Does Conformational Free Energy Distinguish Loop Conformations in Proteins?

Jean-Luc Pellequer and Shu-wen W. Chen
Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032 USA

ABSTRACT Limitations in protein homology modeling often arise from the inability to adequately model loops. In this paper we focus on the selection of loop conformations. We present a complete computational treatment that allows the screening of loop conformations to identify those that best fit a molecular model. The stability of a loop in a protein is evaluated via computations of conformational free energies in solution, i.e., the free energy difference between the reference structure and the modeled one. A thermodynamic cycle is used for calculation of the conformational free energy, in which the total free energy of the reference state (i.e., gas phase) is the CHARMm potential energy. The electrostatic contribution of the solvation free energy is obtained from solving the finite-difference Poisson-Boltzmann equation. The nonpolar contribution is based on a surface area-based expression. We applied this computational scheme to a simple but well-characterized system, the antibody hypervariable loop (complementarity-determining region, CDR). Instead of creating loop conformations, we generated a database of loops extracted from high-resolution crystal structures of proteins, which display geometrical similarities with antibody CDRs. We inserted loops from our database into a framework of an antibody; then we calculated the conformational free energies of each loop. Results show that we successfully identified loops with a "reference-like" CDR geometry, with the lowest conformational free energy in gas phase only. Surprisingly, the solvation energy term plays a confusing role, sometimes discriminating "reference-like" CDR geometry and many times allowing "non-reference-like" conformations to have the lowest conformational free energies (for short loops). Most "reference-like" loop conformations are separated from others by a gap in the gas phase conformational free energy scale. Naturally, loops from antibody molecules are found to be the best models for long CDRs (≥6 residues), mainly because of a better packing of backbone atoms into the framework of the antibody model.

INTRODUCTION

Homology modeling is a powerful approach that allow us to obtain atomic details of macromolecules simply by using the available structural information in databases. The idea is to relate a molecule with unknown three-dimensional structure to at least one with a known structure by using single or multiple sequence alignments (Chothia and Lesk, 1986). Typically, a percentage higher than 30% of identical residues implies strong structural similarities between two proteins (Hubbard and Blundell, 1987). However, even for highly homologous molecules, there are regions that have unique conformations, and special care is required. Generally such regions are localized in loops on the surface of a protein or in fragment deletions or insertions. Correctly modeling such regions is the most challenging aspect of protein modeling. Several techniques have been developed to tackle this problem, including ab initio (Bruccoleri et al., 1988; Bruccoleri and Karplus, 1987; Fine et al., 1986; Moult and James, 1986; Shenkin et al., 1987) and knowledge-based approaches (Chothia and Lesk, 1987; de la Paz et al., 1986; Feldmann et al., 1981; Jones and Thirup, 1986; Koehl and Delarue, 1995; Mainhart et al., 1984; Martin et al., 1989; Summers and Karplus, 1990). The common point of these approaches is that both require a criterion for identifying when a loop conformation is acceptable or not. Usually this criterion is a minimum in a more or less complex potential energy function.

In this paper we report results concerning the development of a complete physical treatment that allows the screening of loop conformations to identify the most suitable ones for a particular protein model. We established a formalism that allows the computation of the conformational free energies of loops by combining a molecular mechanic treatment of a loop with a continuum treatment of the solvent (Smith and Honig, 1994). To illustrate our method, we decided to apply it to a family of molecules with a very conserved core domain but a large heterogeneity in loop conformations. For this reason we selected antibody molecules that are known to have an overall sequence identity above 60% between members and hypervariable loops that are responsible for a specific recognition of epitopes in antigens. Instead of building loops, we took advantage of a recent contribution showing that complementarity-determining region (CDR) conformations are not specific for antibodies, but can be found in a variety of nonantibody molecules (Tramontano and Lesk, 1992). This led us to build a database of loops that exhibit CDR geometry from known three-dimensional antibody structures. We

Received for publication 30 December 1996 and in final form 11 August 1997.

Address reprint requests to Dr. Jean-Luc Pellequer, Department of Molecular Biology, MB4, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037. Tel.: 619-784-8572; Fax: 619-784-2289; E-mail: pelleque@scripps.edu.

Dr. Chen's present address is Department of Molecular Biology, MB4, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037.

0006-3495/97/11/2359/17 $2.00

TABLE 1 X-ray crystal structure of Fab molecule used in our antibody database

Codes		Resolution	CDR from light chain*						CDR from heavy chain*						
PDB	Antibody	(in Å)	L1[#§]	Length	L2[§]	Length	L3[§]	Length	H1[§]	Length	H2[§]	Length	H3[§]	Length	Source
1FAI	R19.9	2.7	A25–Y32	8	Y50–S56	7	G91–R96	6	S31–G33	3	P52A –G55	4	F96–Y100F	11	Lacombe et al., 1992
1FDL	D1.3 cplx	2.5	S26–Y32	7	Y50–D56	7	F91–T97	7	G31–G33	3	W52–N56	5	R96–R99	4	Fischmann et al., 1991
1REI	REI	2	S26–Y32	7	E50–A56	7	Y91–T97	7	—		—		—		Epp et al., 1975
2FB4	Kol	1.9	G25–T32	10	R50–S56	7	W91–N96	6	S31–A33	3	W52–D56	6	G97–F100G	12	Marquart et al. 1980
2FBJ	J539	1.95	A25–S30	6	E50–S56	7	T91–L94	4	K31–W33	3	H52–T56	6	H96–G99	4	Bhat et al., manuscript in preparation
2HFL	HyHEL5	2.54	A25–Y31	7	D50–S56	7	W91–P95	5	D31–W33	3	P52A –G55	4	G95–G102	7	Sheriff et al., 1987
2IGF	B13I2	2.8	N26–Y32	12	K50–S56	7	G91–P96	6	S30–A33	4	S52–Y56	6	S96–Y100A	6	Stanfield et al., 1990
2MCP	McPC603	3.1	S26–F32	13	G50–S56	7	H92–L96	5	D31–Y33	3	S51–T56	10	Y97–W100A	5	Padlan et al., 1985
2RHE	RHE	1.6	A26–S32	8	Y50–P56	7	D93–D96	4	—		—		—		Furey et al., 1983
3FAB	New	2	G25–V33	12	—		Y91–V97	7	D31–W34	4	F52–T56	5	L96–I100A	6	Saul et al., 1978
3HFM	HyHEL10	3	S26–N32	7	Y50–S56	7	S91–T97	9	S31–Y33	3	S52–S56	5	W95–W103	6	Padlan et al., 1989
4FAB	4-4-20	2.7	S25–Y32	13	K50–S56	7	S91–W96	6	S30–W33	4	R52–E56	8	Y96–Y102	6	Herron et al., 1989
6FAB	36-71	1.9	S26–F32	7	F50–S56	7	G91–R96	6	S31–G33	3	N52–Y56	6	Y97–S100A	4	Strong et al., 1991
R45	R45-45-11	2.65	S26–Y32	7	Y50–S56	7	G91–P96	6	D31–Y33	3	S52–S56	6	T100 –G100C	4	Altschuh et al., 1992
			26–32	7	50–52	3	91–96	6	26–32	7	52–56	5	96–101	6	Chothia and Lesk, 1987
			24–34	11–17	50–56	7	89–97	9–15	31–35	5	50–65	16–19	95–102	8–16	Kabat et al., 1991[¶]

*Definition of CDR according to this paper in Kabat's numbering.

[#]Letter corresponds to the amino acid one-letter code.

[§]Definition of CDR according to Chothia et al. (1987).

[¶]Definition of CDR according to Kabat et al. (1991). The variation in length is due to additional inserted residues: CDRL1 (27A → 27F), CDRL3 (95A → 95F), CDRH2 (52A → 52C), and CDRH3 (100A → 100H).

simulate a modeling study by removing the three light chain (Vl) and three heavy chain (Vh) CDRs, one at a time, from a recently solved crystal structure of an antibody in a bound conformation, namely Fab R45-45-11 (R45) (Altschuh et al., 1992; Vix et al., 1993), then replacing loops from our database and calculating the conformational free energies for each conformation.

In the following sections we report the construction of our loop database and assess the conformational variability of loops therein. Then we describe results concerning the conditions of our simulation and the computation of the conformational free energies of loops inserted into the R45 framework. Our results reveal that loops in the database with the lowest conformational energy are the loops with the smallest root mean square deviation (RMSD) compared to CDRs of R45.

MATERIALS AND METHODS

F_v database

Only F_v domains are considered in our calculation. Table 1 lists detailed information about 13 F_v molecules used in our study, including R45 from the database available at the beginning of this work (release Sept 1993; Bernstein et al., 1977). Amino acids of each F_v molecule have Kabat's numbering (Kabat et al., 1991). The F_v database allows us to geometrically describe CDRs from known antibody structures.

Protein data bank files used to generate the loop database

We selected 162 x-ray crystal structures determined at 2.5-Å resolution or better from the Protein Data Bank (PDB). Sequences that have the same name or length or are members of the same superfamilies (cytochromes, ribonucleases, hemoglobins, dihydrofolate reductases, lysozymes, and proteinase acids) were checked for sequence homology by the GCG package programs (Devereux et al., 1984). Except for antibody molecules, only proteins exhibiting a sequence identity below 20% were selected.

Resolution less than 2 Å

1AAJ, 1ACX, 1AK3, 1AKE, 1ALC, 1ALD, 1COX, 1CPC, 1CRN, 1CSE, 1CTF, 1ECA, 1FKF, 1GCR, 1GD1, 1GOX, 1GP1, 1GPB, 1HIP, 1HOE, 1IFB, 1LZ1, 1MBD, 1OMD, 1OVA, 1PAL, 1PAZ, 1PCY, 1PGX, 1R69, 1RBP, 1REI, 1RNB, 1RNH, 1RNS, 1SAR, 1SN3, 1TON, 1UBQ, 1UTG, 1XY1, 256B, 2ACT, 2ALP, 2AZA, 2CBA, 2CCY, 2CDV, 2CGA, 2CI2, 2CNA, 2CPP, 2CSC, 2CYP, 2FB4, 2FBJ, 2FCR, 2GBP, 2HAD, 2LH1, 2LTN, 2MCM, 2MHR, 2MLT, 2OVO, 2PAB, 2PRK, 2PTN, 2RHE, 2RSP, 2SCP, 2SGA, 2SNS, 2SOD, 2WRP, 3APP, 3B5C, 3C2C, 3CBH, 3CHY, 3CLA, 3EBX, 3EST, 3FAB, 3FGF, 3GRS, 3IL8, 3MCG, 3PSG, 3RP2, 3SGB, 451C, 4BLM, 4BP2, 4ENL, 4FD1, 4HHB, 4I1B, 4ICB, 4INS, 4XIS, 5CPA, 5CYT, 5HVP, 5P21, 5PTI, 5RUB, 5RXN, 5TIM, 5TNC, 6FAB, 6LDH, 6TMN, 7AAT, 7RSA, 8ABP, 8ACN, 8DFR, 9PAP, 9RNT, 9WGA.

Resolution higher than 2 Å but less than 2.5 Å (chain)

1ADA, 1AVR, 1CBP, 1CC5, 1CDT(A), 1COL(A), 1FCB(A), 1FHA, 1FNR, 1HRH(A), 1HSA(A, B), 1LE4, 1LIG, 1LRD, 1MSB(A), 1MVP(A), 1NN2, 1PFK(A), 1PHH, 1PRC(C, L, M), 1RHD, 1TIE, 1TPK(A), 1WSY(A, B), 2AAA, 2ABX(A), 2CRO, 2GN5, 2LIV, 2STV, 2TS1, 3CD4(A), 3GAP(A), 3PGK, 3TMS, 4MDH(A), 5FBP(A), 7ICD, 8ADH, 8ATC(A, B), 8CAT(A).

Complementarity-determining regions

Complementarity-determining regions (CDRs) were originally defined based on sequence variability (Kabat et al., 1977). Furthermore, Chothia and Lesk (1987) pointed out that CDRs may be considered from a structural point of view by selecting residues that have high RMSD from antibody superimposition. However, this definition was not quantitative and has been established with only a limited number of antibodies. In this study we need to extract geometrical properties from antibody CDRs to screen the PDB. Consequently, a rigorous and objective method is presented here for delimiting CDR in antibodies.

We consider a CDR as a loop characterized by the absence of secondary structure. To define the length in residues of each CDR, we applied several screening parameters. The starting length of a CDR is determined by the standard definition described by Kabat et al. (1991). First, residues located at either the N- or C-terminus of a CDR were excluded if they were in β-sheet conformation, as determined by the program DSSP (Kabsch and Sander, 1983). Second, residues at the extremities of a CDR were excluded if they had a null (<1 Å^2) accessible surface area, as determined by the program SURFV with a probe radius of 1.4 Å (Sridharan et al., 1992). Third, the minimum length of a CDR is three residues, or four for hairpin loops (as defined by DSSP). The final demarcation of each CDR from all 13 known F_v structures is presented in Table 1. For the sake of comparison, CDR definitions according to Kabat and Chothia are added to this table. It can be observed that our definitions are very similar to those of Chothia and Lesk (1987), except those for CDRL2 and CDRH1. CDRL1, CDRL2 and CDRL3 denote CDR 1, 2, and 3 of the antibody light chain and CDRH1, CDRH2, CDRH3 denote CDR 1, 2, and 3 of the heavy chain.

Definition of the geometric criteria

Once the residue length of each CDR was determined, its backbone geometry was characterized by three variables: 1) the distance between the N- and C-terminal atoms, 2) linearity, and 3) flatness. The backbone is defined by the following atoms: N, CA, C. The first variable is simply derived from the coordinates of the terminal atoms (N and C). Linearity and flatness are derived from the spatial distribution of atoms in a loop (Ring et al., 1992) from geometric extensions of a loop. We built a modified inertia tensor matrix in which the mass of each atom was set to 1. We obtained the principal axes of the coordinate system by computing the eigenvectors of the inertia tensor matrix. The eigenvalues define the extension of atom distributions along the corresponding axis. The extension along the n axis (E_n) is defined as (Ring et al., 1992)

$$E_n = \frac{1}{N}\sum_{i=1}^{N}|r_{i,n}|, \quad n = x, y, z;$$

where $r_{i,n}$ is the coordinate of atom i in the nth eigenvector's basis and N is the number of atoms. The largest extension (E_n) (smallest eigenvalue) was assigned to the x axis, and the smallest extension (E_n) to the z axis.

Linearity is defined as

$$\frac{E_y}{E_x}$$

and flatness as

$$\frac{\sqrt{2}E_z}{\sqrt{E_x^2 + E_y^2}}$$

Because $E_x \geq E_y \geq E_z$, linearity ranges from 0 (fully extended on the x axis) to 1 (symmetrical arc), and flatness ranges from 0 (fully flat) to 1 (sphere where $E_x = E_y = E_z$). We also used a ϕ-ψ criterion to define α-helices, assuming that α-helices are composed of consecutive ϕ-ψ angles ranging from [−83, −61] to [−43, −21] (Presta and Rose, 1988).

Insertion of a loop from the database into the F_v R45

Replacing loop side chains

Side chains were substituted to match the sequence of the F_v R45. We set up the dihedral angles $\chi 1$ and $\chi 2$ (Table 2) according to the highest probability found in the rotamer library established by Tuffery et al. (1991). Dihedral angles $\chi 3$ and $\chi 4$ were set to 180°. Consequently, each amino acid side chain displays the same starting conformation.

Optimizing the inserted loop

As compared to loop building, the use of a database requires an additional step, which is the insertion of a loop into the antibody framework. For example, we could superimpose residues from the framework on the flanking residues of each loop ($N - 1$ and $C + 1$). However, this would require an additional constraint on the distances between atoms from residues $N - 1$ and $C + 1$ during the building of the database. Such a constraint is inconsistent because of our definition of antibody CDRs. Indeed, this constraint could have been adopted only for loops in which residues $N - 1$ and $C + 1$ were in β-conformations in their original molecule (see Materials and Methods for definitions of CDR in this study). Moreover, a recent study concluded that superimposing flanking peptides of loops on an antibody framework failed to provide an accurate model for CDRs (Tramontano and Lesk, 1992). An alternative method is a docking procedure (e.g., Carlacci and Englander, 1993). However, this method is computationally costly and introduces an additional variable, which is the identification of the best docking orientation. Although we could have used the conformational free energy calculation to identify this best docking orientation, we found experimentally that it is more appropriate to start with a simple insertion protocol, such as a least-squares superimposition of the backbone atoms of each loop on the loop from the reference crystal structure (N, CA, C). A similar approach was used, but instead of a direct superimposition, each loop was anchored at the extremity of the reference

TABLE 2 Original and rotamer dihedral angles of light chain CDRs from R45

		$\chi 1$*		$\chi 2$*	
		Fab R45	Rotamer#	Fab R45	Rotamer#
CDRL1	Ser^{26}	58.81	65	—	—
	Gln^{27}	166.61	−63	−164.22	178
	Asp^{28}	−171.73	−70	132.32	−32
	Ile^{29}	59.22	−62	168.06	163
	Ser^{30}	−63.97	65	—	—
	Thr^{31}	−83.74	−61	—	—
	Tyr^{32}	−46.51	−64	−46.33	102
CDRL2	Tyr^{50}	−144.37	−64	−23.9	102
	Thr^{51}	−155.63	−61	—	—
	Ser^{52}	65.03	65	—	—
	Arg^{53}	−168.13	−176	−176.31	156
	Leu^{54}	−74.7	−62	62.12	170
	Arg^{55}	100.98	−176	110.83	156
	Ser^{56}	−54.98	65	—	—
CDRL3	Gly^{91}	—	—	—	—
	Ser^{92}	−176.11	65	—	—
	Arg^{93}	−55.36	−176	−60.19	156
	Ile^{94}	−60.92	−62	176.79	163
	Pro^{95}	35.86	27§	−45.97	−29§
	Pro^{96}	25.61	27§	−42.22	−29§
CDRH1	Asp^{31}	−53.0	−70	−20.6	−32
	Tyr^{32}	−73.1	−64	79	102
	Tyr^{33}	−90.5	−64	57.7	102
CDRH2	Ser^{52}	43.5	65	—	—
	Asn^{52A}	62.3	−72	−45.2	−44
	Gly^{53}	—	—	—	—
	Gly^{54}	—	—	—	—
	Gly^{55}	—	—	—	—
	Ser^{56}	−54.9	65	—	—
CDRH3	Thr^{100}	−65.7	−61	—	—
	Leu^{100A}	−144.9	−62	58.0	170
	Tyr^{100B}	−56.7	−64	−24.9	102
	Gly^{100C}	—	—	—	—

*Only $\chi 1$ and $\chi 2$ were modified according to the rotamer library. $\chi 3$ and $\chi 4$, when present, were set to 180°.
#Rotamer library from Tuffery et al. (1991).
§Values for Pro were obtained from Ponder and Richards (1987).

loop and then rotated around the plane of a loop to minimize the distance between the two coplanar centers of mass (Martin et al., 1989). Our method is not totally unrealistic, because in the modeling of antibodies one begins with a template model from the PDB in which CDR loops are present. Remodeling a CDR by superimposing a loop from a database onto the existing CDR template is reasonable, as long as they have the same number of residues.

Explicit hydrogen atoms were built on each loop when they were inserted into the F_v R45. We used the HBUILD command (Brünger and Karplus, 1988) provided with the Xplor program (Version 3.0) (Brünger, 1992). The Xplor topology and parameter files were, respectively, TOPALLH6x.PRO and PARMALL3x.PRO. Sixty cycles of conjugate gradient minimization (Powell method) were then carried out while fixing all atoms of the F_v and all heavy atoms of the loop. We used a dielectric constant of 1 and a nonbonded cut-off of 9 Å. For both the VdW switch and electrostatic shift functions, the cut-on and cut-off were, respectively, 6.5 Å and 8 Å. The total energy used in this evaluation includes bond distance, valence angle, dihedral and improper angle, van der Waals, and electrostatic energies. We then carried out two types of optimization (including hydrogen and heavy atoms): 1) side chains only and 2) all atoms:

1. Side-chain conformations were minimized by 600 cycles (for light chain CDRs) and 300 cycles (for heavy chain CDRs) of conjugate gradient minimization (Powell method) and saved. We observed that 600/300 cycles of minimization allows convergence in a reasonable time with respect to the length of the CDRs.

2. Starting from the conformations in the type 1 optimization, we applied 600 cycles (for light chain CDRs) and 300 cycles (for heavy chain CDRs) of conjugate gradient minimization to all atoms of the loop.

Optimization of the loop closure

For both minimization procedures described above, the loop closure occurs as follows. The peptide bond atoms located at both extremities of a loop were allowed to move during minimization (even when only side chains were optimized). Only four atoms per extremity were needed for loop closure because of the restricted distances between *N* and *C* termini in the establishment of the database (Table 3).

F_v R45 reference conformations

Starting from the x-ray crystal structure of F_v R45-45-11, four "reference" structures were generated: 1) side chains minimized; 2) all atoms minimized; 3) same as 1), and 4) same as 2), but using rotamer dihedral angle values for side chains (as in Table 2). Minimizations were performed as described above for loops from the database. In our study, loops with side chains minimized were only compared with the reference structures with side chains minimized, and similarly, loops with all atoms minimized were compared with the corresponding all-atom minimized reference structures.

Relative stability of loops

We assessed the stability of loops by evaluating their relative conformational free energy compared with CDRs from F_v R45. The conformational free energy in solution can be described in terms of a thermodynamic cycle and Eq. 1, where gas, sol, ref, mod, solv, conf, ele, and np are, respectively, the abbreviations for gas phase, solution, reference, model, solvation, conformation, electrostatic, and nonpolar.

$$\begin{array}{ccc} \mathrm{Fab}_{\mathrm{ref}}^{\mathrm{gas}} & \xrightarrow{\Delta G_{\mathrm{conf}}^{\mathrm{gas}}} & \mathrm{Fab}_{\mathrm{mod}}^{\mathrm{gas}} \\ \Delta G_{\mathrm{ref}}^{\mathrm{solv}} \downarrow & & \downarrow \Delta G_{\mathrm{mod}}^{\mathrm{solv}} \\ \mathrm{Fab}_{\mathrm{ref}}^{\mathrm{sol}} & \xrightarrow[\Delta G_{\mathrm{conf}}^{\mathrm{sol}}]{} & \mathrm{Fab}_{\mathrm{mod}}^{\mathrm{sol}} \end{array}$$

This thermodynamic cycle does not require knowledge of the correct loop conformation, because we only calculate a relative stability. Consequently, any loop from the database could have been chosen to be the "reference loop."

$$\Delta G_{\mathrm{conf}}^{\mathrm{sol}} = \Delta G_{\mathrm{conf}}^{\mathrm{gas}} + \Delta\Delta G^{\mathrm{solv}} \quad (1)$$

The conformational free energy in the gas phase is obtained from a molecular force field (CHARMm): it includes internal coordinate energies as well as nonbonded interactions (van der Waals and electrostatic). The solvation free energy for transferring a molecule from a gas phase to an aqueous phase is calculated with a continuum model of the solvent (Jean-Charles et al., 1991; Honig et al., 1993). Addition of the gas phase conformational free energy to the solvation free energy gives the conformational free energy in solution. It should be stressed that in this thermodynamic cycle, there are double counting interactions due to the combination of a continuum model and a molecular force field (Smith and Honig, 1994).

The solvation free energy difference between modeled and reference loops is

$$\Delta\Delta G^{\mathrm{solv}} = \Delta G_{\mathrm{mod}}^{\mathrm{solv}} - \Delta G_{\mathrm{ref}}^{\mathrm{solv}} \quad (2)$$

The solvation free energy change for a model loop (mod) or the reference loop (ref) can be decomposed approximately into electrostatic and nonpolar contributions as follows:

$$\Delta G^{\mathrm{solv}} = \Delta G_{\mathrm{ele}}^{\mathrm{solv}} + \Delta G_{\mathrm{np}}^{\mathrm{solv}} \quad (3)$$

where $\Delta G_{\mathrm{ele}}^{\mathrm{solv}}$ is the difference in electrostatic free energy of transferring F_v from gas to water, obtained from finite-difference Poisson-Boltzmann calculations (Delphi Version 3.0; Nicholls and Honig, 1991; Sridharan et al., manuscript in preparation), which is the difference between the reaction field energies in vacuum and in water (Gilson and Honig, 1988; Jean-Charles et al., 1991). $\Delta G_{\mathrm{np}}^{\mathrm{solv}}$ is the transfer free energy of an uncharged molecule of the same size and shape as the F_v from gas to water. It is commonly assumed that $\Delta G_{\mathrm{np}}^{\mathrm{solv}}$ is proportional to the total accessible surface area of the F_v:

$$\Delta G_{\mathrm{np}}^{\mathrm{solv}} = \gamma A_{\mathrm{T}} \quad (4)$$

where γ is the vacuum-to-water transfer free energy coefficient. In our study we used a value of 5 cal/mol/Å^2, as determined from solubility experiments (Ben-Naim and Marcus, 1984).

TABLE 3 Summary of geometrical properties of CDRs

	Light chain CDRs			Heavy chain CDRs		
	L1*	L2	L3	H1	H2	H3
Distance (Nt-Ct)#	11.40 ± 0.84	13.44 ± 0.70	6.40 ± 1.10	7.90 ± 0.38	4.27 ± 0.96	4.75 ± 1.74
Linearity§	0.404 ± 0.098	0.309 ± 0.038	0.675 ± 0.168	0.218 ± 0.070	0.718 ± 0.110	0.707 ± 0.242
Flatness§	0.196 ± 0.050	0.178 ± 0.028	0.221 ± 0.082	0.211 ± 0.070	0.294 ± 0.052	0.259 ± 0.194

*Average ± standard deviation.

#Distances between N- and C-termini are in Å.

§Linearity and flatness are unitless (see Materials and Methods).

The reaction field energy in vacuum and in water was calculated with the Delphi program, using a 129 cube grid size, three focusing runs per calculation (24%, 48%, and 96%), and a dipolar boundary condition for the first run. The final resolution was two grid points per angstrom, which has been shown to be sufficient for convergence. At such a resolution, the relative energy is almost insensitive to the orientation of the molecule inside the grid (Smith and Honig, 1994). The internal dielectric constant was 2 and the external dielectric was 80 for water. Although small variations in the solvation free energy are observed (in the range of a few kcal/mol with higher internal dielectric constants; unpublished data), we adopted the value of 2 because it is the value used during the development of atom radii and charges in the newly derived PARSE parameters that have been optimized for accurate reproduction of the hydration free energy of amino acids upon transfer from gas phase to water (Sitkoff et al., 1994). These PARSE parameters allow an assignment of particular values for N- and C-terminal residues as well as for disulfide bridges. Only Asp, Glu, Lys, and Arg were charged. In the gas phase calculation, the external dielectric constant was 1.

RESULTS

Loop database

Table 4 shows the number of loops bearing a backbone geometry similar (as defined by our three criteria; see Materials and Methods) to that of crystallographically determined antibody CDRs, like those in Table 1. Six main databases are constructed, one for each CDR. Subdatabases are created for each loop length encountered in CDRs (Table 1). Each subdatabase is built independently. As the first step for a particular subgroup, we extracted fragments from the PDB whose distance between the N- and C-termini was in the range shown in Table 3. This resulted in a fast extraction of 75,696 loops (54,328 for V_l and 21,368 for V_h). These extracted loops were screened by applying the linearity and flatness criteria. Loops that had ϕ-ψ angles characteristic of a regular α-helix conformation for at least two consecutive residues were eliminated. The final database consists of 14,558 loops (5510 for V_l and 9048 for V_h). The number of extracted loops varies with each CDR, even those of identical length. CDRL1 and CDRH3 have a large number of subgroups representing the structural diversity of these CDRs in antibodies. CDRL2 is fully conserved in length, whereas CDRL3 shows a little variation in length and consists only of short loops (below seven residues). CDRH1 was found to have two subdatabases (Table 1). However, based on the number of accepted loops with four residues, this particular subdatabase is not statistically significant. CDRH2 shows a large variation in length; however, only a few loops were extracted and accepted. The power of screening is expressed as the percentage loop accepted/loop extracted (Table 4). This ratio shows the efficiency of the first screening using our first geometric criterion, the distance between the N- and C-terminal atoms. It is interesting to note that the ratio decreases when the length of CDRs increases (CDRL1, CDRH2, CDRH3). It indicates that selecting loop conformation based on a distance criterion is less efficient for long loops and that the additional screening steps presented here (linearity and flatness) are necessary.

TABLE 4 Summary of the contents of our loop database

	Length in residues											
	3	4	5	6	7	8	9	10	11	12	13	Total
CDRL1												
Loop extracted*				4917	9744	4646	3057	2602	2380	2196	2083	31625
Loop accepted#				974	406	249	232	169	172	113	104	2419
Ratio§				19.8	4.2	5.4	7.6	6.5	7.2	5.2	5.0	7.7
CDRL2												
Loop extracted					3866							3866
Loop accepted					196							196
Ratio					5.1							5.1
CDRL3												
Loop extracted		12819	2725	1863	1430							18837
Loop accepted		1268	702	497	428							2895
Ratio		9.9	25.8	26.7	29.9							15.4
CDRH1												
Loop extracted	4318	1991										6309
Loop accepted	2205	2										2207
Ratio	51.1	0.1										35.0
CDRH2												
Loop extracted			395	251	143	216	216	153				1374
Loop accepted			121	62	41	71	38	14				347
Ratio			30.6	24.7	28.7	32.9	17.6	9.2				25.3
CDRH3												
Loop extracted		6724	1551	1096	910	723	731	653	671	626		13685
Loop accepted		1920	998	737	710	557	562	456	333	221		6494
Ratio		28.6	64.4	67.2	78.0	77.0	76.9	69.8	49.6	35.3		47.5

*Number of loop selected according to the distance criterion.

#Number of extracted loop that passed through the screening of linearity and flatness criteria.

§Corresponds to the number of accepted loops divided by the number of extracted loops times 100.

Modeling side chains with a rotamer library

It was observed that the dihedral angles (χ angles) of side chains in proteins located in the PDB adopt only a few possible conformations (called rotamers). These preferred side-chain conformations have been grouped into rotamer libraries (Ponder and Richards, 1987; Tuffery et al., 1991; Dunbrack and Karplus, 1993). Because of the extremely high combinatorial possibilities of dihedral angles for a seven-residue loop, a complete systematic search is not possible. We therefore decided to use only one configuration per side chain. We selected the combination of χ angles with the highest probability of occurrence found in the rotamer library of Tuffery et al. (1991). The main reason for assigning only one rotamer conformation per side chain is the assumption that CDR loops are highly exposed to solvent and have no external constraints. Indeed, for light chain CDRs, during our minimization procedures (both side chains and all-atom optimization), an average of 65% of the side chains kept their original rotamer configuration (Table 2). Twenty percent of side chains adopted another highly probable rotamer configuration, and ~15% adopted an unusual (nonrotamer) configuration. In Table 2 we show side-chain dihedral angles from the crystal structure of R45 with the ones from rotamer libraries. It can be seen that rotamer values do not systematically match side-chain dihedral angle values from the crystal structure. Thus rotamer side-chain dihedral angle values cannot be responsible for the low RMSDs observed in this work.

Geometric comparison between modeled and reference loop

We calculated the root mean square deviation (RMSD) between a modeled loop and the corresponding reference loop by superimposing the corresponding backbone atoms (N, CA, C), even when we compute the RMSD values for all heavy atoms. As described in Materials and Methods, each superimposition was consistent, meaning that side-chain or all-atom minimized CDRs are compared with reference side-chain or all-atom minimized CDRs, respectively.

The superimposition of every loop from our database on the respective reference CDR from R45 gives an estimate of the structural diversity encountered in our database. The RMSD for backbone atoms varied from 0.19 to 3.78 Å for V_l and from 0.05 to 1.92 Å for V_h, whereas RMSD for all heavy atoms varies from 0.81 to 5.02 Å for V_l and from 0.16 to 8.12 Å for V_h when only side chains were minimized, for a total of 1099 loops for V_l and 4187 for V_h. To estimate the effect on the conformation due to side-chain dihedral angles (rotamer conformations), we minimized the reference CDRs in which crystallographic side-chain conformations were replaced by rotamer conformations similar to those of loops from the database. The RMSD for the same set of 1099 loops for V_l and 4187 for V_h, for all heavy atoms, with such a reference minimized CDR, ranges from 0.7 to 5.0 Å for V_l and from 0.7 to 7.9 Å for V_h, which is similar to values of reference CDR side chains in their crystal structures (0.81 to 5.02 Å for V_l and 0.16 to 8.16 Å for V_h). The RMSD values reported here attest to the fact that loops in our database have diverse conformations, indicating that we sampled a reasonably large conformational space. In the subsequent paragraphs, we give RMSD values for all heavy atoms. However, for the purpose of comparison with studies in the literature, RMSD values for backbone atoms (N, CA, C) are presented in the last RMSD column of Tables 5–10. However, care should be taken in comparing published RMSDs, because the superimposition schemes used for calculating these values may be different.

Analysis of the stability of loops inserted into the F_v

Two ways of optimizing CDR modeled loops were tested: 1) only the side chains of the loops were minimized, and 2) all atoms of loops were minimized. In the first case, the backbone was kept fixed in the original loop conformation. In the second case, all atoms were minimized to obtain a "clash-free" loop conformation. To reduce computational requirements, only a subset of the lowest conformational free energy loops (<1500 kcal/mol) from 1) were optimized in 2) for light chain CDRs, whereas all loops were optimized for heavy chain CDRs (shorter length). Loop closure occurs identically in both approaches and therefore does not contribute to the energy difference observed between them. Modeled loops were compared with the reference loops treated with identical minimization procedures.

CDRL1: side-chain optimization only (406 loops)

The range in gas phase conformational free energy is wide: 33 loops have energies lower than 1000 kcal/mol. The main high-energy component came from steric clashes (vdW energy) among backbone atoms or between side chains that could not be fully optimized because of the fixed backbone orientation. The loop with the lowest conformational free energy (in gas phase and in solution) has the lowest RMSD (Table 5 *A*). This loop was extracted from an antibody molecule (6fab, residues L26_L32). We noticed a gap (more than 100 kcal/mol) in the conformational gas phase free energy between the lowest energy loop (6fab.L26_L32) and the one with the next best energy. Because of the relatively high conformational free energy in the gas phase, the solvation free energy term did not change the ordering of the loops.

CDRL1: all-atom optimization (40 loops)

As expected for such an optimization procedure, most of the steric clashes are removed and the conformational free energy was greatly reduced, ranging from 7.9 to 488.37 kcal/mol for the 40 selected loops. The best conformation observed when only side chains were minimized is also the

TABLE 5 Five lowest conformational free energy loops for CDRL1

	Loop*	ΔG^{sol}_{conf} (kcal/mol)	ΔG^{gas}_{conf} (kcal/mol)	RMSD (Å)#	RMSD (Å)§	Origin of the loop¶
A. CDRL1 Side chains minimized	R45 CDRL1	0	0	0	0	Antibody R45-45-11
	6fab.L26_L32	135.50	161.23	1.31	0.18	Antibody 36-71
	1msb.A73_A79	275.16	265.87	3.43	1.52	Mannose binding protein A
	3tms.64_70	380.96	387.21	3.61	1.47	Thymidylate synthase
	2aaa.453_459	425.98	427.11	2.81	1.32	Alpha amylase
	1cox.47_53	460.32	440.38	3.26	1.66	Cholesterol oxidase
B. CDRL1 All atoms minimized	6fab.L26_L32	7.90	35.30	1.41	0.36	Antibody 36-71
	4enl.287_293	33.03	35.68	3.23	1.73	Enolase
	4enl.374_380	34.93	69.54	3.72	1.54	Enolase
	1rei.A26_A32	38.67	60.56	1.53	0.65	Bence-Jones immunoglobulin
	8adh.128_134	51.85	80.86	2.30	1.16	Alcohol dehydrogenase

*Loop names are defined as follows: PDB code of the protein + extension showing the first and last residue of the loop. When several chains are present in one PDB file, the chain letter is the residue number prefix.
#RMSD values between all heavy atoms of loops and minimized reference CDR. Side-chain minimized loops are compared to side-chain minimized reference CDR (reciprocally for all-atom minimized loops).
§RMSD values between backbone atoms of loops and reference CDR from crystal structure (atoms N, CA, C).
¶Origin of the loop as listed in the PDB file.

best in all-atom minimization (Table 5 *B*, 6fab.L26_L32). Backbone atoms of this loop are superimposed on the minimized reference CDRL1 in Fig. 1. Two loops had very close conformational gas phase free energies (6fab.L26_L32 and 4enl.287_293), but 6fab.L26_L32 displayed a lower conformational free energy in solution, indicating the beneficial effect of the solvation free energy when loop stability is assessed. This is a clear case where the addition of solvation free energy to the conformational gas phase free energy allowed the identification of the correct conformation. Furthermore, when the solvation free energy is included, the difference in conformational free energy between the reference and modeled loops is reduced, as shown in Fig. 2 for CDRL1. Similar plots were obtained for CDRL2 and CDRL3 (data not shown). The gap in energy observed when only side chains were optimized is reduced but still present for CDRL1 (Table 5 *B*). There is no modeled CDRL1 loop that had a lower conformational gas phase free energy than the reference CDR loop.

CDRL2: side-chain optimization only (196 loops)

As for CDRL1, the range in gas phase conformational free energy is wide. Only 20 loops had lower energies than 1000 kcal/mol. The main component from such high-energy conformations came from atom clashes (vdW energy). Several loops were found to have "reference-like" CDR geometry (RMSD < 1.5 Å). Although the loop with the lowest conformational free energy did not have the lowest RMSD (Table 6 *A*), its conformation was almost as close to the reference as the best fit structure. The three lowest conformational free energy loops are from antibody molecules (2rhe.51_57, 2fb4.L49_L55, 1rei.A50_A56). The conformational free energies in both the gas phase and solution of 2rhe.51_57 were lower than those calculated for the reference structure. The electrostatic contribution of the two arginine residues is 40 kcal/mol lower for this loop than the reference. We observed a gap (more than 100 kcal/mol) in the conformational gas phase free energy between this lowest energy CDRL2 loop and the consecutive one (Table 6 *A*). However, this gap did not correlate with a gap in the RMSD, because most of the low-energy CDRL2 loops have a "reference-like" conformation.

CDRL2: all-atom optimization (39 loops)

The conformational free energy ranged from −6.24 to 1995.12 kcal/mol for the 39 selected conformations. The

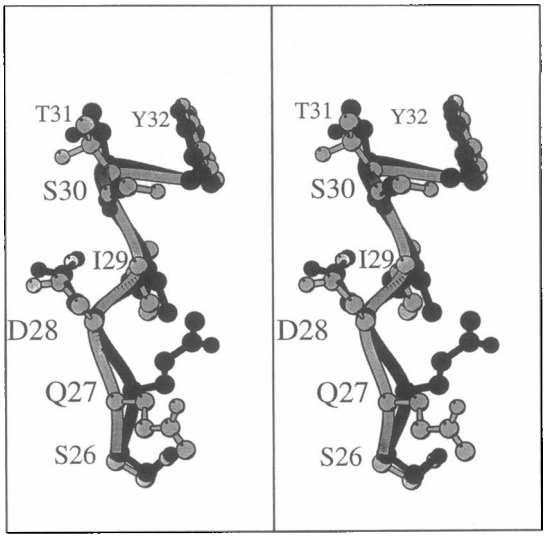

FIGURE 1 Stereo backbone atom superimposition of the lowest conformational energy loop (6fab.L26_L32 *in light gray*, all-atom minimized) on the minimized reference CDRL1 of R45 (*in black*). The RMSD is 0.36 Å for atoms N, CA, C. Plots were drawn with Molscript (Kraulis, 1991).

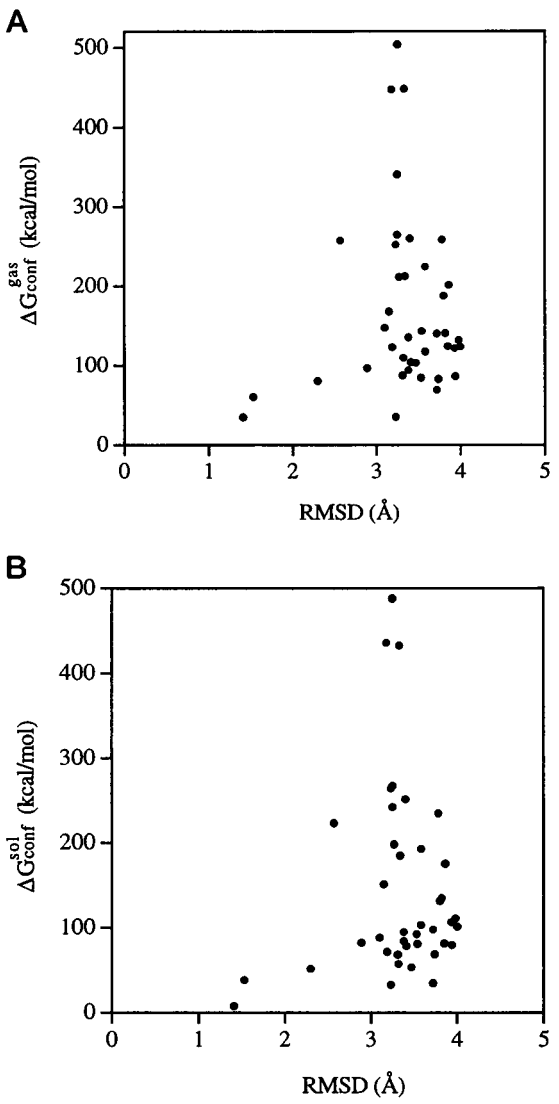

FIGURE 2 Conformational free energy in (*A*) gas phase and (*B*) solution for all-atom minimized loops modeling CDRL1 plotted against heavy atom root mean square deviation (RMSD) from the minimized crystal structure. Similar plots were obtained for CDRL2 and CDRL3 (not shown).

three conformations with the lowest RMSD were the same as those found when only side chains were minimized, but the ordering is changed (Table 6 *B*). The backbone atoms of these three loops are superimposed on the minimized reference CDRL2 in Fig. 3. However, when all atoms are minimized, nonantibody loops are found to have low energy, with an RMSD in the range of that of loops originating from antibody molecules. Inclusion of the solvation free energy allowed a more accurate description of the free energy, although 1rei.A50_A56 still has a lower energy than the reference CDRL2 loop (Table 6 *B*). Gaps in the energy previously observed in Table 6 *A* have decreased in magnitude but are still present. Although nine loops were found to have a "reference-like" CDR geometry (RMSD < 1.5 Å), two subgroups could be distinguished, both by their RMSD and by their conformational free energy in solution. One group has a RMSD < 0.9 Å and ΔG_{conf}^{sol} < 1.5 kcal/mol (three loops), and the second has a RMSD range from 1.13 to 1.39 Å and a ΔG_{conf}^{sol} from 19.35 to 34.39 kcal/mol (mainly nonantibody loops).

CDRL3: side-chain optimization only (497 loops)

Like those of CDRL1 and CDRL2, the range in gas phase conformational free energy is wide: 229 loops had energies lower than 1000 kcal/mol. There is no large gap in energy between the best conformation and the next lowest one. Indeed, the second lowest free energy conformation also has a "reference-like" CDR conformation (RMSD < 1.5 Å; Table 7 *A*). A large gap energy is found between the second loop and the third one, which is not "reference-like" (gap of ~200 kcal/mol). The two loops with the lowest energy are equally close to the reference conformation (Table 7 *A*).

CDRL3: all-atom optimization (47 loops)

The conformational free energy ranged from −2.25 to 443.17 kcal/mol for the 47 selected conformations. The two lowest RMSD conformations observed in all-atom minimization were the same when only side chains were minimized (Fig. 4). Because of the solvation free energy term, the loop 451c.21_26 has conformational free energy close to the two best conformations, but it has non-reference-like CDR conformation. This particularly low solvation free energy is due to increases in the solvent-accessible surface area: 50 $Å^2$ of arginine 93 and 13 $Å^2$ for the polar atoms of backbone, compared with other loops shown in Table 7 *B*. Although the inclusion of solvation free energy did not exclude a wrong conformation, the difference in energy between reference and minimized loops was reconciled. It is interesting that the all-atom minimization improved the conformation of a nonantibody loop toward a "reference-like" CDR geometry (2cba.166_171, Table 7 *B*) with a RMSD of 0.97, compared with 1.69 Å when only side chains were optimized. The improvement of the geometry occurs in the last two prolines of CDRL3. Despite its fourth position in the conformational free energy scale in solution, it occupies the third place when energy is calculated in the gas phase.

CDRH1: side-chain optimization only (2205 loops)

In contrast to light chain CDRs, the range in gas phase energies is narrow: 1062 loops have energies lower than 1000 kcal/mol. This is due mainly to the increase in redundancy in the short loop database. Consequently, there is no gap in gas phase energy, as observed for light chain CDRs. Nevertheless, the optimized reference CDR has the lowest gas phase conformational energy (Table 8 *A*). The next five lowest ones in the energy scale have similar stabilities (<0.5 kcal/mol), and all display a "reference-like" CDR geometry, as judged by the low RMSD (<0.23 Å) for backbone atoms (Table 8 *A*). However, the RMSD for heavy atoms is slightly higher (~1.9 Å; Table 8 *A*) than

TABLE 6 Five lowest conformational free energy loops for CDRL2

	Loop*	ΔG^{sol}_{conf} (kcal/mol)	ΔG^{gas}_{conf} (kcal/mol)	RMSD (Å)#	RMSD (Å)§	Origin of the loop¶
A. CDRL2 Side chains minimized	R45 CDRL2	0	0	0	0	Antibody R45-45-11
	2rhe.51_57	−5.00	−14.13	0.92	0.51	Bence-Jones immunoglobulin
	2fb4.L49_L55	115.92	109.00	0.83	0.27	Antibody Kol
	1rei.A50_A56	122.26	157.01	0.97	0.48	Bence-Jones immunoglobulin
	9wga.A13_A19	292.04	280.9	1.51	0.89	Wheat germ agglutinin
	2fbj.H195_H201	341.22	339.46	1.40	0.61	Antibody J539
B. CDRL2 all atoms minimized	1rei.A50_A56	−6.24	3.10	0.90	0.29	Bence-Jones immunoglobulin
	2rhe.51_57	0.92	1.76	0.85	0.31	Bence-Jones immunoglobulin
	2fb4.L49_L55	1.48	1.30	0.89	0.30	Antibody Kol
	1pcy.22_28	19.35	47.25	1.36	0.39	Plastocyanin
	4mdh.A30_A36	19.52	27.92	1.39	0.37	Malate dehydrogenase

*Loop names are defined as follows: PDB code of the protein + extension showing the first and last residue of the loop. When several chains are present in one PDB file, the chain letter is the residue number prefix.
#RMSD values between all heavy atoms of loops and minimized reference CDR. Side-chain minimized loops are compared to side-chain minimized reference CDR (reciprocally for all-atom minimized loops).
§RMSD values between backbone atoms of loops and reference CDR from crystal structure (atoms N, CA, C).
¶Origin of the loop as listed in the PDB file.

expected for "reference-like" loops (1.5 Å). This is due to an unusual tyrosine side-chain conformer (Fig. 5) observed in minimized reference CDRH1 (around −90°). In this CDR, we observed 464 loops that have the lowest gas phase conformational energies that display a "reference-like" CDR geometry (average backbone RMSD = 0.25 Å and 1.9 Å for all heavy atoms). Because of the large number of loops, we decided to compute solvation free energy only on fully optimized loops.

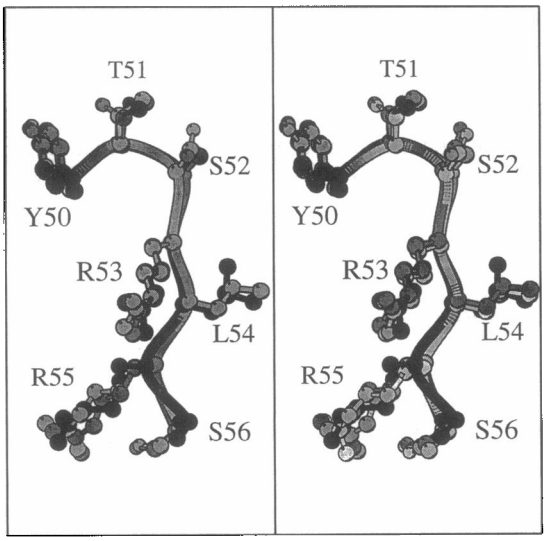

FIGURE 3 Stereo backbone atom superimposition of the three lowest conformational energy loops (1rei.A50_A56, 2fb4.L49_L55, and 2rhe.51_57, all-atom minimized; *lighter gray* corresponds to a higher conformational energy) on the minimized reference CDRL2 of R45 (*in black*). The RMSD values are, respectively, 0.29, 0.30, 0.31 Å for atoms N, CA, C.

CDRH1: all atom optimized (2205 loops)

Because CDRH1 is short (three residues), all loops are fully minimized. We find that 139 loops have a gas phase conformational energy lower than 5 kcal/mol (376 loops have gas phase conformational energies lower than 10 kcal/mol). No conformations are found to have lower conformational energy, either in gas phase or in solution, than the reference CDR. A small energy gap of ~3 kcal/mol is found between the reference CDR and the next loop in the energy scale (both in gas phase and in solution; Table 8 *B*). Although the conformation with the lowest energy does not have the lowest RMSD (Table 8 *B*), it exhibits "reference-like" CDR geometry, as do the next 500 loops, with an average RMSD of 0.14 Å for backbone atoms and 1.85 Å for all heavy atoms.

CDRH2: side-chain optimization only (62 loops)

The results for CDRH2 are similar to those for light chain CDRs, in the sense that only one loop is found to have very low gas phase conformational energy (−3 kcal/mol). This loop (2fbj.H52_H57, Table 9 *A*) has a lower energy than the reference CDR and originates from an antibody molecule. As in light chain CDRs, a large gap in energy is found (~50 kcal/mol) between the lowest energy conformation and the subsequent one.

CDRH2: all-atom optimization (62 loops)

Because of the low number of loops, all of them were fully optimized. Only one loop has a lower gas phase conformational energy than the reference CDR (2fbj.H52_H57, −3.4 kcal/mol), and it displays the lowest RMSD (Table 9 *B*). This loop is also the one that has the lowest conformational energy when only side chains are optimized. The gap in the

TABLE 7 Five lowest conformational free energy loops for CDRL3

	Loop*	ΔG_{conf}^{sol} (kcal/mol)	ΔG_{conf}^{gas} (kcal/mol)	RMSD (Å)#	RMSD (Å)§	Origin of the loop¶
A. CDRL3 Side chains minimized	R45 CDRL3	0	0	0	0	Antibody R45-45-11
	1rei.A91_A96	6.22	13.82	0.92	0.19	Bence-Jones immunoglobulin
	6fab.L91_L96	20.35	19.52	0.81	0.25	Antibody 36-71
	1fcb.A217_A222	208.43	238.45	1.91	0.59	flavocytochrome B2
	2scp.A104_A109	246.00	248.74	1.94	1.20	Calcium-binding protein
	3app.186_191	255.33	239.78	2.05	1.40	Penicillopepsin
B. CDRL3 All atoms minimized	1rei.A91_A96	−2.25	18.74	0.97	0.29	Bence-Jones immunoglobulin
	6fab.L91_L96	0.03	17.59	0.92	0.23	Antibody 36-71
	451c.21_26	2.59	34.66	1.93	0.82	Cytochrome C551
	2cba.166_171	8.05	22.25	0.97	0.27	Carbonic anhydrase
	3psg.316_321	15.42	46.44	1.93	0.86	Pepsinogen

*Loop names are defined as follows: PDB code of the protein + extension showing the first and last residue of the loop. When several chains are present in one PDB file, the chain letter is the residue number prefix.
#RMSD values between all heavy atoms of loops and minimized reference CDR. Side-chain minimized loops are compared to side-chain minimized reference CDR (reciprocally for all-atom minimized loops).
§RMSD values between backbone atoms of loops and reference CDR from crystal structure (atoms N, CA, C).
¶Origin of the loop as listed in the PDB file.

gas phase energy scale between this loop and the subsequent one still exists, although it is reduced to ~7 kcal/mol. However, in solution, a non-"reference-like" CDR loop (3grs.321_326, Fig. 6) is found to have the lowest conformational energy (Table 9 *B*). This loop is ranked third in gas phase, but has a large favorable solvation free energy (~10 kcal/mol lower than the reference CDR). Such a lower solvation free energy is correlated to a significant increase in the surface area accessible to the solvent of this loop conformation.

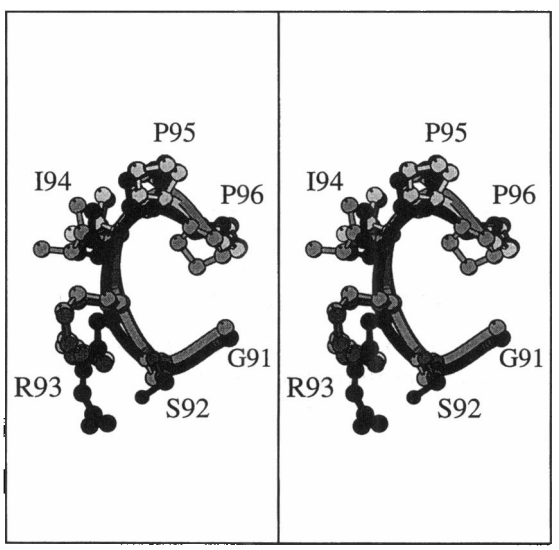

FIGURE 4 Stereo backbone atom superimposition of the three lowest conformational energy loops (6fab.L91_L96, 2cba.166_171, and 1rei.A91_A96, all-atom minimized, same color coding as in Fig. 3) on the minimized reference CDRL3 of R45 (*in black*). The RMSD values are, respectively, 0.23, 0.27, 0.29 Å for atoms N, CA, C.

CDRH3: side-chain optimization only (1920 loops)

Like that of CDRH1, the range in gas phase conformational energy is narrow (1329 loops have energies lower than 1000 kcal/mol). Six loops have lower conformational gas phase energy than the reference CDR (Table 10 *A*). Moreover, these six loops display a "reference-like" CDR geometry (RMSD ≤ 0.3 Å). No significant gap in the energy scale is observed,

CDRH3: all-atom optimization (1920 loops)

Because of the short length of CDRH3 (four residues), all loops are fully optimized. We find that 144 loops have gas phase conformational energies lower than 5 kcal/mol (190 loops have gas phase conformational energies lower than 10 kcal/mol). We observed that 59 loops have gas phase conformational energies lower than the reference CDR, and that the 124 lowest conformational energy loops have a "reference-like" CDR geometry (RMSD < 0.5 Å for backbone atoms with an average of 0.2 Å; Fig. 7 *A*). We count 32 loops that have lower solution conformational energies than that of the reference CDR, with an average RMSD for backbone atoms of 0.4 Å (1.7 Å for all heavy atoms). However, we find that, as in CDRH2, the lowest conformational energy loops in solution have a "non-reference-like" geometry (the average RMSD for 10 lowest energy conformations is 0.7 Å for backbone atoms and 3.5 Å for all heavy atoms; Fig. 7 *B*). The reordering of loops in the energy scale is due to a strong favorable solvation free energy (from −8 to −27 kcal/mol lower than the reference CDR). As in CDRH2, a larger accessible surface area exposed to solvent is responsible for these lower solvation free energies. Interestingly, we notice that these 10 "non-reference-like" CDRs have systematically higher bonded energy (mainly covalent

TABLE 8 Five lowest conformational free energy loops for CDRH1

	Loop*	ΔG_{conf}^{sol} (kcal/mol)	ΔG_{conf}^{gas} (kcal/mol)	RMSD (Å)#	RMSD (Å)§	Origin of the loop¶
A. CDRH1 Side chains minimized	R45 CDRH1	0	0	0	0	Antibody R45-45-11
	2cpp.86_88	—	−40.58	1.89	0.12	Cytochrome C peroxidase
	4icb.43_45	—	−40.28	1.95	0.18	Calbindin D9K
	2ts1.142_144	—	−40.24	1.83	0.10	Tyrosyl tRNA synthetase
	1tie.107_109	—	−40.13	1.84	0.22	Erythrina trypsin inhibitor
	2alp.86_88	—	−40.11	2.12	0.24	Alpha-lytic protease
B. CDRH1 All atoms minimized	2cdv.98_100	3.16	6.26	1.81	0.08	Cytochrome C3
	3fab.L134_136	3.43	7.573	1.38	0.07	Antibody New
	2ts1.42_44	3.50	5.634	1.45	0.07	Tyrosyl tRNA synthetase
	2ts1.255_257	4.79	6.652	1.86	0.11	Tyrosyl tRNA synthetase
	1ton.124_126	5.15	7.978	1.50	0.07	Tonin

Symbols have the same meanings as in Tables 5–7.
—, Not computed.

angle energy; data not shown) than the reference CDR or loops with "reference-like" CDR geometry.

DISCUSSION

A database of loops

We created a database of loops to generate reasonable backbone conformations by extracting loops from high-resolution protein crystal structures. The aim was for the database to contain loops displaying enough diversity in their conformation. We believe that sufficient diversity is achieved in our database because of the range of RMSD values obtained in optimized loops (0.16–8.12 Å for heavy atoms). Although our database does not encompass the complete conformational space for each loop, it does reflect conformational variation within a protein environment.

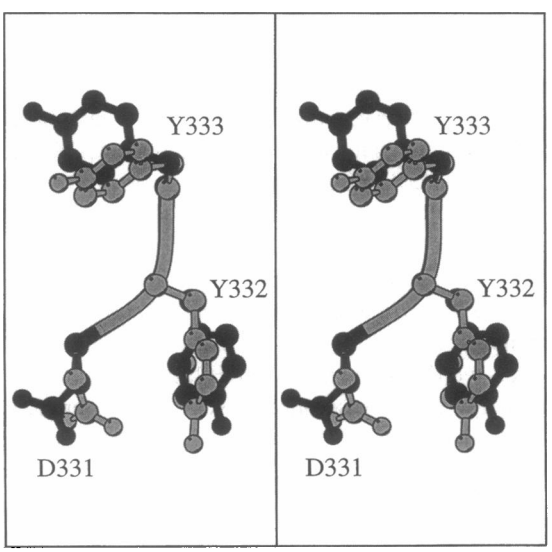

FIGURE 5 Stereo backbone atom superimposition of the lowest conformational energy loop (2cpp.86_88, side-chain atom minimized, same color coding as in Fig. 3) on the minimized reference CDRH1 of R45 (*in black*). The RMSD values are, respectively, 0.12 Å for atoms N, CA, C and 1.89 Å for all heavy atoms, because of the Tyr^{333} (H33 in Kabat's numbering).

Our database was built to reflect the diversity of antibody CDRs by generating subdatabases for each of them. We increased the flexibility by dividing these subdatabases into groups corresponding to the length of the segment to be modeled. The conformation of loops in our database was not restricted solely to CDRs, as would have occurred if loops were extracted using a RMSD criterion (Tramontano and Lesk, 1992). Additional advantages of our database as a generator of loop conformations are that the process is relatively fast, new loops are easy to add, it is not limited to CA atoms, and there is no need to spend time on generating loops when changing the antibody to be modeled.

Loops constituting our database came from a wide variety of protein family topologies. Loops from antibody families were generally the best candidates for modeling (discussed below). However, in the case of CDRL2 and CDRL3, and more significantly for CDRH1 and CDRH3, loops extracted from nonantibody molecules were also found to be good models, as shown by their low RMSD (<1.5 Å for all atoms and <0.5 Å for backbone atoms). This result is consistent with the fact that CDR conformation can be found in non-antibody-related molecules (Tramontano and Lesk, 1992).

Modeling side chains of a loop

The database of loops allowed us to generate a large diversity of backbone conformations. To evaluate the stability of each loop from the database, the original side chains were replaced to match those of R45. Several more or less complex techniques have been developed for modeling side-chain conformations (Lee and Subbiah, 1991; Roitberg and Elber, 1991; Tuffery et al., 1991; Desmet et al., 1992; Dunbrack and Karplus, 1993; Eisenmenger et al., 1993; Lasters and Desmet, 1993; Wilson et al., 1993; Koehl and Delarue, 1994). In this study, we selected only a single conformation for each side chain, based on the highest probability of occurrence in the PDB according to rotamer

TABLE 9 Five lowest conformational free energy loops for CDRH2

	Loop*	ΔG_{conf}^{sol} (kcal/mol)	ΔG_{conf}^{gas} (kcal/mol)	RMSD (Å)#	RMSD (Å)§	Origin of the loop¶
A. CDRH2 Side chains minimized	R45 CDRH2	0	0	0	0	Antibody R45-45-11
	2fbj.H52_H57	—	−3.08	0.79	0.21	Antibody J539
	1tie.4_9	—	44.79	1.74	0.58	Erythrina trypsin inhibitor
	1fnr.84_89	—	66.34	1.86	0.57	Ferredoxin reductase
	4enl.12_17	—	84.20	1.97	0.59	Enolase
	1fnr.122_127	—	108.55	2.21	0.65	Ferredoxin reductase
B. CDRH2 All atoms minimized	3grs.321_326	−4.47	9.16	1.43	0.96	Glutathione reductase
	2fbj.H52_H57	−2.59	−3.39	1.06	0.20	Antibody J539
	1fnr.122_127	−0.19	8.88	1.43	0.88	Ferredoxin reductase
	1wsy.B557_B562	0.40	4.90	1.32	0.31	Tryptophan synthase
	1tie.4_9	4.47	8.25	1.29	0.75	Erythrina trypsin inhibitor

Symbols have the same meanings as in Tables 5–7.
—, Not computed.

library data (Tuffery et al., 1991). We used this strategy for two reasons. First, a huge amount of computational time is required to generate and evaluate the conformational free energy of the 2187 (3^7) conformations for a seven-residue loop (assuming only three rotamer values are selected per side chain). Second, the high solvent exposure of CDR loops results in fewer geometrical constraints than for buried side chains. Most side chains, even after extensive minimization, kept their originally assigned rotamer dihedral angle values. We emphasize that assigned rotamer values were not necessarily similar to values in the reference R45 (Table 2). In addition, conformational changes during antigen binding are commonly found, especially for side chains (Bhat et al., 1990; Stanfield et al., 1990; Rini et al., 1992; Arevalo et al., 1993; Braden et al., 1996), meaning that a realistic modeling of side chains can only be undertaken during docking of the antigen with the antibody. However, when a side chain of a particular residue in a CDR is restricted to one rotamer in known antibody structures, this value should be used as the starting guess.

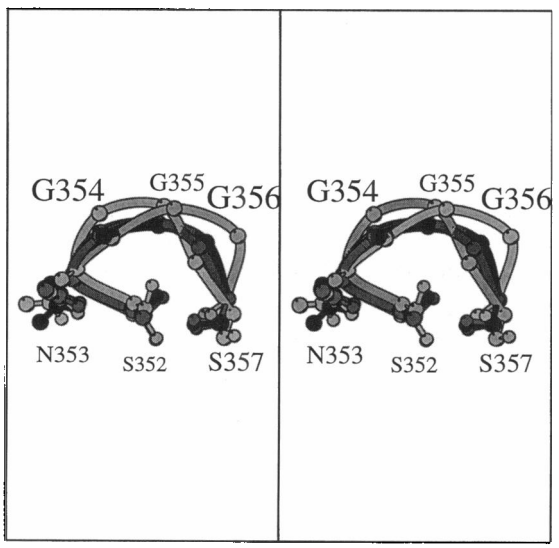

FIGURE 6 Stereo backbone atom superimposition of the four lowest conformational energy loops in solution (3grs.321_326, 2fbj.H52_H57, 1fnr.122_127, 1wsy.B557_B562, all-atom minimized) on the minimized reference CDRH2 of R45 (*in black*). Darker gray loops have the lowest RMSD, whereas the light gray loops have high RMSDs. The RMSD values are, respectively, 0.96, 0.20, 0.88, 0.31 Å for atoms N, CA, C. In Kabat's numbering, 352 is H52 and 357 is H56.

What is a good loop?

The success of any predictive method relies on the ability to predict the conformation of a loop similar to a reference one (usually from x-ray crystal structure). The root mean square deviation is commonly used to assess the similarity between two conformations of a protein (Rao and Rossmann, 1973; Cohen and Sternberg, 1980; Remington and Matthews, 1980). Although the RMSD can distinguish two different conformations, it is a more subjective criterion for recognizing similar conformations (Maiorov and Crippen, 1994). A recent study by Tramontano and Lesk (1992) shows that CDRs from antibodies can model other antibody CDRs within an average RMSD of 0.5 Å for backbone atoms. This implies that if a method can identify loops that had a RMSD below 0.5 Å, these loops are certainly the best ones to fit a current model. In the present study we found that loops with the lowest conformational free energy display a RMSD lower than 0.5 Å for backbone atoms and lower than 1.5 Å for all heavy atoms (although up to 2 Å is consistently observed for CDRH1 "reference-like" CDR geometry, because of a tyrosine side chain). Consequently, in this study a loop was considered to be "reference-like" if its RMSD value for backbone atoms was below 0.5 Å and below 1.5 Å for all heavy atoms (~2 Å for CDRH1).

What are the determinant parameters for a "reference-like" CDR loop?

In this section we assess contributions that allow the discrimination of correct from incorrect loops (only all-atom minimized loops) and discuss the fact that loops extracted

TABLE 10 Five lowest conformational free energy loops for CDRH3

	Loop*	ΔG^{sol}_{conf} (kcal/mol)	ΔG^{gas}_{conf} (kcal/mol)	RMSD (Å)#	RMSD (Å)§	Origin of the loop¶
A. CDRH3 Side chains minimized	R45 CDRH3	0	0	0	0	Antibody R45-45-11
	1mvp.A74_A77	—	−7.33	1.11	0.33	Myeloblastosis viral protease
	2cna.160_163	—	−5.30	1.10	0.19	Concanavalin A
	3cd4.A132_A134	—	−5.16	0.46	0.17	CD4
	1prc.C295_C298	—	−4.65	0.62	0.20	Photosynthetic reaction center
	4mdh.A201_A204	—	−4.64	1.13	0.22	Malate dehydrogenase
B. CDRH3 All atoms minimized	4mdh.A256_A259	−13.95	3.49	2.34	0.69	Malate dehydrogenase
	8acn.423_426	−5.33	3.44	2.45	0.65	Aconitase
	3mcg.1128_1131	−5.33	10.77	6.47	0.48	Light chain dimer MCG
	1fcb.A422_A425	−4.24	18.02	2.26	0.79	Flavocytochrome B2
	4mdh.A27_A30	−4.23	17.42	2.15	0.88	Malate dehydrogenase

Symbols have the same meanings as in Tables 5–7.
—, Not computed.

from antibodies were always found to be the best candidates. We want to emphasize that this section reports various enthalpy contributions for loops that have been fully minimized and that conclusions are drawn from a comparative study between all such loops. A careful statistical analysis was carried out for light chain CDRs; however, conclusions reported in this section applied exactly to heavy chain CDRs.

First, we found a strong correlation ($r \geq 0.95$) between ΔG^{gas}_{conf} and the sum of internal bonded energy (Brünger, 1992), as well as between ΔG^{gas}_{conf} and van der Waals energy ($r \approx 0.90$). This indicates that high energy loops have high van der Waals energies as well as high internal bonded energies. All of these loops have been extensively minimized, with no bad contacts or clashes remaining. Consequently, a higher van der Waals energy signified imperfect packing rather than bad steric contacts. We checked all loops for imperfect closure with the framework (internal bonded energies). However, it is an intrinsic property of current force fields that they minimize the total energy of a system, so low energy conformations can be achieved in different ways. In particular, we found that van der Waals energies can be overminimized by sacrificing internal bond energy, such as bond angle energies. Although a few loops displayed some high bond angle energies (specially in CDRL1), it was not the general trend and therefore could not be responsible for the correlation observed between ΔG^{gas}_{conf} and the internal bonded energy. For instance, CDRL3 showed no bad geometric closures (no deviation around an average bond distance (<0.3 Å), no deviation around the average bond angle (<20°), and no deviation around the average improper angle (<30°)), but displayed a strong correlation between free energy and internal bonded energy ($r = 0.96$). A dramatic example is seen with CDRH3 loops. Indeed, loops that have lower gas phase conformational energy than the reference R45 CDRH3 are found to have a "reference-like" geometry. In contrast, the 10 loops that have the lowest conformational energy in solution do not display "reference-like" CDR geometry. We found that these 10 conformations systematically have a covalent bond angle energy that on average is 11 kcal/mol higher (~1356 kcal/mol) than the reference (1345 kcal/mol). In the gas phase these loops were ranked as low as 399 in the energy scale. We conclude that the difference in energy among loops comes from a concomitant effect of internal bonded energy and packing (van der Waals energy). Second, why were antibody loops found to be the best models? Without exception, they have the lowest van der Waals energy as well as the lowest internal bonded energies. We suggest that antibody loops are the best candidates because their backbones pack better into the highly conserved FR compared with a nonantibody loop. Because the orientation of side chains was of minor importance, the most crucial interaction appears to be between the backbone of a loop and the FR. Because the FR is structurally conserved, we are not surprised that loops coming from a similar environment (another antibody) were found to dock better to the R45 framework. Therefore local environment appears to be the dominant factor in CDR modeling, because nonantibody loops were found to be incapable of accurately modeling long CDRs (≥6 residues) in general. Although it may not be surprising that CDRs are the best candidates for modeling CDRs, our conformational free energy screening is highly accurate, because no "false positives" were found (using gas phase conformational energies). More interestingly, our energy screening discriminates correct from incorrect conformation, even when all atoms of loops were minimized.

Conformational free energy analysis

Side chains versus all-atom optimization

Because the computational time required for minimizing seven-residue loops is not negligible, it is fair to ask whether side-chain optimization alone is sufficient during a large screening of candidates. Nevertheless, we agree that complete optimization is required at the ultimate stage of modeling, to remove any bad intramolecular contacts. According to our results, the same loops were found to be correct in both optimization methods (Tables 5–7, 9; compare side

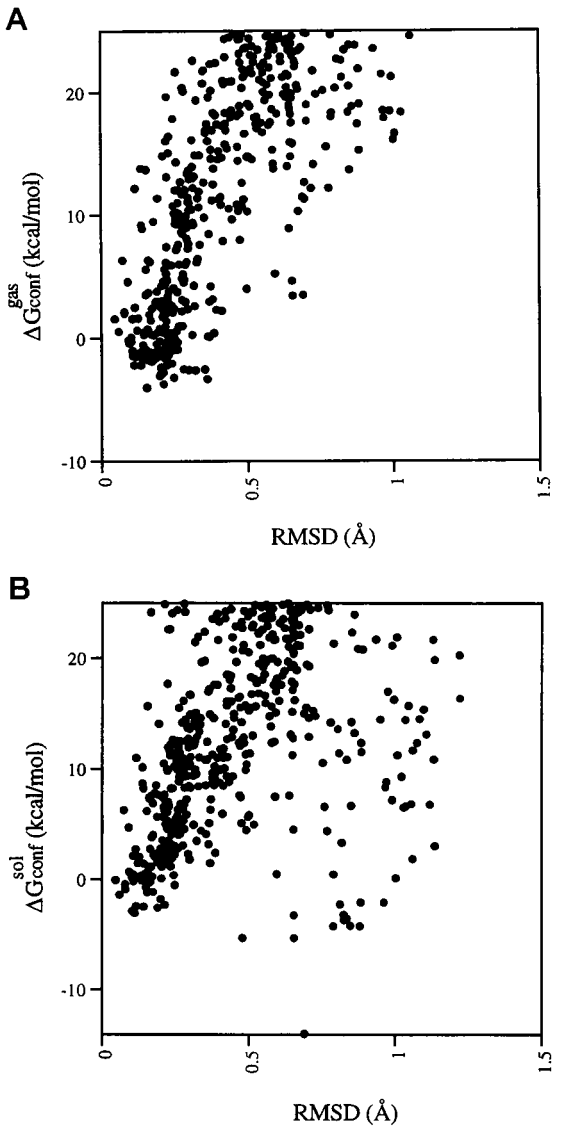

FIGURE 7 Conformational free energy in (*A*) gas phase and (*B*) solution for all-atom minimized loops modeling CDRH3 plotted against backbone atom RMSD from the minimized reference CDRH3. It is clear that solvation free energy confuses the identification of "reference-like" conformations by lowering the total conformational energies from non-"reference-like" loop conformation.

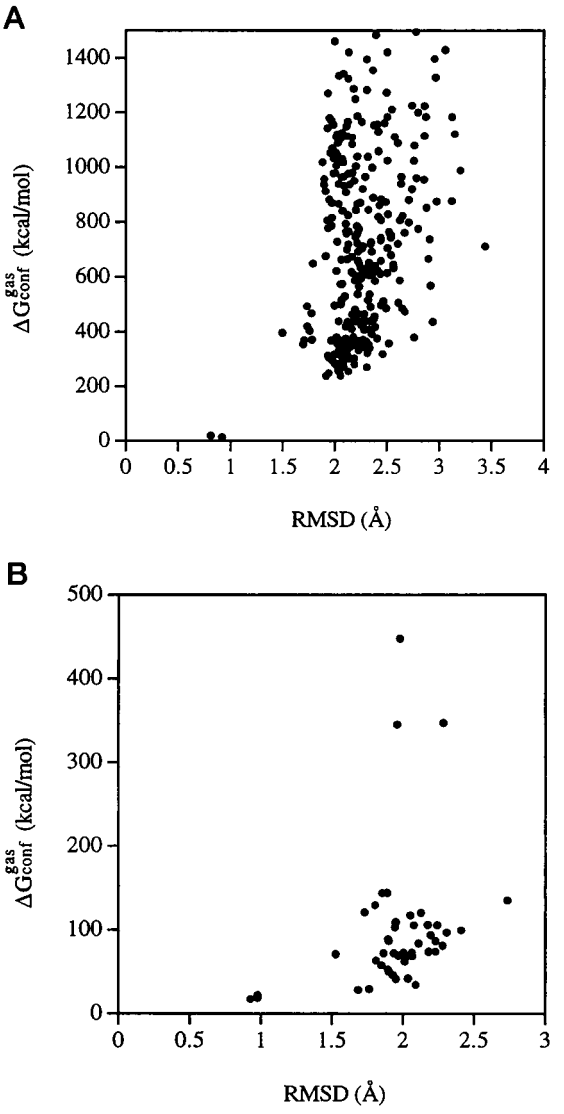

FIGURE 8 Conformational free energy in gas phase of loops modeling CDRL3, where (*A*) only side-chain atoms or (*B*) all atoms were minimized and plotted against heavy atoms RMSD from the minimized crystal structure. Fewer loops appear in *B* because only 47 loops are fully minimized for CDRL3 (see text).

chain versus all-atom minimized loops). We found that the best loops (RMSD < 1.5Å) had a much lower free energy than incorrect loops when only side chains were minimized (Fig. 8). Such a separation in the energy makes it easier to identify correct loops (Fig. 8 *A*). When all atoms of loops are minimized, this gap in energy was reduced or disappeared (Fig. 8 *B*). Nevertheless, the lowest conformational free energies (in the gas phase) always correlated with "reference-like" CDR conformations. A second aspect of this comparison should be considered: Does a complete optimization improve the geometry toward the reference structure? As shown in Tables 5–10 and Fig. 2, there is no significant improvement when all atoms are minimized, except for one loop in CDRL3. An extended minimization (all atoms) led to a compaction on the conformational energy scale, but did not improve loop conformations toward the reference state (RMSD similar to that of only side-chain minimization). Therefore a large and fast screening may not require full optimization of all loop atoms. A complete minimization is therefore best carried out on a few selected conformations at a latter stage.

Gas phase versus solution conformational free energy

An important question is: Does inclusion of the solvation free energy improve the identification of correct loops? Based on the results presented in Tables 5 *B*–10 *B*, the conformational free energy in gas phase alone can be used

to identify "reference-like" conformations. When only side chains were minimized, the inclusion of solvation free energy (Vl CDRs) did not help distinguish candidates, because of the very high free energy of such conformations. When loops were completely optimized (all-atom minimization), the solvation free energy sometimes allowed a more accurate description of the near reference conformation (solution conformational energies of loops are closer to the reference CDR than gas phase conformational energies; Tables 5 *B*–9 *B*). However, for short loops (4–6 residues) solvation free energies misled the identification of a "reference-like" CDR geometry (Tables 9 *B* and 10 *B*). Therefore, it is clear that a low solvation free energy is not correlated with a "reference-like" geometry. Moreover, during a fast screening test, calculation of the solvation free energies for each loop is computationally expensive. We suggest that solvation free energies should only be calculated for the conformations with the lowest range of gas phase energies.

CONCLUSIONS

In this paper we have developed a method for identifying the "reference-like" conformation of loops when inserted into an antibody framework. The stability of each loop was assessed by calculating the conformational free energy difference between the model and the reference structure. Results showed that the most stable loops (in the gas phase) gave the best models (RMSD < 0.5 Å for backbone atoms and 1.5 Å for all heavy atoms). Naturally, the most stable loops exclusively originated from antibody molecules for long CDRs ($\geq$7 residues), mainly due to local environment effects. Whereas, for shorter CDRs ($\leq$6 residues), the PDB contains enough redundant structures to encompass CDR conformations for loops not originating from antibodies. Our study reveals that extensive optimization (as opposed to only side-chain minimization) did not improve the geometry of the model toward the reference. We found that most of the long loops (more than 6 residues) with the lowest conformational free energy, with only side-chain optimization, were also the lowest in all-atom minimization. The inclusion of the solvation free energy in the evaluation of loop stability is not required to identify "reference-like" CDR geometry, even when loops are fully minimized. Further developments will include a decomposition of solvation free energy terms to understand, at the atomic level, the bias of solution conformational energies toward "non-reference-like" CDRs.

This work has been carried out in the laboratory of Prof. Barry Honig in the Department of Biochemistry and Molecular Biophysics at Columbia University. We acknowledge D. Altschuh and J.-C. Thierry for providing the coordinates of the Fab R45-45-11. JLP thanks V. A. Roberts for critical reading of the manuscript, J. A. Tainer for fruitful comments and for partial support from grant NSF BIR963146, and M. Israel for editing. We are indebted to all x-ray crystallographers who submitted coordinates to the Brookhaven Data Bank. Part of this work has been presented at the 10th Symposium of the Protein Society at San Jose, CA, August 3–7, 1996.

REFERENCES

Altschuh, D., O. Vix, B. Rees, and J. C. Thierry. 1992. A conformation of cyclosporin A in aqueous environment revealed by the X-ray structure of a cyclosporin-Fab complex. *Science.* 256:92–94.

Arevalo, J. H., E. A. Stura, M. J. Taussig, and I. A. Wilson. 1993. Three-dimensional structure of an anti-steroid Fab′ and progesterone-Fab′ complex. *J. Mol. Biol.* 231:103–118.

Ben-Naim, A., and Y. Marcus. 1984. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* 81:2016–2027.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Bhat, T. N., G. A. Bentley, T. O. Fischmann, G. Boulot, and R. J. Poljak. 1990. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature.* 347:483–485.

Braden, B. C., B. A. Fields, X. Ysern, F. A. Goldbaum, W. Dall'Acqua, F. P. Schwarz, R. J. Poljak, and R. A. Mariuzza. 1996. Crystal structure of the complex of the variable domain of antibody D1.3 and turkey egg white lysozyme: a novel conformational change in antibody CDR-L3 selects for antigen. *J. Mol. Biol.* 257:889–894.

Bruccoleri, R. E., E. Haber, and J. Novotny. 1988. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature.* 335:564–568.

Bruccoleri, R. E., and M. Karplus. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers.* 26:137–168.

Brünger, A. T. 1992. X-PLOR Manual, Version. 3.0. Yale University, New Haven, CT.

Brünger, A. T., and M. Karplus. 1988. Polar hydrogen positions in proteins: empirical energy function placement and neutron diffraction comparison. *Proteins.* 4:148–156.

Carlacci, L., and S. W. Englander. 1993. The loop problem in proteins: a Monte Carlo simulated annealing approach. *Biopolymers.* 33: 1271–1286.

Chothia, C., and A. M. Lesk. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5:823–826.

Chothia, C., and A. M. Lesk. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* 196:901–917.

Cohen, F. E., and M. J. E. Sternberg. 1980. On the prediction of protein structure: the significance of the root-mean-square deviation. *J. Mol. Biol.* 138:321–333.

de la Paz, P., B. J. Sutton, M. J. Darsley, and A. R. Rees. 1986. Modelling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO J.* 5:415–425.

Desmet, J., M. De Maeyer, B. Hazes, and I. Lasters. 1992. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature.* 356:539–542.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analyses program for the VAX. *Nucleic Acids Res.* 12:387–395.

Dunbrack, R. L., Jr., and M. Karplus. 1993. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* 230:543–574.

Eisenmenger, F., P. Argos, and R. Abagyan. 1993. A method to configure protein side-chains from the main-chain trace in homology modelling. *J. Mol. Biol.* 231:849–860.

Epp, O., E. E. Lattman, M. Schiffer, R. Huber, and W. Palm. 1975. The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution. *Biochemistry.* 14:4943–4952.

Feldmann, R. J., M. Potter, and C. P. J. Glaudemans. 1981. A hypothetical space-filling model of the V-regions of the galactan-binding myeloma immunoglobulin J539. *Mol. Immunol.* 18:683–698.

Fine, R. M., H. Wang, P. S. Shenkin, D. L. Yarmush, and C. Levinthal. 1986. Predicting antibody hypervariable loop conformations. II. Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. *Proteins.* 1:342–362.

Fischmann, T. O., G. A. Bentley, T. N. Bhat, G. Boulot, R. A. Mariuzza, S. E. V. Phillips, D. Tello, and R. J. Poljak. 1991. Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution. *J. Biol. Chem.* 266:12915–12920.

Furey, W., Jr., B. C. Wang, C. S. Yoo, and M. Sax. 1983. Structure of a novel Bence-Jones protein (Rhe) fragment at 1.6 Å resolution. *J. Mol. Biol.* 167:661–692.

Gilson, M. K., and B. Honig. 1988. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins.* 4:7–18.

Herron, J. N., X.-M. He, M. L. Mason, E. W. Voss, Jr., and A. B. Edmundson. 1989. Three-dimensional structure of a fluorescein-fab complex crystallized in 2-methyl-2,4-pentanediol. *Proteins.* 5:271–280.

Honig, B., K. Sharp, and A.-S. Yang. 1993. Macroscopic models of aqueous solutions: biological and chemical applications. *J. Phys. Chem.* 97:1101–1109.

Hubbard, T. J. P., and T. L. Blundell. 1987. Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. *Protein Eng.* 1:159–171.

Jean-Charles, A., A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T. F. Hendrickson, and W. C. Still. 1991. Electrostatic contributions to solvation energies: comparison of free energy perturbation and continuum calculations. *J. Am. Chem. Soc.* 113:1454–1455.

Jones, T. A., and S. Thirup. 1986. Using known substructures in protein model building and crystallography. *EMBO J.* 5:819–822.

Kabat, E. A., T. T. Wu, and H. Bilofsky. 1977. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J. Biol. Chem.* 252:6609–6616.

Kabat, E. A., T. T. Wu, H. M. Perry, K. S. Gottesman, and C. Foeller. 1991. Sequences of Proteins of Immunological Interest. U.S. Department of Health and Human Services, Washington, DC.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

Koehl, P., and M. Delarue. 1994. Application of a self-consistent mean field theory to predict side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* 239:249–275.

Koehl, P., and M. Delarue. 1995. A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modeling. *Nature Struct. Biol.* 2:163–170.

Kraulis, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.

Lacombe, M.-B., P. M. Alzari, R. J. Poljak, and A. Nisonoff. 1992. Three-dimensional structure of two crystal forms of FabR19.9 from a monoclonal anti-arsonate antibody. *Proc. Natl. Acad. Sci. USA.* 89: 9429–9433.

Lasters, I., and J. Desmet. 1993. The fuzzy-end elimination theorem: correctly implementing the side-chain placement algorithm based on the dead-end elimination theorem. *Protein Eng.* 6:717–722.

Lee, C., and S. Subbiah. 1991. Prediction of protein side-chain conformation by packing optimisation. *J. Mol. Biol.* 217:373–388.

Mainhart, C. R., M. Potter, and R. J. Feldmann. 1984. A refined model for the variable domains (Fv) of the J539 β(1,6)-D-galactan-binding immunoglobulin. *Mol. Immunol.* 21:469–478.

Maiorov, V. N., and G. M. Crippen. 1994. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J. Mol. Biol.* 235:625–634.

Marquart, M., J. Deisenhofer, and R. Huber. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol.* 141:369–391.

Martin, A. C. R., J. C. Cheetham, and A. R. Rees. 1989. Modeling antibody hypervariable loops: a combined algorithm. *Proc. Natl. Acad. Sci. USA.* 86:9268–9272.

Moult, J., and M. N. G. James. 1986. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins.* 1:146–163.

Nicholls, A., and B. Honig. 1991. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* 12:435–445.

Padlan, E. A., G. H. Cohen, and D. R. Davies. 1985. On the specificity of antibody/antigen interactions: phosphocholine binding to McPC603 and the correlation of three-dimensional structure and sequence data. *Ann. Inst. Pasteur.* 136C:271–276.

Padlan, E. A., E. W. Silverton, S. Sheriff, and G. H. Cohen. 1989. Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. USA.* 86:5938–5942.

Ponder, J. W., and F. M. Richards. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* 193:775–791.

Presta, L. G., and G. D. Rose. 1988. Helix signals in proteins. *Science.* 240:1632–1641.

Rao, S. T., and M. G. Rossmann. 1973. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* 76:241–256.

Remington, S. J., and B. W. Matthews. 1980. A systematic approach to the comparison of protein structures. *J. Mol. Biol.* 140:77–99.

Ring, C. S., D. G. Kneller, R. Langridge, and F. E. Cohen. 1992. Taxonomy and conformational analysis of loops in proteins. *J. Mol. Biol.* 224:685–699.

Rini, J. M., U. Schulze-Gahmen, and I. A. Wilson. 1992. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science.* 255:959–965.

Roitberg, A., and R. Elber. 1991. Modelling side-chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing method to find minimum energy conformations. *J. Chem. Phys.* 95:9277–9287.

Saul, F. A., L. M. Amzel, and R. J. Poljak. 1978. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 Å resolution. *J. Biol. Chem.* 253:585–597.

Shenkin, P. S., D. L. Yarmush, R. M. Fine, H. Wang, and C. Levinthal. 1987. Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. *Biopolymers.* 26: 2053–2085.

Sheriff, S., E. W. Silverton, E. A. Padlan, G. H. Cohen, S. Smith-Gill, B. C. Finzel, and D. R. Davies. 1987. Three-dimensional structure of an antibody-antigen complex. *Proc. Natl. Acad. Sci. USA.* 84:8075–8079.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Smith, K. C., and B. Honig. 1994. Evaluation of the conformational free energies of loops in proteins. *Proteins.* 18:119–132.

Sridharan, S., A. Nicholls, and B. Honig. 1992. A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* 61:A174.

Stanfield, R. L., T. M. Fieser, R. A. Lerner, and I. A. Wilson. 1990. Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. *Science.* 248:712–719.

Strong, R. K., R. Campbell, D. R. Rose, G. A. Petsko, J. Sharon, and M. N. Margolies. 1991. Three-dimensional structure of murine anti-*p*-azophenylarsonate Fab 36–71. 1. X-ray crystallography, site-directed mutagenesis, and modeling of the complex with hapten. *Biochemistry.* 30:3739–3748.

Summers, N. L., and M. Karplus. 1990. Modeling of globular proteins. A distance-based data search procedure for the construction of insertion/deletion regions and Pro↔non-Pro mutations. *J. Mol. Biol.* 216: 991–1016.

Tramontano, A., and A. M. Lesk. 1992. Common features of the conformations of antigen-binding loops in immunoglobulins and application to modelling loop conformations. *Proteins.* 13:231–245.

Tuffery, P., C. Etchebest, S. Hazout, and R. Lavery. 1991. A new approach to the rapid determination of protein side chain conformations. *J. Biomol. Struct. Dyn.* 8:1267–1289.

Vix, O., B. Rees, J.-C. Thierry, and D. Altschuh. 1993. Crystallographic analysis of the interaction between cyclosporin A and the Fab fragment of a monoclonal antibody. *Proteins.* 15:339–348.

Wilson, C., L. M. Gregoret, and D. A. Agard. 1993. Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. *J. Mol. Biol.* 229:996–1006.