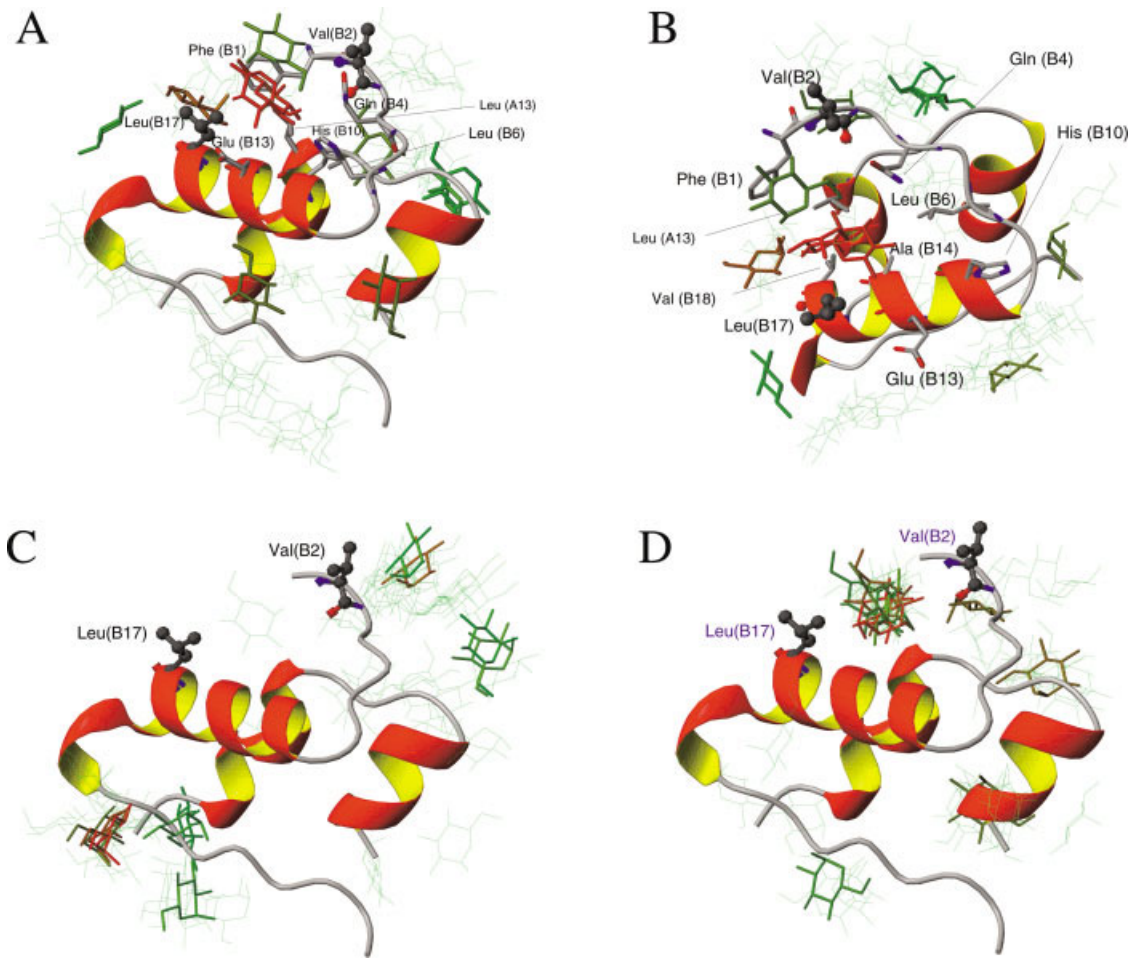

Figure 3

TABLE I. Atom Types, Partial Atomic Charges and Atomic Radii for Glucose, for All-Atom and Polar Hydrogen Parameter Sets[†]

	All-atom set			Polar hydrogens set		
Atom	Atom type	Radius	Partial charge	Atom type	Radius	Partial charge
C1	CTS	2.2750	0.200	CH1E	2.3650	0.290
H1	HAS	1.3200	0.090	—	—	—
O1	OHS	1.7700	−0.660	OH1	1.6000	−0.660
HO1	HOS	1.7700	0.430	H	0.8000	0.430
C5	CTS	2.2750	0.250	CH1E	2.3650	0.340
H5	HAS	1.3200	0.090	—	—	—
O5	OES	1.7700	−0.400	OE	1.6000	−0.400
C2	CTS	2.2750	0.140	CH1E	2.3650	0.230
H2	HAS	1.3200	0.090	—	—	—
O2	OHS	1.7700	−0.660	OH1	1.6000	−0.660
HO2	HOS	0.2245	0.430	H	0.8000	0.430
C3	CTS	2.2750	0.140	CH1E	2.3650	0.230
H3	HAS	1.3200	0.090	—	—	—
O3	OHS	1.7700	−0.660	OH1	1.6000	−0.660
HO3	HOS	0.2245	0.430	H	0.8000	0.430
C4	CTS	2.2750	0.140	CH1E	2.3650	0.230
H4	HAS	1.3200	0.090	—	—	—
O4	OHS	1.7700	−0.660	OH1	1.6000	−0.660
HO4	HOS	0.2245	0.430	H	0.8000	0.430
C6	CTS	2.2750	0.050	CH2E	2.2350	0.230
H61	HAS	1.3200	0.090	—	—	—
H62	HAS	1.3200	0.090	—	—	—
O6	OHS	1.7700	−0.660	OH1	1.6000	−0.660
HO6	HOS	0.2245	0.430	H	0.8000	0.430

[†]Atomic radii are in Å and partial charges in e. The polar hydrogen parameter set was used with MCSS and the all-atom set was used for the ranking of the replicas, the molecular dynamics simulations, and the normal mode analysis. See Figure 2 for atom names.

ranking function accounting for solvent effects.[10] The stability of the proposed binding pocket for the glucose molecule, as well as possible correlated motions between insulin and glucose in the calculated binding positions, are investigated by molecular dynamics simulations and normal mode analysis of the complex.

COMPUTATIONAL METHODS

The MCSS[7] and SEED[8,9] programs determine energetically favored positions and orientations for small molecules or functional groups on macromolecules of known three dimensional structure. For a detailed description of the two related but complementary approaches, the reader is referred to the corresponding papers.[7,8,9]

Since no coordinates for monomeric insulin are available, the starting structure was taken from the X-ray structure of the insulin dimer at 1.5 Å (Protein Data Bank (PDB[11]), entry 4INS[12]). Zinc atoms were removed and hydrogen atoms were added using the HBUILD[13] module of CHARMM.[14] It was assumed that the monomer insulin structure is close to that of the dimer. The monomer insulin was used for the calculations since the NMR experiments were carried out with the monomeric form of the hormone.

The glucose molecule was described by an all-atom force field for the ranking of the replicas, molecular dynamics simulations and normal mode analysis. The all-atom parameters and partial atomic charges developed for glucose by G. Liang and J. Brady, and provided with the CHARMM c29 distribution, were used for that purpose. This parameter set, and the polar hydrogen parameter set for proteins, TOPH19, were used to derive the polar hydrogen parameter set for glucose used with MCSS. Atom types, radius and partial atomic charges of glucose, for the two parameter sets, are provided in Table I. The atom names for glucose are provided in Figure 2. Insulin was described by the all-atom CHARMM22 parameter set.[15]

Fig. 1. Cross-eye stereo-view of the leucine (grey), isoleucine (orange) and valine (magenta) side chains of the insulin monomer. The latter is in ribbon representation and the side chains are shown in ball and stick representation. All figures of molecular structures were prepared with the MOLMOL program.[31]

Fig. 3. **A, B:** The positions of the 50 best glucose replicas obtained with MCSS, according to of Eq. (1) The score ranges from −1.3 kcal/mol (green) to −3.5 kcal/mol (red). Important residues defining the best binding site are shown in thick lines, except Val(B2) and Leu(B17), which are shown in ball and stick. B is the same as A, but seen from above. **C:** The positions of the 55 best glucose replicas, obtained with SEED and $\epsilon = 1$ for the solute, according to the SEED score. The score ranges from −3 kcal/mol (green) to −9.11 kcal/mol (red). **D:** The positions of the 50 best glucose replicas, obtained with SEED and $\epsilon = 4$ for the solute, according to the SEED score. The score ranges from −8.51 kcal/mol (green) to −11.51 kcal/mol (red). The insulin monomer is in ribbon representation. Glucose replicas are shown in thin lines, except the ten best ones, which are in thick lines.

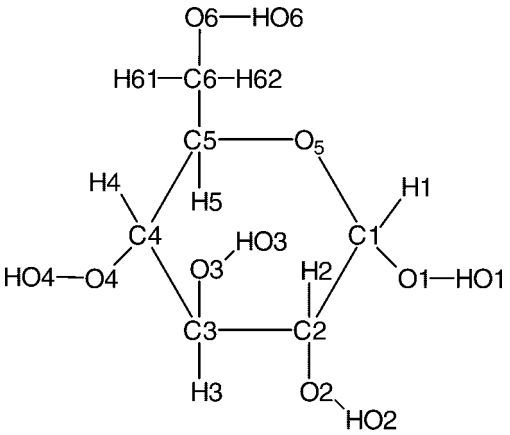

Fig. 2. Atom names for the glucose.

MCSS

Docking With MCSS

The MCSS calculations were performed as follows. To search for possible binding sites, the entire insulin monomer molecule (783 atoms) was used. Within a 20 Å radius sphere centered at the center of gravity of insulin, 10,000 glucose replicas were randomly placed outside the protein. The structure of the replicas was minimized and gathered to find those in energetically favorable positions and orientations (MCSS protocol). The minimizations consisted of 500 steps of steepest descent (SD) followed by 300 steps of additional steepest descent; positions of glucose replicas before and after the final 300 steps of minimization were used to make a first gathering of the replicas. Finally 20 repetitions of 500 steps of conjugated gradient minimization were performed. If the RMS energy gradient fell below a 0.0001 kcal/mol.Å tolerance, the minimization was stopped before the specified number of steps was reached. A 7.5 Å cutoff was applied on the shifted nonbonded interactions. Replica gathering was carried out after each minimization cycle, except the first one. If two replicas were positioned within 4 Å RMSD and if this RMSD decreased relative to the last minimization cycle, then they were considered to have converged to a common binding position.

At the end of the MCSS run, the RMS coordinate separation between two replicas is at least 4 Å. This is more coarse grained than in standard drug design problems where a value of 0.2 Å is generally used for the geometrical convergence criterion.[7,10,16] The large value for the convergence criterion was used here to limit the number of replicas obtained. After each minimization cycle, except the first, the replicas whose interaction energy with the protein were above a given cutoff were discarded. This energy cutoff was set to 500 kcal/mol for the first gathering (after the second minimization cycle) and was geometrically decreased by a factor of 0.5 at each cycle to a final value of 3 kcal/mol for the last one (i.e., 500 for gathering 1, 248.5 for gathering 2, 125.75 for gathering 3, etc. . .). The minimizations were carried out in vacuo, with a distance-dependent dielectric function $\epsilon = r$, to limit the computation time required. Solvent effects for evaluating the binding free energy are calculated with the UHBD program[17] (See the following section on the scoring function). During the minimization, the protein is kept fixed, whereas the conformation of the glucose can distort to adapt to its environment. The "fixed protein" approximation is in part justified by the NMR TOCSY experiments which showed that, for low molar ratios of D-glucose to insulin, D-glucose does not induce local conformational changes of the hormone.[6] D-glucose only influences the side chains of the residues involved.[6]

Van der Waals and electrostatic Coulomb interaction energies of insulin with each replica obtained by MCSS were calculated using the CHARMM22 all-atom force field.[15] For each insulin-glucose complex, HBUILD[13] was used to explicitly add all hydrogen atoms of the glucose molecule and to optimize the position of the hydrogen atoms linked to heavy atoms which were closer than 4 Å to any glucose atom. Van der Waals and electrostatic interaction energies were calculated after 150 steps of SD minimization using a (distance-independent) dielectric constant of 1 and a 20 Å cutoff for the non-bonded interactions. A harmonic constraint of 0.2 kcal/mol.Å^2 was applied to all heavy atoms of the replicas. This minimization was necessary because the all-atom CHARMM22 force field was introduced at this point for the ligand. Clashes due to hydrogen atom placement were released while the position of the replica was maintained. The RMSD between the positions of the replicas before and after minimization was generally less than 0.1 Å.

Scoring function

To rank the proposed glucose binding positions a scoring function was used. We note that the scoring calculation, which introduces the solvent effect, is considerably more computer intensive, so that it is limited to positions obtained with a faster algorithm. This scoring function is based on a study of the binding free energy for HIV-1 protease ligands.[18] The equation to estimate the binding free energy after minimization of the ligand in a fixed protein is

$$\text{score} = 2.11 + 0.0434 \times E_{elec} + 0.145 \times \Delta G_{solv} - 0.0196 \times SAS_{bur} \quad (1)$$

The electrostatic Coulomb interaction energy between the ligand and the protein, E_{elec}, is calculated using CHARMM and the electrostatic solvation free energy contribution to the binding free energy, ΔG_{solv}, is calculated using UHBD.[19] This latter is equal to the difference of the solvation free energies between the complex and the separated ligand and protein, $\Delta G_{solv}^{complex} - (\Delta G_{solv}^{protein} + \Delta G_{solv}^{ligand})$. The non-polar contribution to the free energy of binding is assumed to be proportional to the loss of solvent accessible surface, $SAS_{bur} = SAS_{ligand} + SAS_{protein} - SAS_{complex}$.[20]

SEED

In SEED, the small molecules are placed in the region of interest according to certain geometry criteria and to the

polar or apolar nature of both the receptor and the ligand.[8] After each placement, the binding energy is estimated. Two placement procedures are used. For polar docking, the small molecule is placed so that it makes at least one H-bond with a protein polar group at optimal distance. Predefined rules allow the distribution of vectors of unitary length on all H-bond groups of the receptor and the ligand, in the direction of an ideal H-bond geometry. For the receptor, additional vectors are distributed uniformly on a spherical region around the ideal direction to increase the sampling. The docking of polar molecules is achieved by matching a H-bond vector of the receptor with a H-bond vector of a fragment, at the ideal distance. The small molecule is then rotated around that direction to improve the sampling. For apolar docking, points are placed on the solvent-accessible surface (SAS) of the receptor. Vectors are defined by joining the points, corresponding to the most hydrophobic part of the SAS, to the corresponding atom center. Similarly, vectors are also determined for points uniformly distributed on the small molecule SAS. Apolar molecules are then docked by matching a vector of the molecule with a vector of the receptor at the optimal van der Waals distance. To increase the sampling, additional rotations of the small molecule, around the joining axis, are performed. No minimization of the small molecule position or of its conformation is done.

As we are interested in all possible binding positions of the glucose molecule on the insulin surface, we included all the insulin residues in the "binding site" list. The maximal numbers of polar and apolar vectors for docking were 3000 and 500 respectively. Both polar and apolar docking procedures were used. The "fixed protein" approximation was used as for MCSS.

The screening of the binding positions by SEED were performed by a two-step process.[9] In the first step, a fast model with an approximate treatment of solvation was used to estimate the binding energies and discard unfavorable positions. In the fast model the electrostatic term is based on the assumption that the electrostatic desolvation can be approximated by removing the first layer of water molecules at the binding site and on the use of the Coulomb approximation for the electric displacement.[9] The positions were then clustered and sorted according to the estimated binding energy. In the second step, the binding energies of the 10 best binding positions of each cluster were re-calculated using a more accurate solvation model.[8] In this latter model, the electrostatic desolvation energy of the receptor was calculated by the Coulomb field approximation, whereas the partial electrostatic desolvation of the fragment and the screened fragment–receptor electrostatic interactions were calculated using the generalized Born approximation. Four different runs were performed, using values of 1, 2, 3, and 4 for the dielectric constant of the solute (insulin and glucose). Since, with SEED, the solvent effect is accounted for by continuum models, no distance-dependent dielectric constant was used for the docking, unlike MCSS.

Molecular Dynamics Simulations

MD simulations with explicit solvent and stochastic boundary conditions

To test the best binding positions obtained from MCSS and SEED, molecular dynamics simulations were performed using stochastic boundary conditions[21] starting from different positions of the glucose ligand with respect to insulin. The three energetically most favored binding positions of glucose according to MCSS/Eq. (1), and to SEED/$\epsilon = 4$ were used as initial structures. The insulin/glucose complex was solvated with a 17 Å sphere of TIP3P[22] water molecules. This sphere was centered on the C5 glucose atom and contained between 447 and 504 water molecules, depending on the initial position of the glucose. Insulin heavy atoms in the buffer region (between 13 and 17 Å from the C5 glucose atom) were restrained by a harmonic constraint of 5 kcal/mol.Å^2 applied on the heavy atoms. Insulin heavy atoms in the buffer region were coupled to a heat bath, using the Langevin equation of motion and a 250 ps^{-1} friction constant.[23] Insulin atoms outside the buffer region were held fixed. The water molecules were kept within the sphere through the use of a solvent boundary potential[24] (SBOUND module of CHARMM) and a friction constant of 62 ps^{-1} was applied to the water oxygens.[23] A 12 Å cutoff was applied to the switched van der Waals and shifted electrostatic interactions, with $\epsilon = 1$. The SHAKE[25] method was used to fix the length of the bonds involving a hydrogen atom and a 1 fs time step was used to propagate the equations of motion. The all-atom CHARMM22 force field was used to describe the system. Starting from 0 K, the system was gradually heated and equilibrated at 300 K during 150 ps. At the starting of the heating process, the glucose and insulin heavy atoms in the reaction region were restrained by a harmonic constraint of 5 kcal/mol.Å^2. This harmonic constraint was gradually removed during the heating/equilibration process. Then, the molecular dynamics production run was performed at 300 K for 500 ps. During the production run, all atoms in the reaction region were unconstrained.

MD simulation with explicit solvent and periodic boundary conditions

One molecular dynamics simulation was performed using explicit solvent and periodic boundary conditions[21] (PBC) starting from the best glucose binding position according to MCSS/Eq. (1). The all-atom CHARMM22 force field was used to describe the system. The insulin-glucose complex was solvated in a $46.6 \times 43.5 \times 34.1$ Å^3 water box previously equilibrated at 300 K and 1 atm. Water molecules overlapping with insulin or glucose were removed; the system contained 1970 water molecules and a total of 6717 atoms. For all the subsequent MD simulation, the Verlet leapfrog integrator was used and a 12 Å cutoff was applied to the switched van der Waals and shifted electrostatic interactions, with $\epsilon = 1$. A 1 fs time step was used. The solvent was first equilibrated at 300 K during 100 ps in the presence of the fixed insulin. Then, the entire system was heated to 300 K during 6 ps, while

the glucose and insulin heavy atoms were restrained by a harmonic constraint of 2 kcal/mol.Å^2. This harmonic constraint was gradually removed during the following 100 ps equilibration run. Finally the molecular dynamics production run was performed at 300 K during 1 ns. During the production run no constraints were applied to any atoms.

Normal Mode Analysis

Normal modes were calculated for monomeric insulin alone and for the energetically most favored insulin-glucose complexes according to MCSS/Eq. (1), the same as used for the molecular dynamics simulations. In each case, the system was minimized in vacuo, with $\epsilon = 4r$, using the Adopted Basis Newton-Raphson minimization algorithm, until the root mean square of the energy gradient reached a value of 10^{-7} kcal/mol.Å. This gradient value has been shown to be satisfactory for calculating normal modes and is expected to yield only real frequency modes.[26] The RMS deviation between the structures before and after minimization is about 1.1 Å for insulin backbone atoms and 1.7 Å for glucose heavy atoms, when present. The regions showing the greatest deviation are residues of the N and C-terminus of chain B, with RMSD up to 2–3 Å for backbone atoms. The important interactions between the glucose and insulin residues Leu(A13), Gln(B4), Ala(B14), Leu(B17) and Val(B18) are conserved.

The VIBRAN module of the CHARMM program was used to calculate the force constant matrix and to diagonalize it to determine the first 100 normal modes and normal mode frequencies for the fully minimized structures. The normal modes were calculated in vacuo, with $\epsilon = 4r$ and the same cutoff conditions on the non-bonded energy terms as those used in the MD study.

Cross Correlated Motions Between Glucose and Insulin Residues

The cross-correlation coefficient C_i between the displacement of insulin residue i and glucose is given by

$$C_i = (\Delta r_i \cdot \Delta r_{glucose})/(\langle \Delta r_i^2 \rangle \langle \Delta r_{glucose}^2 \rangle)^{1/2} \qquad (2)$$

where Δr_i and $\Delta r_{glucose}$ are respectively the displacement from the mean position calculated for the backbone atoms of insulin residue i and for the heavy atoms of glucose, during a MD trajectory or for a given normal mode.

To calculate the cross-correlation coefficients from the MD simulation in PBC, the following protocol was used. The overall translation and rotation of the system was removed by minimizing the RMSD, calculated using all the C^α atoms, of all conformations with respect to the starting one.[27,28] The time window of interest for the trajectory was divided into successive blocks. For each of these blocks, a mean structure was determined and the C_i coefficients were calculated from the fluctuations for the time interval. The global C_i were obtained by averaging over all blocks. Correlation coefficients were calculated for time blocks rather than directly from the entire trajectory for two reasons. First, since correlated motions depend on the time scale over which data are gathered, different types of correlations can be demonstrated using different time scales.[29] Second, the correlation coefficients are calculated from a mean position of the atoms during the considered block. This mean position becomes less meaningful if it is calculated for a long block where large scale motions can occur. Different block lengths, ranging from 10 to 50 ps, were examined and all provided correlated motions similar to those described in the Results section. C_i correlation coefficients were also calculated for each normal mode and averaged over the 10 normal modes with the lowest frequencies, which were found to account for most of the large scale motions.

RESULTS AND DISCUSSION

We first describe the results from the MCSS calculations. 271 positions were found by MCSS, with a scoring ranging from −3.51 to +3.65 kcal/mol. Table II gives the calculated energies for the 15 best replicas according to Eq. (1). Figure 3A and 3B show the MCSS results, with the replicas colored according to the score from Eq. (1). The large pocket situated between Val(B2) and Leu(B17) is found to have a larger affinity for glucose than the rest of the insulin surface. This pocket is defined by the side chains of residues Phe(B1), Val(B2), Gln(B4), Leu(B6), His(B10), Glu(B13)—mainly C_β, Ala(B14), Leu(B17), Val(B18), and Leu(A13). Replicas found in this pocket are ranked, 1, 2, 3, 7, 16, 21, 34, and 42 according to Eq. (1). These replicas have non-polar interactions with the side chains. For example, the first replica has van der Waals interactions with Leu(A13), Gln(B4) (no hydrogen bond), Leu(B6), Ala(B14), Leu(B17), and Val(B18) side chains. The second replica has van der Waals interactions with Leu(A13), Gln(B4) (no hydrogen bond), Leu(B6), Ala(B14), and Leu(B17) side chains. The third replica has van der Waals interactions with Leu(A13), Phe(B1), Leu(B17), and Val(B18) side chains. The seventh replica has van der Waals interactions with Leu(A13), Phe(B1), and Val(B2) side chains (see Table III).

These results are in agreement with the limited experimental results obtained by NMR spectroscopy.[6] They suggested that glucose interacts with insulin methyl protons, which could be attributed to surface Val, Ile, or Leu residues. The authors retained Val(B2) and Leu(B17) as possible interacting residues with the glucose molecule. The different positions found in this pocket agree with the existence of multiple binding sites suggested by fluorescence experiments.[6] The estimated binding free energies of the glucose molecules positioned in this pocket calculated from Eq. (1) range from −3.5 kcal/mol to −1.4 kcal/mol. These values are close to the experimental value of -2 ± 0.5 kcal/mol. It should be noted that Eq. (1) has been developed to evaluate the binding free energy of HIV-1 protease inhibitors, and that no insulin ligand has been used to refine its parameters.

Figure 3C and 3D show the results obtained with SEED for different values of the solute dielectric constant. In Figure 3C the positions of the 55 best replicas found with SEED and $\epsilon = 1$ for the solute are shown. They have estimated binding free energies ranging from −9.11 to −3 kcal/mol. The glucose replicas ranked 1, 2, and 4 make

TABLE II. CHARMM Vacuum Interaction Energies, Electrostatic Solvation Free Energy Contribution to the Binding Free Energy, Buried Surface Upon Complexation, and Values of the Scoring Function for the 15 Best Glucose Replicas According to Equation (1)[†]

MCSS					CHARMM IE		Score	
rank	E_{vdW}	E_{elec}	ΔG_{solv}	SAS_{bur}	Energy[a]	rank[b]	*score*[c]	rank[d]
292	−11.21	−4.09	9.12	344.95	−15.30	217	−3.51	1
135	−11.16	−4.80	14.52	351.74	−15.96	212	−2.89	2
293	−9.15	−24.17	16.65	306.41	−33.32	56	−2.53	3
343	−6.34	−1.56	6.04	260.50	−7.90	263	−2.19	4
51	−8.24	−35.11	28.74	351.92	−43.35	22	−2.14	5
321	−7.27	−5.49	6.20	249.00	−12.76	238	−2.11	6
138	−7.68	−12.79	14.47	292.89	−20.47	165	−2.09	7
132	−4.84	−16.04	14.63	278.67	−20.88	161	−1.93	8
30	−1.20	−27.14	20.08	294.02	−28.34	87	−1.92	9
251	−7.59	−5.95	12.64	283.38	−13.54	228	−1.87	10
304	−7.27	−5.53	10.30	267.10	−12.80	237	−1.87	11
165	−7.15	−20.07	17.16	283.87	−27.22	96	−1.84	12
56	−5.21	−35.25	27.95	330.47	−40.46	28	−1.84	13
180	−7.29	−6.22	10.61	265.07	−13.51	229	−1.82	14
212	−8.18	−8.38	12.98	277.09	−16.56	205	−1.80	15

[†]All energies are in kcal/mol. SAS_{bur} is in Å^2.
[a]Interaction energy calculated with CHARMM (sum of van der Waals, E_{vdW}, and electrostatic, E_{elec}, terms).
[b]Rank according to this interaction energy.
[c]Score calculated using Eq. (1).
[d]Rank according to Eq. (1).

TABLE III. Interactions Between Insulin and the Most Favored Replicas Found by MCSS/Equation (1), or with SEED ($\epsilon_1 = 1$ and $\epsilon = 4$ for the solute)[†]

Ranking	Interactions	Scoring energy
MCSS/Eq. (1)		
1	VW: Leu(A13), Gln(B4) (no HB) Leu(B6), Ala(B14), Leu(B17), Val(B18)	−3.5
2	VW: Leu(A13), Gln(B4) (no HB) Leu(B6), Ala(B14), Leu(B17)	−2.9
3	VW: Leu(A13), Phe(B1), Leu(B17) Val(B18)	−2.5
7	VW Leu(A13), Phe(B1) and Val(B2)	−2.1
SEED/$\epsilon = 1$		
1, 2 and 4	HBs: Phe(B24) [BB], Asn(A21) [SC]	−9.1, −7.2, −5.8
3	HBs: Asn(B3) [BB] VW: Val(B2) [SC]	−7.1
SEED/$\epsilon = 4$		
1	HBs: Ala(B14) backbone CO VW: Leu(A13), Gln(B4), Leu(B6), Glu(B13), Ala(B14), Leu(B17) [SC]	−11.5
2	HBs: His(B10) VW: Leu(A13), Gln(B4), Glu(B13), Ala(B14), Leu(B17)	−11.1
3	HBs: Gln(B4) [SC] VW: Leu(A13), Gln(B4), Ala(B14), Leu(B17)	−10.5

[†]HB: H-bonds. VW: van der Waals interactions. [SC] side chain. [BB] backbone. Energies (in kcal/mol) are calculated according to the scoring function used to rank the binding positions. Energies calculated with the different methods (and different dielectric constant for the solute) should not be compared.

hydrogen bonds with Phe(B24) backbone atoms and the Asn(A21) side chain. The third best replica makes hydrogen bonds with Asn(B3) backbone atoms and van der Waals interactions with the Val(B2) side chain. Only one glucose replica, ranked 20 (−3.83 kcal/mol), was found in the Val(B2)-Leu(B17) pocket, the preferred binding site suggested by the experiment. It makes van der Waals interactions with Leu(A13), Gln(B4), Leu(B6), Ala(B14), Leu(B17) and Val(B18) side chains. Figure 3D shows the positions of the 50 best replicas found with SEED from calculations with $\epsilon = 4$ for the solute. The energetically most favorable replica obtained with SEED ($\epsilon = 4$) makes

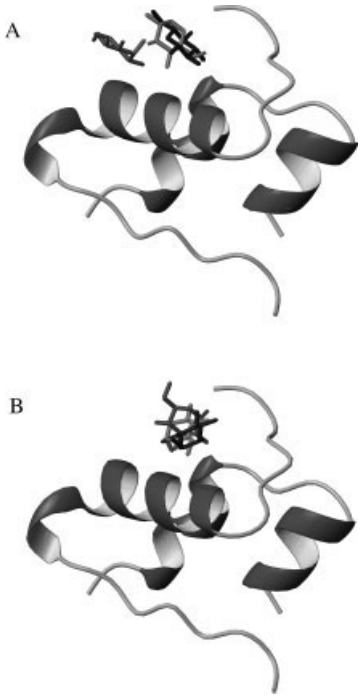

Fig. 4. **A:** The starting positions of the glucose molecules (heavy atoms only), relative to insulin, for the three molecular dynamics simulations of complexes proposed by MCSS. **B:** The starting positions of the glucose molecules (heavy atoms only), relative to insulin, for the three molecular dynamics simulations of complexes proposed by SEED.

Fig. 5. Frames of the MD run with stochastic boundary conditions starting from the MCSS glucose position ranked 1 by Eq. (1). For clarity, only heavy atoms are shown, and only the starting conformation of insulin is given in ribbon representation. The starting position of glucose is in thick lines, colored according to the atom type. Positions separated by 50 ps and explored by the glucose molecule during the MD simulation are shown in thin black lines. Val(B2) and Leu(B17) are shown in ball and stick representation.

van der Waals interactions with Leu(A13), Gln(B4), Leu(B6), Glu(B13), Ala(B14), and Leu(B17) side chains and a hydrogen bond with the Ala(B14) backbone CO atoms and HO1. The second best SEED replica has van der Waals interactions with Leu(A13), Gln(B4), Glu(B13), Ala(B14), and Leu(B17) and a hydrogen bond with the ND1 atom of His(B10) and O2. The third best SEED replica makes van der Waals interactions with Leu(A13), Gln(B4), Ala(B14), and Leu(B17) and a hydrogen bond with the CO atoms of the Gln(B4) side chain and HO1 (see Table III). Using the scoring function from SEED, the 50 best replicas shown in Figure 3D have estimated binding free energies ranging from −11.51 to −8.51 kcal/mol, which are significantly larger than the experimental values. Using the scoring function of Eq. (1) the estimated binding free energy of the 50 best SEED replicas ranges from −3.12 to −0.12 kcal/mol. The difference between the two sets of results is likely to be due to the fact that Eq. (1), which was developed to estimate the binding free energy of HIV-1 protease inhibitors, takes the entropic penalty of binding into account implicitly and has a term proportional to the buried surface, whereas the SEED energy function contains only intermolecular van der Waals interactions and continuum electrostatics. The best SEED replicas according to Eq. (1) are those bound in the Val(B2)-Leu(B17) pocket. We can see in Figure 3D that this pocket has also the largest affinity for the glucose molecule according to the scoring function from SEED. Similar results were found using a dielectric constant of 2 or 3 for the solute (data not shown). This is in agreement with results obtained from MCSS with Eq. (1): replicas situated in the hydrophobic pocket between Val(B2) and Leu(B17) are favored by elevated dielectric constants.

Molecular dynamics simulations. Investigations using MCSS and SEED give information about the *static* preference of particular binding sites around the insulin monomer. However, for the situation in the in vivo environment it is also important to investigate the *dynamic* stability of a particular site. For this purpose, six 500 ps molecular dynamics simulations with stochastic boundary conditions were performed at 300 K. The initial structures were taken from the three most favored glucose positions relative to insulin according to MCSS/Eq. (1), and from the three most favored ones according to SEED/$\epsilon = 4$. Figure 4 gives the six starting structures. Figure 5 shows the positions of the glucose molecule during the 500 ps of the MD simulation, starting from the best position and orientation calculated by MCSS and scored with Eq. (1). The overall position of the glucose molecule shows little change during the MD run. Thus, D-glucose is stable within the binding pocket during the length of the simulation. The fluctuation of the glucose heavy atoms with respect to their average position during the 500 ps is around 1.8 Å, and the average RMSD from the starting position calculated for the heavy atoms is about 2.1 Å. All the positions explored by the glucose molecule during the MD run are within the space filled by the MCSS glucose replicas. No alternative positions proposed by MCSS/Eq. (1) are explored during this particular MD simulation. The MD simulations started from the second and third best glucose positions according to MCSS/Eq. (1) show larger fluctuations in the position of the ligand. The fluctuation of the glucose heavy atoms around their average position and the average RMSD from the starting position are 1.8 Å and 2.8 Å, and 2.5 Å and 2.8 Å for the second and third best glucose replicas, respec-

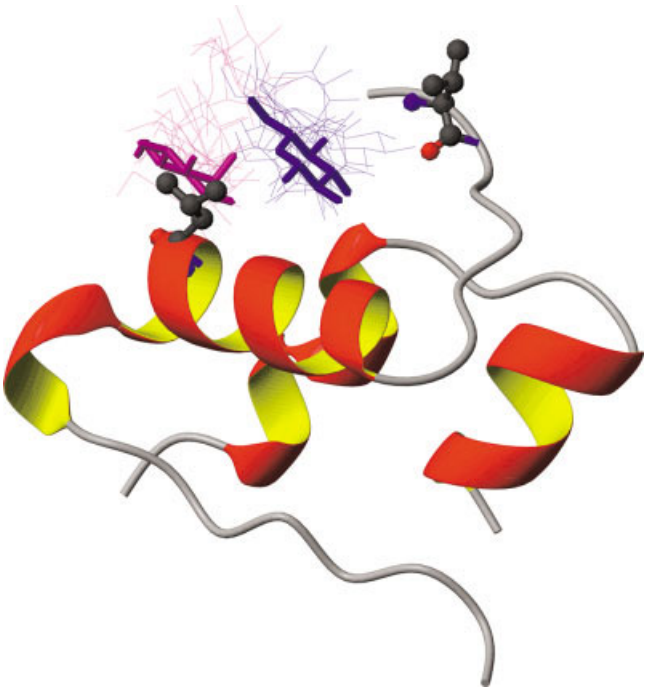

Fig. 6. Frames of the MD runs with stochastic boundary conditions starting from the MCSS glucose positions ranked 2 and 3 by Eq. (1). For clarity, only heavy atoms are shown, and only the starting conformation of insulin is given in ribbon representation. The starting positions of glucose are given in blue and magenta thick lines for the replicas ranked 2 and 3, respectively. Positions separated by 50 ps and explored by the glucose molecule during the MD simulation are shown in blue and magenta thin lines. Val(B2) and Leu(B17) are shown in ball and stick representation.

Fig. 7. Frames of the MD run with periodic boundary conditions starting from the MCSS glucose position ranked 1 by Eq. (1). For clarity, only heavy atoms of glucose and backbone atoms of insulin chain B are shown, and the backbone of chain A has been removed.

tively. However, the glucose molecule remains in the defined binding pocket and explores alternative positions proposed by MCSS. Figure 6 shows the positions of the glucose molecule during the two 500 ps of the MD simulations starting from the positions calculated by MCSS and scored 2 and 3 with Eq. (1). Similar results are found for the MD simulations started from the three most favored SEED/ϵ = 4 glucose binding positions. The fluctuation of the glucose heavy atoms around their average position and the average RMSD from the starting position are 1.4 Å and 2.4 Å, and 1.9 Å and 2.2 Å for the first and second best glucose replicas, respectively. The glucose shows a more dynamical behavior in the MD simulation started from the third most favored position; i.e., the fluctuation of the glucose heavy atoms around their average position and the average RMSD from the starting position are respectively 1.9 Å and 3.2 Å. However, in all cases, the glucose molecule stays within the binding pocket defined by Val(B2) and Leu(B17) and explores different binding positions found by SEED.

During the molecular dynamics simulations with stochastic boundary conditions, insulin residues in the reservoir region are fixed whereas residues in the buffer region are harmonically restrained. This limits the overall dynamical behavior of the insulin molecule. Therefore, to study possible correlations between the insulin and glucose motions, we performed a more computer intensive molecular dynamics simulation with periodic boundary conditions during 1 ns at 300 K. The simulation was started from the most favored position determined by MCSS/Eq. (1) for glucose. The entire complex was solvated in a box with explicit solvent molecules and no constraints were applied to the system during the 1 ns production run (see Methods). Figure 7 shows a series of frames of the MD simulation. Figure 8A gives the RMSD from the starting position calculated along the trajectory for the heavy atoms and for the center of gravity of the glucose molecule. As can be seen from these figures, the glucose molecule remains in its starting position during the first 230 ps of the simulation. During that period, the glucose molecule has van der Waals interactions with the Leu(A13), Gln(B4), Leu(B6), Ala(B14), Leu(B17), and Val(B18) side chains, as was the case for the starting conformation. Moreover, the glucose and insulin fluctuations during that period allow additional van der Waals contacts with the Phe(B1) and Val(B2) side chains and transient hydrogen bonds between the O6 oxygen atom of glucose (and to a lesser extent HO1 and O5) and the Gln(B4) side chain. A H-bond between HO6 and the Ala(B14) carbonyl takes place between 75 and 230 ps, and replaces the former H-bond between HO6 and a water molecule [see Figure 8(C)]. Between 150 and 300 ps the insulin chain B N-terminus shows a transition between the starting "closed" conformation and a more "open" conformation characterized by a larger distance between the backbone atoms of residues B17–B18 and B1–B3 [see Figs 7 and 8(B)]. During that transition, after 230 ps, the glucose molecule leaves the starting position and reaches a position, whose center of gravity is closer to that of the seventh best calculated position by MCSS/Eq. (1) (replica number 138), and where it still interacts with Leu(A13), Phe(B1) and, to a lesser extent, with Val(B2). The interactions with Gln(B4), Leu(B6), Ala(B14), Leu(B17), and Val(B18) are interrupted. Between 400 and 600 ps, the glucose molecule progressively returns to a position close to the starting one. This motion is accompanied by a decrease of the distance between residues B17–B18 and B1–B3. A transient H-bond between the O2 glucose atom and the Gln(B4) side chain exists between 490 and 520 ps. The van der Waals interaction between

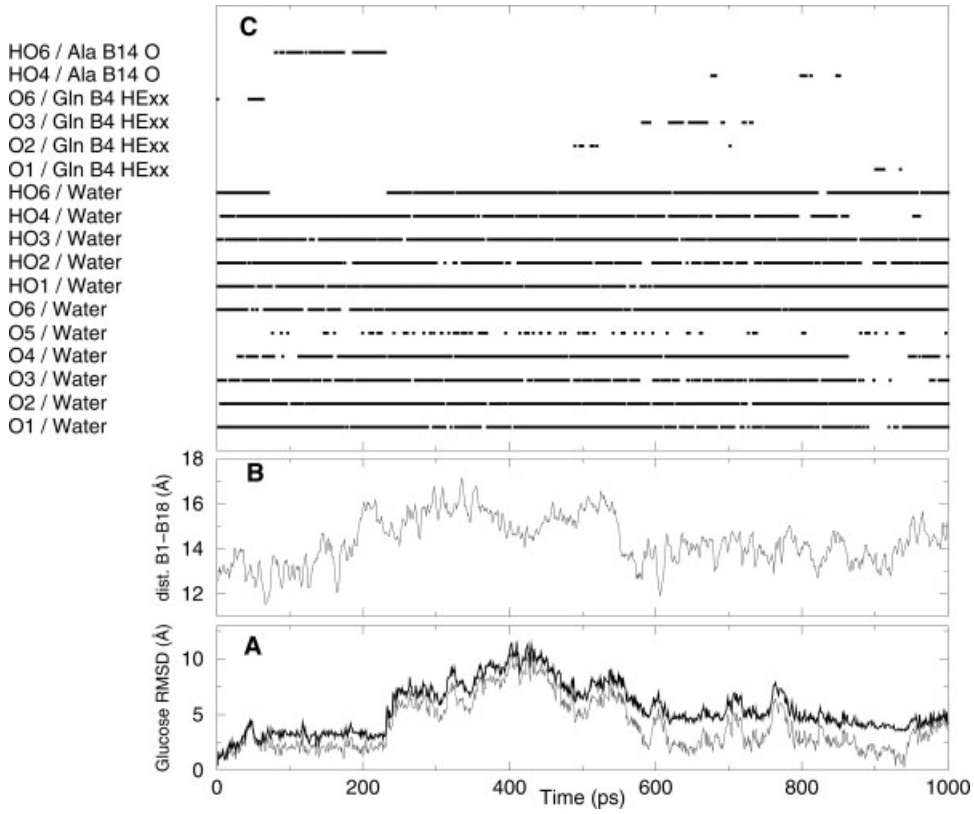

Fig. 8. MD run with periodic boundary conditions starting from the MCSS glucose position ranked first by Eq. (1). **A:** RMSD from the starting position calculated along the trajectory for the heavy atoms (thick line) and the glucose ring center of gravity (thin line) of the glucose molecule. **B:** Distance between the CA atoms of residues Phe(B1) and Val(B18) of insulin chain B during the MD simulation. **C:** Hydrogen bonds formed between the glucose molecule and water or insulin atoms during the MD simulation.

glucose and the Phe(B1) side chain is absent. After 600 ps, the glucose molecule is in a position close to the starting one, but in a different orientation. During that period, van der Waals interactions with Leu(A13), Val(B2), Gln(B4), Ala(B14), Leu(B17), and Val(B18) are dominant. Additional van der Waals interactions with His(B10) and Glu(B13) appear during the last 50 ps. Transient hydrogen bonds between the HO4 glucose atom and the Ala(B14) carbonyl, and between the O1, O2, and O3 glucose atoms and the Gln(B4) side chain are formed during that period and replace the corresponding hydrogen bonds between glucose and water molecules [see Fig. 8(C)]. Some very labile hydrogen bonds are formed between HO2, HO3, and O4 glucose atoms and the Gln(B4) side chain. With the exception mentioned above, all hydrogen bonds involving glucose and water molecules are present during the entire trajectory [see Fig. 8(C)]. Water molecules making hydrogen bonds with given glucose atoms exchange during the trajectory. For example, more than 70 water molecules exchanged a hydrogen bond with the HO6 glucose atom at different periods of the 1 ns trajectory. Figure 9 shows the hydrogen bonds network between insulin and water molecules observed for the starting conformation. The low number and the transient nature of the hydrogen bonds between glucose and insulin observed during this MD simulation are in agreement with the low binding free energy (≈ -2 kcal/mol) measured by fluorescence quenching experiments.[6] The MD simulation demonstrates that, whereas the glucose molecule remains within the binding site described above, it shows large motions and explores alternative binding positions. Transitions between alternate positions are correlated with large scale motions of the N-terminus of chain B relative to the C-terminal end of the chain B helix. All the insulin residues interacting with glucose during the trajectory were also found to interact with the best replicas found by MCSS/Eq. (1) or SEED/ϵ = 4 (see Table III).

Correlated motions between glucose and insulin residues were observed during the first 200 ps of the MD simulation. This window was chosen because the glucose molecule remains close to the starting position and the interactions between the glucose and insulin remain unchanged. Figure 10 gives the correlation coefficients between glucose and insulin residues. As can be seen, the motions of glucose are correlated to motions of residues that are in contact with it. Correlated motions are found with Ala(A13) and its flanking residues, with Gln(B4), and with residues Ala(B14), Leu(B17), and Val(B18) and their flanking residues.

To test the preference for the Leu(B17)/Val(B2) binding pocket, molecular dynamics simulations in explicit solvent with periodic boundary conditions were also carried out for 1 ns for two other binding positions found with MCSS outside the dominant binding pocket (Fig. 11). Both in-

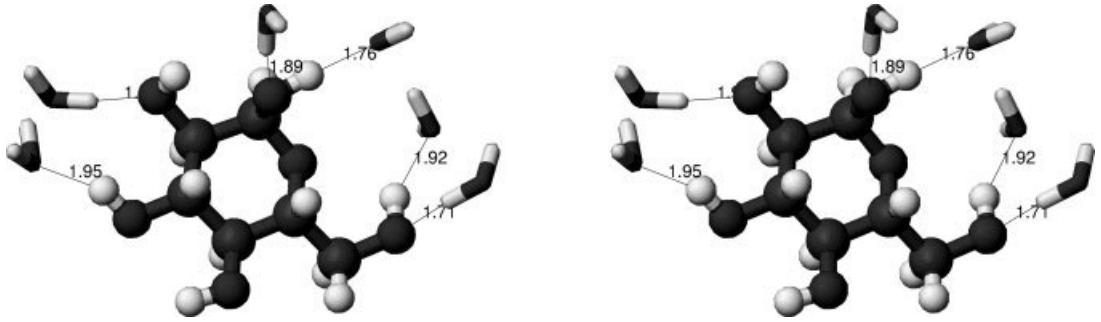

Fig. 9. Cross eye stereo view of the hydrogen bond network between insulin and water molecules observed for the starting conformation [MCSS replica ranked first by Eq. (1)] of the 1 ns molecular dynamics simulation of the insulin/glucose complex with periodic boundary conditions. For clarity, the insulin molecule, which is below the glucose molecule shown, is not displayed. Thin lines and numbers give hydrogen bonds and interatomic distances. The glucose molecule is in ball and stick representation and water molecules are in thick lines.

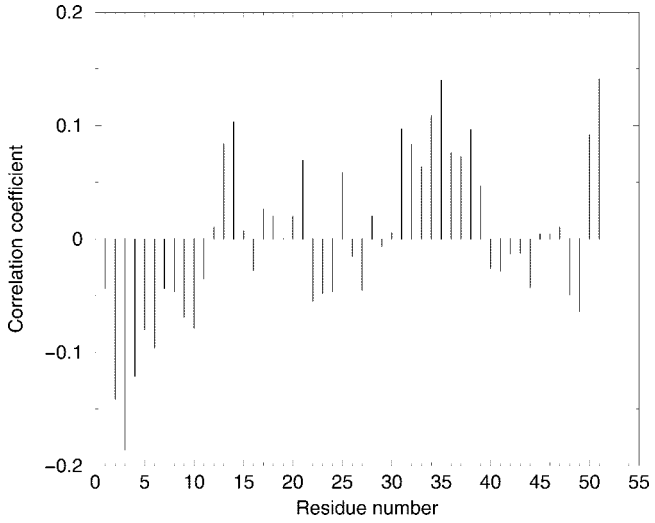

Fig. 10. Correlation coefficients between glucose and insulin residues motions during the first 200 ps of the MD simulation with periodic boundary conditions, averaged over blocks of 20 ps. Residues B1 to B30 are numbered 22 to 51. Negative C_i coefficients indicate anticorrelated motions (coupled but in opposite sense).

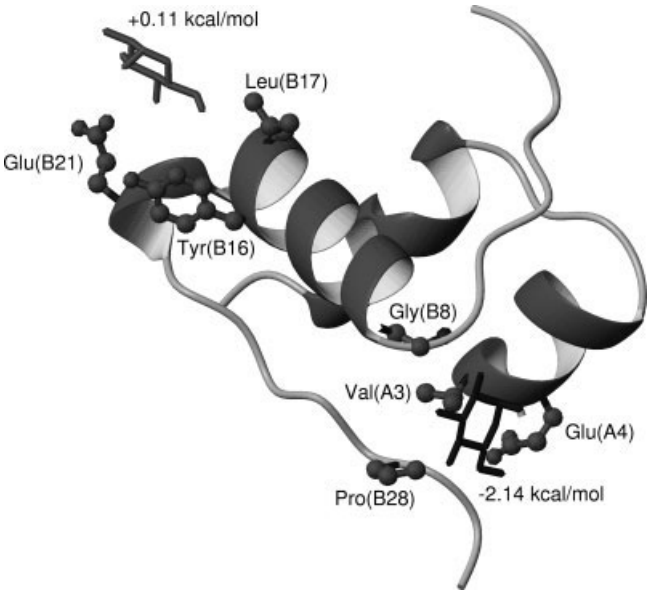

Fig. 11. The starting positions of the glucose molecules (heavy atoms only), relative to insulin, for the two molecular dynamics simulations with periodic boundary conditions carried out for the binding positions found with MCSS and having calculated binding free energies of −2.14 kcal/mol and +0.11 kcal/mol. Important insulin residues are shown in ball and stick representation.

volve contacts with leucine and valine side chains. One has a favorable binding free energy of −2.14 kcal/mol according to Eq. (1) (ranked 5), whereas the binding of the other one is unfavorable (i.e., +0.11 kcal/mol). The first replica makes van der Waals contacts with Val(A3) and Pro(B28) and has hydrogen bonds with the Glu(A4) side chain and the Gly(B8) backbone NH atoms. Binding positions close to this replica were also proposed by SEED/$\epsilon = 4$. During the MD simulation, the glucose molecule leaves its starting position, and after 350 ps, stays between 5 and 12 Å from its initial position. Van der Waals interactions with Val(A3) and Pro(B28), as well as the hydrogen bond with Gly(B8) backbone are lost. Transient hydrogen bonds with the Glu(A4) side chain are found. Van der Waals interactions with Pro(A8) appear after 500 ps. A high flexibility of the C-terminal part of the B chain, which includes Pro(B28), is unfavorable for the dynamic stability of this complex. The second replica selected to perform an MD simulation with PBC is ranked 185 by Eq. (1), with an unfavorable binding free energy of +0.11 kcal/mol. It was selected to test the relevance of Eq. (1) since a replica with an unfavorable binding free energy is expected to be expelled from its initial-binding position during an MD simulation. This replica makes hydrogen bonds with the Glu(B21) side chain and the Tyr(B16) carbonyl, and exchanges van der Waals contacts with Tyr(B16) and Leu(B17). No favorable positions were found for glucose in this region of the insulin surface by MCSS/Eq. (1) or SEED. During the MD simulation, the glucose molecule leaves its starting position after 300 ps. The hydrogen bonds with the Glu(B21) side chain and the Tyr(B16) carbonyl are lost and transient hydrogen bonds with the Leu(B17) and Val(B18) carbonyls appear, as well as additional van der Waals interactions with Leu(B17), Val(B18) and Phe(B1) side chains. After 800 ps, the glucose molecule leaves the insulin surface and does not have any more contacts with insulin. Thus proposed glucose binding positions outside

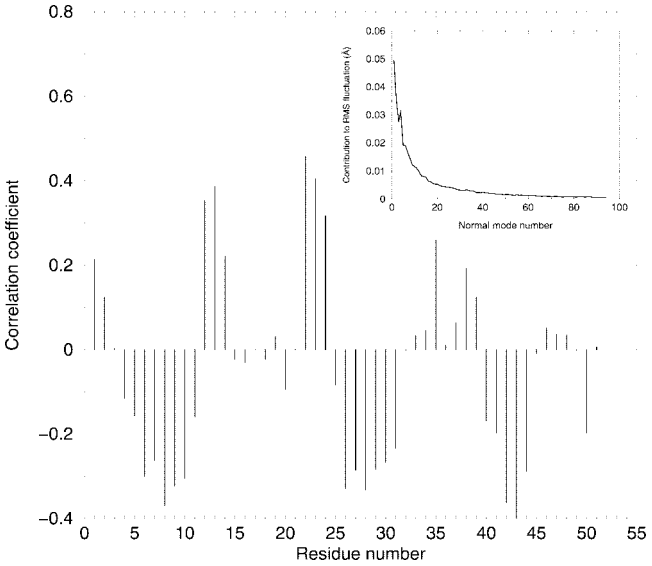

Fig. 12. Correlation coefficients between the motions of the glucose molecule and the insulin residues calculated from the ten normal modes with the lowest frequencies and for the most favored glucose position according to MCSS/Eq. (1). Inset gives the contribution of each mode to the atomic fluctuations. Residues B1 to B30 are numbered 22 to 51.

the Leu(B17)/Val(B2) pocket; especially the one with an unfavorable calculated binding free energy, appear to be less stable during MD simulations, in agreement with the ranking provided by Eq. (1).

MD simulations with PBC and explicit solvent have been performed for the insulin monomer without glucose (V. Zoete et al., in preparation). During these simulations, the chain B N-terminus shows essentially the same dynamical behavior as described above for the insulin/glucose complex, i.e., the results are compatible with glucose binding in the Leu(B17)/Val(B2) pocket.

Normal mode simulations. The 100 first normal modes were calculated for the most favored insulin-glucose complexes according to MCSS/Eq. (1), as well as for the monomeric insulin alone. We investigated the normal modes since they provide information about the atomic fluctuations and dynamical behavior of the system around the equilibrium state. No normal modes with imaginary-frequencies were obtained; six normal modes with zero frequencies, which correspond to translations and rotations of the entire system, were obtained. In the case of the complex between insulin and the most favored glucose replica according to MCSS/Eq. (1) the frequencies of the first 94 normal modes ranges from 4.7 to 45.2 cm^{-1}. The inset of Figure 12 shows the contribution of each mode to the atomic fluctuations summed over all the atoms. As can be seen, the observed contributions have the expected inverse frequency dependence[30] and the first ten normal modes are responsible for most of the large scale motions. In all cases (complexes or monomeric insulin alone), the three normal modes with lowest frequencies account for one of the following types of motion: large motions of residues B29 to B30, or B20 to B23, and another motion that can be described mainly as an increase of the distance between the N-terminus of chain B and the C-terminal end of the chain B helix (see Fig. 13). This latter normal mode, that also exists in the case of insulin alone, is correlated with a large-scale motion of the glucose molecule in the case of the insulin-glucose complexes. Correlated motions between the glucose heavy atoms and the insulin backbone atoms were calculated for each normal mode and averaged over the ten with the lowest frequencies. Figure 12 shows the correlated coefficients between the glucose and insulin residues motions. As can be seen, glucose motions are correlated with motions of residues A12–A14, B1–B3, B14 and B17–B18. These results are in agreement with those obtained from the molecular dynamics simulation of the complex in periodic boundary conditions.

CONCLUSION

This paper describes the investigation of putative binding positions of D-glucose on the insulin surface. Both the MCSS and SEED programs were used for this purpose. MCSS revealed a large number of possible binding sites. These positions were ranked using a scoring function developed to estimate the binding free energy of HIV-1 protease inhibitors. The scoring function emphasizes the influence of non-polar interactions. According to it, mainly replicas grouped in a hydrophobic pocket defined by Val(B2) and Leu(B17), and by other essentially hydrophobic residues [Phe(B1), Leu(B6), Ala(B14), Val(B18), Leu(A13) and the aliphatic portions of Gln(B4), His(B10), and Glu(B13)] are found. These results are in agreement with previous NMR NOESY data, which suggest that D-glucose interacts with insulin methyl protons that can be attributed to surface Val, Ile or Leu residues.[6] Val(B2) and Leu(B17) were selected as possible residues that interact with glucose by the authors, although no experimental data suggested this choice. The calculated binding free energy of -1.4 to -3.5 kcal/mol agrees with the experimentally measured value of -2.0 ± 0.5 kcal/mol (although Eq. (1) was developed for HIV-1 protease inhibitors and has not been refined for insulin) and the relevance of the binding site is supported by the results from MD simulations in explicit solvent.

Results obtained with SEED confirm the MCSS results. Using a low dielectric constant for the solute ($\epsilon = 1$) only one glucose binding position was found in the pocket defined by Val(B2) and Leu(B17). However, if the dielectric constant of the solute is increased ($\epsilon = 2$, 3, or 4) several favorable positions are found in the hydrophobic pocket. They also include the positions found by MCSS. Increasing the dielectric constant decreases the contribution of polar contacts to the binding free energy. This result confirms that glucose is predominantly found in apolar regions of insulin.

It is also relevant to determine the kinetic stability of the observed D-glucose binding positions using molecular dynamics simulations. We found that starting from the energetically most favored insulin-D-Glucose complex from MCSS calculations the system is stable during 500 ps of molecular dynamics simulation with stochastic boundary conditions and 1 ns of dynamics simulation with periodic

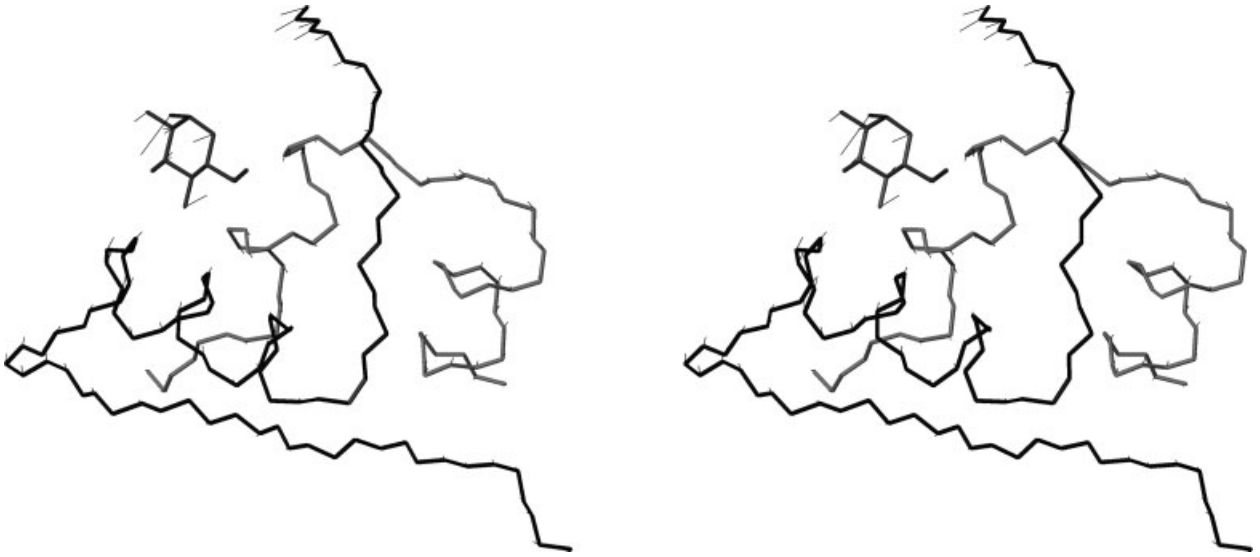

Fig. 13. Cross eye stereo view of the second normal mode calculated for the complex of insulin with the most favored glucose replica according to MCSS/Eq. (1). The glucose molecule (heavy atoms only) and the insulin backbone of chains A and B are shown in thick lines. Thin lines show the motions of insulin backbone atoms or glucose heavy atoms. The magnitude of the motion, indicated by the length of the thin lines, was amplified eight times relative to the amplitude calculated at 300 K.

boundary conditions. The simulations were carried out in explicit solvent at 300 K. The glucose molecule explores the region of the hydrophobic pocket and samples different favored binding positions found by MCSS or SEED. However, it does not leave the binding pocket. This shows that the glucose molecule binding modes are favored under ambient conditions including full solvation. The motions of the bound glucose are correlated with the motions of the insulin side chains that are in contact with it, and also with larger scale motions between the N-terminus of chain B and the C-terminal end of the chain B helix. Other insulin/glucose complexes suggested by MCSS, in which the glucose molecule is situated outside the pocket described above, appear less stable during MD simulations. These results show that molecular dynamics simulations and normal mode analysis of selected binding positions proposed by small fragment docking programs can be used to provide additional information about their relevance and stability.

At present there is no experimental evidence that glucose binding to insulin plays a role in its function. However, since the mechanism of insulin activation is not clear, the fact that such binding does occur is of interest. In the insulin hexamer, the Val(B2)/Leu(B17) pocket is situated at the interface between the dimers. Actually, each Leu(B17) of an insulin dimer is occupying a Val(B2)/Leu(B17) pocket of another dimer. Therefore, the docking of glucose in the Val(B2)/Leu(B17) pocket could compete with insulin hexamer formation. This would be of interest since the hexamer is inactive. Further experimental studies are necessary to test this possibility and determine the detailed structure of the glucose/insulin complex.

ACKNOWLEDGMENTS

We are grateful to Roland Stote and Annick Dejeagere for helpful discussions and thank Michael Schaefer for providing the CHARMM-UHBD interfaces. V.Z. was supported by Enanta Pharmaceuticals. Research in Basel (V.Z. and M.M.) was in part supported by a grant from the Insulin Dependent Diabetes Trust (IDDT). M.M. acknowledges financial support from the Schweizerischer Nationalfonds for a Förderungsprofessur. All calculations were performed at the Centre Informatique National de l'Enseignement Supérieur (CINES, Montpellier, France) and the Centre d'Etude du Calcul Parallèle de Strasbourg. (Illkirch, France).

REFERENCES

1. De Meyts P, Whittaker J. Structural biology of insulin and IGF1 receptors: implications for drug design. Nat Rev Drug Discov 2002;1:769–783.
2. Weiss MA, Nakagawa SH, Jia W, Xu B, Hua QX, Chu YC, Wang RY, Katsoyannis PG. Protein structure and the spandrels of San Marco: insulin's receptor-binding surface is buttressed by an invariant leucine essential for its stability. Biochemistry 2002;41: 809–819.
3. Ludvigsen S, Olsen HB, Kaarsholm NC. A structural switch in a mutant insulin exposes key residues for receptor binding. J Mol Biol 1998;279:1–7.
4. Weiss MA, Nguyen DT, Khait I, Inouye K, Frank BH, Beckage M, O'Shea E, Shoelson SE, Karplus M, Neuringer LJ. Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics. Biochemistry 1989;28:9855–9873.
5. Hua QX, Weiss MA. Comparative 2D NMR studies of human insulin and des-pentapeptide insulin: sequential resonance assignment and implications for protein dynamics and receptor recognition. Biochemistry 1991;30:5505–5515.
6. Falconi M, Bozzi M, Paci M, Raudino A, Purrello R, Cambria A, Sette M, Cambria MT. Spectroscopic and molecular dynamics simulation studies of the interaction of insulin with glucose. Int J Biol Macromol 2001;29:161–168.
7. Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins 1991;11:29–34.
8. Majeux N, Scarsi M, Apostokalis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. Proteins 1999;37:88–105.

9. Majeux N, Scarsi M, Caflisch A. Efficient electrostatic solvation model for protein-fragment docking. Proteins 2001;42:256–268.
10. Sirockin F, Sich C, Improta S, Schaefer M, Saudek V, Froloff N, Karplus M, Dejaegere A. Structure activity relationship by NMR and by computer: a comparative study. J Am Chem Soc 2002;124: 11073–11084.
11. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr Biol Crystallogr 2002;58:899–907.
12. Baker EN, Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Hodgkin DM, Hubbard RE, Isaacs NW, Reynolds CD, et al. The structure of 2Zn pig insulin crystals at 1.5 Å resolution. Philos Trans R Soc Lond B Biol Sci 1988;319:369–456.
13. Brunger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. Proteins 1988;4:148–156.
14. Brooks BR, Bruccoleri R, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. J Comput Chem 1983;4: 187–217.
15. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem 1998; B102:3586–3616.
16. Caflisch A, Schramm HJ, Karplus M. Design of dimerization inhibitors of HIV-1 aspartic proteinase: a computer-based combinatorial approach. J Comput Aided Mol Des 2000;14:161–179.
17. Schaefer M, Zoete V, Karplus M. Fully automated, efficient procedure for predicting the binding mode of small ligands. 2004. Submitted for publication.
18. Zoete V, Michielin O, Karplus M. Protein-ligand binding free energy estimation using continuum electrostatics. J Comp-Aided Mol Des 2004. In press.
19. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz-J, Gilson MK, Bagheri B, Scott LR, McCammon JA. Electrostatics and diffusion in solution: simulations with the University of Houston Brownian Dynamics program. Comp Phys Comm 1995;91:57–95.
20. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand–receptor complexes. J Med Chem 1995;38:4953–4967.
21. Brooks III CL, Karplus M, Pettitt BM. Protein. A theoretical perspective of dynamics, structure, and thermodynamics, volume 71 of Advances in chemical physics. New York: Wiley-Interscience; 1988. p 38–44.
22. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
23. Brooks III CL, Karplus M. Solvent effects on protein motion and protein effects on solvent motion. J Mol Biol 1989;208:159–181.
24. Brooks III CL, Karplus M. Deformable stochastic boundaries in molecular dynamics. J Chem Phys 1983;79:6312–6325.
25. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion if a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–341.
26. Tidor B, Karplus M. The contribution of vibrational entropy to molecular association. J Mol Biol 1994;238:405–414.
27. Karplus M, Ichiye T. Comment on a "fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations." J Mol Biol 1996;263:120–122.
28. Zhou Y, Cook M, Karplus M. Protein motions at zero angular momentum: the importance of long range correlations. Biophys J 2000;79:2902–2908.
29. Harte WE, Swaminathan S, Beveridge DL. Molecular dynamics of HIV-1 protease. Proteins 1992;13:175–194.
30. Brooks BR, Janezic D, Karplus M. Harmonic analysis of large systems. I. Methodology. J Comput Chem 1995;16:1522–1542.
31. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.