Calculation of binding energies using a robust molecular mechanics technique: Application to an antibody–antigen complex

I. Haneef

Astbury Department of Biophysics, University of Leeds, Leeds, UK

A novel molecular mechanics technique, which is both computationally efficient and robust, for calculation of relative stability of macromolecules and binding energies is presented. The technique delivers exact results for a number of hypothetical systems; the technique can be used to energy minimize a number of similar macromolecules simultaneously; simultaneous minimization of many structures requires computer time only fractionally over that needed to energy minimize one such structure. The method has been used to successfully calculate the relative stability and binding of two avian lysozymes to the monoclonal antibody D1.3.

Keywords: antibody–antigen interactions; binding energies; molecular mechanics; molecular recognition; relative stability of molecules

INTRODUCTION

There now exist a large number of exquisite examples of molecular recognition processes ranging from the highly specific binding of small ligands to proteins, to the association of pairs of large macromolecules such as nucleic acids and proteins or antigens and antibodies. Site-directed mutagenesis[1] has provided a very powerful approach to testing structure–function hypotheses by introducing new amino acids in positions thought to be responsible for molecular recognition or catalysis. Such experiments provide, in principle, a powerful approach for the design of genetically engineered molecules with modified or novel properties of clinical or industrial importance.

Address reprint requests to Dr. Haneef at the Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, UK.
Received 27 July 1989; accepted 6 September 1989

Despite a considerable body of structural and biochemical data, in many cases the exact nature of the interactions responsible for highly specific recognition processes are not well understood.[2] This lack of understanding of the basic interactions at the atomic level currently poses a major obstacle to the design of novel proteins with specific properties. Here we describe a computationally efficient computer simulation technique that can be used to make useful predictions about the effect of site-specific mutations on protein structure and stability and we report results on the application of this technique to study the structural effects and energetics of complex formation between an antibody D1.3[3] and its lysozyme antigens.

METHODOLOGY

In most molecular mechanics[4] studies one is simply interested in finding a conformation of the molecule with low energy; in such cases, minimization is performed to relieve strain or bad steric contacts in model-built or experimentally determined conformations (such as from distance geometry techniques employing NMR data). For such studies the various molecular mechanics techniques are adequate. These techniques can also be used to determine the relative stabilities of different possible structures of the same molecule, by invoking the assumption that the conformations with lower potential energy have higher statistical weight than those with higher energy. However, when one is interested in calculating binding energies or relative stabilities of similar, but not identical, molecules considerable care is required; most methods currently employed for such studies are inadequate and often give grossly misleading results.

Here we introduce the concept of "equivalent energy minima" (see Figure 1). Using the example given in Figure 1, we ascribe conformations X1 and Y1, and X2 and Y2 to be in equivalent energy minima. One can obtain binding

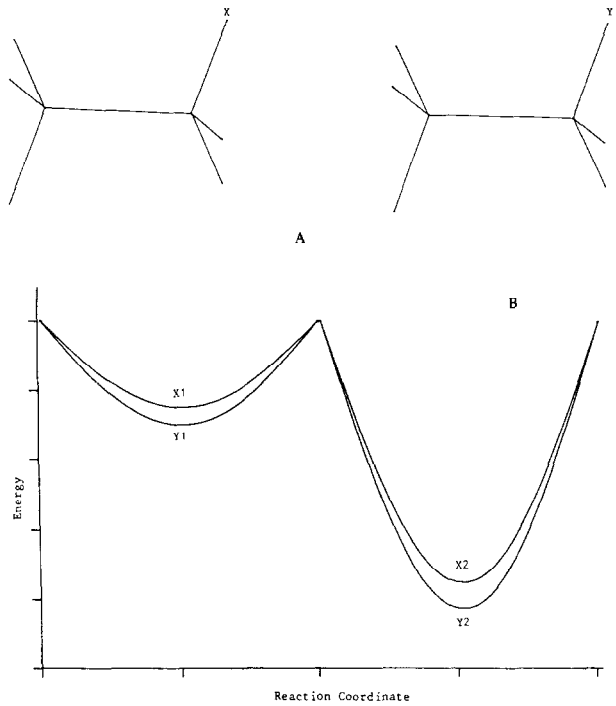

Figure 1. (a) Ethane-like molecules X and Y; they differ in that one has substituent X *while the other has substituent* Y*; (b) potential energy surfaces, and energy minima, for molecules* X *and* Y*. The figure represents a simple situation where one wishes to determine which of the two structures,* X *or* Y*, is more stable. For simplicity, we also assume that the exact potential energy surfaces for these two molecules can be calculated and have the profiles shown in Figure 1b. From the examination of these surfaces it is clear that structure* Y *is more stable than structure* X*. However, for macromolecules, such an examination of the potential energy surfaces is not possible due to the impossibly large computer resources required to generate them over whole of the configurational space. For this simple system, there exist four possible ways of comparing the energies of* X *and* Y*: comparison of energies of* X1 *with* Y1*,* X1 *with* Y2*,* X2 *with* Y1 *and* X2 *with* Y2*. Only the comparison of energies of conformations* X1 *with* Y1 *and* X2 *with* Y2 *give a fair indication of the relative stabilities of molecules* X *and* Y*. Comparisons of energies of conformations* X1 *with* Y2 *and* X2 *with* Y1 *give grossly incorrect results, and indeed the latter would indicate that molecule* X *is more stable than* Y

energies or relative stabilities of two similar molecules only when the comparisons of the energies is carried out for equivalent energy minima. Although this concept shows how the relative stabilities should be calculated, current simulation techniques do not provide a protocol for obtaining energy-minimized structures in equivalent energy minima since such techniques cannot be targeted to sample specified regions of the potential energy surface. We have developed a technique, SIMMIN (simultaneous minimization), that allows two or more similar structures to be targeted to equivalent energy minima.

The importance of attaining equivalent energy minima can be gauged from considering the following case: In studying the interactions between two molecules E and S, the binding energy for the process

$$E + S \rightarrow E.S \quad (1)$$

where E represents, say, an enzyme, S the substrate and $E.S$ the enzyme–substrate complex is obtained from

$$U(binding) = U(E.S) - U(E) - U(S) \quad (2)$$

where $U(X)$ represents the potential energy of the minimized structure X. Now consider the hypothetical case where there are no interactions between E and S; for this system, one would expect $U(binding)$ and RMS(E,E_c) to be exactly zero (where E is the conformation of the isolated energy-minimized E and E_c is conformation for E in the complex, and RMS represents the root mean square (rms) difference between two conformations). We have carried out a number of calculations for such a hypothetical binding of trimethoprim (TMP) to dihydrofolate reductase (DHFR); results from such calculations are presented in Table 1. In all cases, the structures were restrained to their initial conformations using a harmonic restraining term. The restraining force constant was decreased from 100 kcal/mol/Å^2 to zero over 2000 steps. Further minimizations were carried out without the use of any restraining terms. The minimizations were stopped when all components of the gradient vector had fallen below 0.01 kcal/mol/Å. The various calculations differ only in that we have used different search protocols in the conjugate gradient minimization routines.

The connotations from the results in Table 1 are clear. Consider the situation where we are interested in the difference in the binding energy of E to two similar substrates S and S'. If we were to use the method given above for calculating binding energy differences, we would obtain results in the range ± 40 kcal/mol—even if the two substrates did not interact with the enzyme! The only difference in the minimizations of E on the left- and right-hand sides of Equation 1 is the number of degrees of freedom of the system. On the left-hand side, minimization is carried out with N_E degrees of freedom; on the right hand side the total number of degrees of freedom is $N_E + N_S$, where N_E and N_S degrees of freedom are totally independent of each other. Indubitably, the total number of degrees of freedom has a profound effect on numerical minimization algorithms. In an exact analytical approach, the minimum obtained would depend solely on the starting structure of the molecule. Although the total energies of the enzyme and the complex, respectively, are very similar in magnitude in the different calculations, the rms differences in the resulting structures range from 0.5Å–0.8Å for Cα atoms and 0.7Å–1.1Å for all atoms (Table 2). Thus there are a number of factors that determine which particular minimum is obtained during energy minimization; one of these factors is the total number of degrees of freedom used in any calculation. The essential feature of the SIMMIN technique is to simultaneously minimize all molecules in Equation 1 in the same simulation; in this way all structures are subjected to identical conditions during the simulation and guarantees the convergence of structures to their equivalent energy minima. The technique has considerable advantages over currently available methods for calculating binding energies; in particular, the technique delivers exact results for a number of hypothetical situations of the type described above.

Table 1. Total and binding energies for the hypothetical binding of trimethoprim to dihydrofolate reductase where there are no interactions between the two molecules. For each calculation, all molecules were minimized using identical protocols. The three calculations differ in that different search methods were used in conjugate gradient minimization. In all calculations the minimizations were started from the same structures; the starting structures of DHFR in isolation and in the "complex" were also identical. Rms' (in Å) are for all atoms and Cα atoms between isolated and complexed enzyme

Calculation	*U* (DHFR.TMP)	*U* (DHFR)	*U* (TMP)	*U* (*binding*)	RMS
1	−5257	−5248	10	−19	0.39, 0.34
2	−5245	−5268	10	13	0.33, 0.26
3	−5259	−5267	10	−2	0.22, 0.16

Table 2. Rms differences (Å) between crystal (*X*) and energy-minimized isolated (*E*) and complexed (E_c) structures of DHFR from the three calculations referred to in text. Upper triangle, rms' for Cα atoms; lower triangle, rms' for all atoms

	X	*E*(1)	*E*(2)	*E*(3)	E_c(1)	E_c(2)	E_c(3)
X	–	0.78	0.87	0.80	0.84	0.83	0.82
E(1)	0.99	–	0.83	0.76	0.34	0.77	0.77
E(2)	1.11	1.09	–	0.50	0.83	0.26	0.52
E(3)	1.04	0.97	0.73	–	0.80	0.49	0.16
E_c(1)	1.04	0.39	1.10	1.01	–	0.78	0.78
E_c(2)	1.06	1.04	0.33	0.72	1.06	–	0.51
E_c(3)	1.06	0.99	0.75	0.22	1.00	0.75	–

The SIMMIN approach has as its basis the assumption, largely verified by examination of three dimensional structures of homologous proteins[5] and site-specific mutants,[6] that the change in conformation due to insertions or amino acid differences are localized rather than spread through the molecule. Although the technique has a wide range of applications, it is particularly suited to the study of the structure and stability of protein molecules that differ by a few amino acids (such as wild-type and site-specific mutants). In the case of a wild-type and a site-specific mutant, the first step in the SIMMIN approach is to construct the mutant structure based on the wild-type protein. This is done by model-building the appropriate sidechain in (one of) its preferred conformations.[7] We then define three regions in the molecules: the *C* region, which consists of all residues distant (e.g., all residues outside a 10Å sphere) from the site of mutation; and the F region, which comprises all residues in the immediate vicinity of the mutation including the site of mutation. In our studies, the F region is defined by all residues that are within a 5Å contacting distance of any atoms of either the wild-type or the mutated residue. Finally, the *R* region is defined as all residues intermediate between the *C* and *F* regions. The two structures, the wild-type and the mutant, are then energy-minimized simultaneously. During the simulation, all atoms within the *C* region of the wild-type and mutant structures are constrained to move along identical trajectories. Free, independent trajectories are allowed to develop only for the *F* regions. Atoms within the *R* region of the wild-type and mutant structures are restrained to have similar, but not identical, trajectories. Although the wild-type and mutant structures are present simultaneously in the same coordinate space, no interactions are calculated between the two molecules. The total potential function that is minimized can be represented as

$$V = V_w(C,R_w,F_w) + V_m(C,R_m,F_m) + R(R_w,R_m) \quad (3)$$

where $V_x(c,r,f)$ represents the potential energy of structure x, with atomic coordinates defined by variables c, r and f in the constrained, restrained and free regions, respectively; subscripts m and w refer to the mutant or wild-type structures, respectively. The function $R(R_w,R_m)$ in its simplest form is a restraining potential which ensures that the *R* regions in the wild-type and the mutant develop along similar trajectories. When convergence has been achieved using the SIMMIN potential of Equation 3, further minimizations are continued for the molecules independently of each other. In the SIMMIN approach we make the assumption that protein molecules with only a few amino acid differences should have similar structures; however, we do not have an *a priori* knowledge of the extent of any structural differences. The rationale for the independent minimizations at the final stages is to allow any larger scale differences to develop. Thus the technique is of wider applicability than the basic assumptions of this approach at first sight might suggest.

The main advantage of SIMMIN approach is that it generates structures in equivalent energy minima, thus permitting a direct comparison of the energies to establish the relative stability of two or more similar structures. Another advantage is that since large regions of the molecules follow identical trajectories, a substantial saving in computer time is achieved; simultaneous minimization of two molecules in the same simulation requires little over half the computer time required for two independent minimizations. There are no limitations in the SIMMIN approach as to how many molecules are simulated in the same simulation.

For example, in studying the interactions between antibodies and their specific antigens, the binding energy for the process

$$Ab + A \rightarrow Ab.A \quad (4)$$

where *Ab* represents an antibody, *A* the antigen and *Ab.A* the antibody–antigen complex is given by

$$\Delta(binding) = U(Ab.A) - U(Ab) - U(A) \quad (5)$$

where $U(X)$ represents the potential energy of the minimized structure X. The binding energy for a different antigen, A', is obtained similarly. In the SIMMIN approach all five molecules, *Ab.A*, *Ab.A'*, *Ab*, *A* and *A'*, can be simultaneously energy-minimized in the same simulation.

Antibody specificity

The monoclonal antibody D1.3 exhibits fine specificity for avian lysozymes,[3] being able to distinguish between antigens differing by only a single amino acid residue (Figure 2). Fab D1.3 binds hen egg white lysozyme (HEL) with an equilibrium affinity constant (k_{aff}) of *ca.* 4.5×10^7/mol, but if Gln 121 of HEL is replaced by His, the complex formation is effectively abolished. The fine specificity of D1.3, and the pivotal role played by a single antigen amino acid residue in affecting the binding affinity, provides an excellent model system for testing the predictive power of computer simulation techniques.

The antibody–antigen complex of D1.3/HEL

The three-dimensional crystallographic structure of this complex reveals an extensive and relatively flat interface between the two proteins, extending over maximum dimensions of *ca.* 20Å by 30Å; the contacting surfaces are complementary, with protruding side chains of one surface lying in the depressions of the other. The lysozyme antigenic determinant recognized by D1.3 consists of two stretches of polypeptide chain, comprising residues 18 to 27 and 116 to 129. These residues interact with all six complementarity determining regions (CDR's) of the Fab, 17 antibody residues making short contacts (≤4Å) with 16 antigen residues; these interactions comprise of many van der Waals' contacts interspersed with hydrogen bonds. The most striking example of the antibody–antigen complementarity is provided

Lysozymes Recognized by D1.3 Fab

	40	55	68	91	121	Binding
Hen egg	Thr	Ile	Arg	Ser	Gln	X
Bobwhite quail	Ser	Val	Lys	Thr	Gln	X
California quail	Ser	Val	Arg	Thr	His	

Figure 2. Sequence differences between egg lysozymes and effect on binding of Fab D1.3. Bobwhite quail lysozyme, which has four amino acid sequence differences but none in the interface with Fab D1.3, binds with similar affinity to that of HEL. The binding of antibody D1.3 to the lysozymes of partridge (3 amino acid differences), California quail (4 amino acid differences), Japanese quail (6 amino acid differences), turkey (7 amino acid differences), pheasant and guinea fowl (10 amino acid differences each) is undetectable. These lysozymes differ from hen lysozyme at position 121, which makes close contact with the antibody. In all except for Japanese quail and pheasant lysozymes, His replaces Gln at that position; in California quail and turkey lysozymes, the change at position 121 is the only change in the antibody–antigen interface

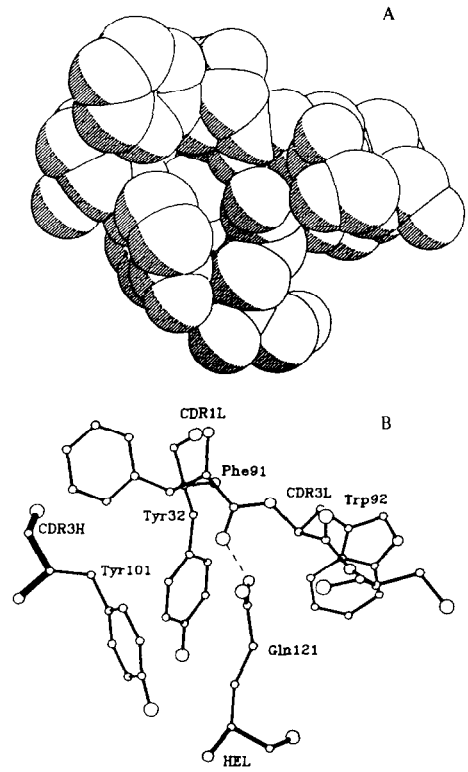

Figure 3. (a) The environment of Gln 121 with atoms drawn with their van der Waals' radii, showing the close packing of antibody residues around the antigen side chain (heavily shaded). (b) Skeletal drawing of Gln 121 and its environment; dotted line indicates the hydrogen bond between Nε of Gln and the main chain carbonyl oxygen of V_L *Phe 91.*

by the side chain of Gln 121 of lysozyme, which penetrates deeply into a depression in the Fab surface (Figure 3a); the amide nitrogen of Gln 121 forming a strong hydrogen bond with the main chain carbonyl oxygen of V_L Phe 91, with the rest of the side chain being surrounded by aromatic side chains of the Fab (Figure 3b).

METHODS

We have used multi-start[8] molecular mechanics simulations to study the structure and the energetics of the D1.3/lysozyme complexes. The actual model system studied consisted of the two variable domains of the Fab (V_L and V_H) complexed with the antigen. Only the polar hydrogen atoms, those bonded to oxygens, nitrogens or sulphurs, were included. The system thus consisted of 351 amino acid residues and a total of 3435 interacting centers, including "lone pairs" on sulphur atoms. Besides the 2.8Å resolution structure of the Fab/HEL complex, two "perturbed" structures were also constructed, generated using SIMPLEX[9] minimization in torsion angle space, where all side chains within 10Å of Cα of residue 121 were included in the minimization. Another three complexes, but with His at position 121, were generated in a similar fashion. Although all six structures are identical in the main chain, the range of RMS differences between them is 0.3–0.5Å measured over all the atoms within a 10Å sphere centered on the Cα of residue 121. In

Table 3. Residues in the *F* regions of dual complexes

Lysozyme residues	Gly 26, Thr 118 to Arg 125
V_L domain	Ile 29 to Tyr 32, Phe 91 to Ser 93, Arg 96
V_H domain	Tyr 101

all six structures, the hydrogen bond between V_L Phe 91 and the side chain of residue 121 is intact.

Each of the six complexes was then converted into a dual complex by placing both His and Gln residues at position 121. For example, a His residue was built on top of Gln in the crystallographically determined structure by placing the main chain atoms of His coincident with the main chain atoms of Gln; the side chain of His was then model built in one of its preferred conformations, with the side chain atoms overlapping those of Gln as far as possible. Finally, a list was made of all residues having any atom within 5.0Å of either the His or Gln residues (Table 3). Formally, the dual complex consists of two identical copies of Fab and two copies of the antigen. The antigens differ only in that one has Gln (**GLN** complex) at 121 while the other has His (**HIS** complex). During the energy minimization, no interactions were calculated between the two copies of the antigen–antibody complexes. However, all atoms listed as being outside a 5.0Å limit referred to earlier were constrained to move in an identical fashion during the minimization; thus the energy minimization allows free, unconstrained trajectories only for residues in the immediate vicinity of residue 121. In these studies we did not employ a restrained region *R*, since our pilot studies for similar sized chains showed it to be unnecessary.

Initial 2000 steps of Cartesian space energy minimization were performed using the modified potential

$$V' = V + K_r(r - r_0)^2 \quad (6)$$

where the second term on the right-hand side is a restraining potential which forces the structure to lie close to the starting structure r_0. In the initial 500 steps, all atoms of the complexes were tethered to their starting positions using a restraining constant K_r of 1 kcal/mol/Å^2. In the following 1500 steps only the main chain N, Cα and C atoms were restrained; here the initial K_r was set at 100 kcal/mol/Å^2 and gradually decreased to 1 kcal/mol/Å^2 over the 1500 steps. A further 6000 steps of minimization were carried out without the use of the restraining potential; at this stage all components of the gradient vector had fallen below 0.05 kcal/mol/Å.

At this stage the dual complex was separated into its two component complexes. Each complex, the free Fab and free antigens, were then subjected to a further 2000 steps of minimization. Binding energies were obtained using Equation 5.

RESULTS

As a first step towards rationalizing the large decrease in the binding affinity of D1.3 for lysozymes with His at position 121, we considered the steric factors that may come into play when Gln is replaced by His. Computer graphics analysis shows that, despite the tight packing in the vicinity of Gln 121, Gln can be replaced by His with only minor displacements of the neighboring antibody side chains. Indeed, it is possible to replace Gln by His such that the Nϵ of His lies on top of Nϵ of Gln without causing any steric or stereochemical strain; thus the His side chain is not prevented from making the hydrogen bond to the main chain of V_L Phe 91. A number of other possibilities had also been considered. In particular, it was suggested that His 121 might be charged and therefore unstable in the hydrophobic pocket normally occupied by Gln. NMR studies[10] have established the pK_a's of the two His residues in turkey lysozyme to be 5.6 and 5.9 respectively, and thus less than 10% protonated at pH7. Recent crystallographic refinement of the turkey lysozyme structure in this laboratory[11] shows that His 121 occupies essentially the same space as Gln 121 in hen lysozyme; although the 2.2Å refinement of this structure is incomplete (crystallographic R-factor, R = 0.19), there are no obvious changes in the conformation of the polypeptide backbone and the antigenic determinant appears to be identical to that of hen lysozyme within experimental error.

The root mean square differences, for all main chain atoms, between the crystal structure and the various energy-minimized structures range from 0.9Å to 1.1Å and are comparable to the expected errors in the positional parameters of the crystallographically determined Fab/HEL complex (*ca.* 0.5Å); the RMS differences between the various energy-minimized structures have a range 0.7–0.9Å. These minimizations represent essentially a random sample of the potential energy surface in the vicinity of the crystal structure from which we propose representative conclusions can be drawn.

While the relationship between affinity constant and free energy $\Delta G = -RT \ln(k_{aff})$ only holds for the biochemical standard state, and does not allow calculation of absolute free energies for realistic concentrations[12] of antigen and antibody, it does hold for free energy differences and, therefore, the ratio of the affinity constants. In the case of D1.3/lysozyme Gln→His 121, this would give a free energy difference of approximately 10kcal/mol if k_{aff} for His 121 were 1.0. Experimental observation indicates only that this constant is below the limit of detectability of the assay.

The calculated energy differences between the **GLN** and **HIS** complexes are listed in Table 4. In all cases studied, lysozyme antigens with Gln at position 121 are predicted to bind more tightly than lysozymes with His at position 121. From the range of configurational space sampled in the six calculations, one may reasonably conclude that the potential energy surface in the vicinity (*ca.* 1Å) of the crystal structure for the **GLN** complex is lower than that for the **HIS** complex. The range of differences between the various energy minimized structures is 6–44 kcal/mol; the binding energy differences, using Equation 5 with free antigens and antibodies as reference states, have a range 7–21 kcal/mol.

The interaction energies between the antigens and the Fab in the complexes are two to three times larger than the binding energies obtained using Equation 5. The differences in interaction energies for **GLN** and **HIS** complexes reflect those for the binding energies, although there is no obvious correlation between the two. The differences for the interaction energies of residue 121 with rest of the complex have a range from 3 to 7 kcal/mol, smaller than that for differences in binding energies and differences in interaction ener-

Table 4. Total, binding and interaction energies for HIS and GLN complexes. Figures in brackets are differences in respective energy terms. Figures in column 2 are RMS differences for main chain atoms with respect to the crystal structure. Energies in kcal/mol; RMS differences in Å

Model	GLN/HIS	GLN/HIS		GLN/HIS	
1	1.14/1.13	−11421/−11415[a]	(6)	−139/−127[b]	(12)
		−301/−288[c]	(13)	−49/−46[d]	(3)
2	1.03/1.05	−11412/−11391	(21)	−127/−106	(21)
		−302/−291	(11)	−52/−45	(7)
3	1.08/1.08	−11395/−11379	(16)	−159/−147	(12)
		−266/−263	(3)	−51/−47	(4)
4	1.10/1.10	−11332/−11313	(19)	−165/−156	(9)
		−319/−292	(27)	−51/−45	(6)
5	1.11/1.12	−11280/−11262	(18)	−131/−114	(17)
		−279/−274	(5)	−50/−47	(3)
6	0.91/0.89	−11188/−11144	(44)	−89/−82	(7)
		−245/−235	(10)	−52/−47	(5)

[a]Total energies of complex
[b]Binding energies calculated using equation 5 of text
[c]Interaction energies between antigen and Fab
[d]Interaction energies between residue 121 and rest of complex

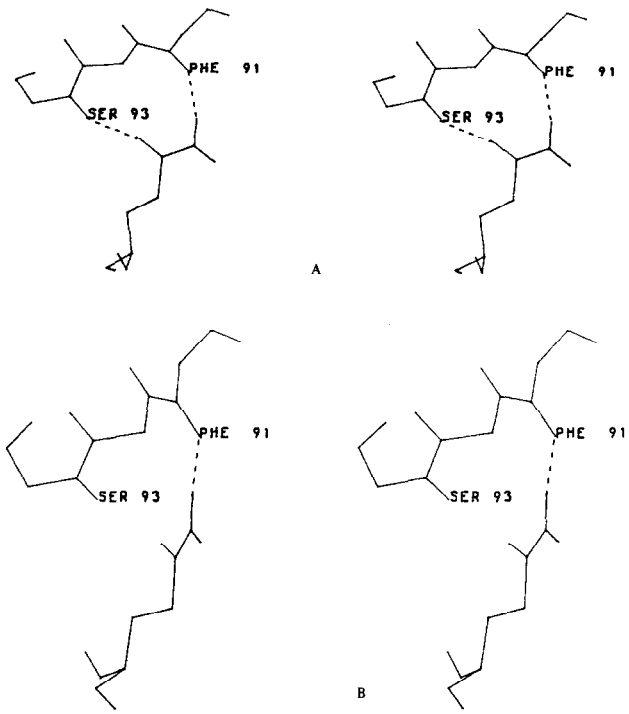

Figure 4. Contacts of Gln 121 with main chain of V_L *91–93. Gln forms two hydrogen bonds in energy-minimized structure of model 6. (b) contacts of Gln 121 with main chain of* V_L *of 91–93. Gln forms only one hydrogen bond (to carbonyl oxygen of* V_L *Phe 91). A short contact exists between* V_L *Ser 93 amide hydrogen and Oϵ of Gln in energy-minimized structure of model 1.*

gies between the antigens and Fab. Thus the interaction of residue 121 with the rest of the complex accounts for only part of the binding energies; small but significant differences are observed in the conformations of **GLN** and **HIS** complexes that further favor complex formation of D1.3 with Gln containing antigen over those containing His.

Detailed analysis of the energy-minimized structures shows that the largest RMS differences between the crystal structure and the minimized structures occur in the main chain of residues of V_L 91–93. Although the changes in the backbone are *ca.* 1Å, the amide hydrogen of V_L Ser 93 moves by *ca.* 2Å, concomitant with small but complementary movements of the side chain of Gln 121, to establish good contacts with carbonyl oxygen of Gln 121 side chain. Oϵ of Gln can form a good hydrogen bond (Figure 4a) with the main chain amide hydrogen of V_L Ser 93, but this requires a partial break in the hydrogen bond from Nϵ to the V_L Phe 91 carbonyl oxygen. In most energy-minimized complexes, this hydrogen bond between Gln 121 and V_L Ser 93 is not established but, rather, constitutes a good short contact (Figure 4b). Since we observe similar conformational changes in the main chain of V_L Ser 93 in both **HIS** and **GLN** complexes and in the uncomplexed Fab, the extra hydrogen bond with Gln 121 in the **GLN** complex is not the origin of these conformational changes.

DISCUSSION

We have used multi-start molecular mechanics simulations to model the binding differences of D1.3 antibody to lysozymes with Gln or His at position 121. Although computer graphics analysis shows that the His side chain can be accommodated into the binding pocket occupied by Gln in the crystal structure of the D1.3/HEL complex, the use of empirical potential energy functions shows that the Gln side chain makes better contacts with the neighboring antibody residues. The origin of the fine specificity of D1.3 for lysozymes with Gln rather than His at position 121 lies in small but significant differences in the contacts the two side chains make with the antibody. The analyses of the energy-minimized structures of the **HIS** and **GLN** complexes show that His and Gln side chains make similar contacts with the

surrounding antibody residues. There exist two possible explanations for the differences in the binding energies: (a) both His and Gln side chains form one good hydrogen bond to V_L Phe 91 of the antibody but the Gln side chain forms many more favorable van der Waals contacts with the neighboring antibody atoms; (b) Gln can form two hydrogen bonds, one with V_L Phe 91 and another with V_L Ser 93, whereas His can form just the one hydrogen bond to V_L Phe 91. Examination of the crystallographically determined structure of the D1.3/HEL complex favors (a), however, possibility (b) cannot be ruled out at the current resolution of the diffraction data. Our recent refinement of the D1.3/HEL complex at 2.8Å resolution, using the simulated anealing method, does indeed suggest that a short contact exists between Gln Oε and the amide hydrogen of V_L Ser 93.

Although in these studies we did not take into account solvation of free His and Gln containing antigens, estimates of free energy of transfer from solvent to a hydrophobic environment using accessible surface areas[13] would suggest that the complex formation would be favored for His containing antigens over those containing Gln by *ca.* 0.5 kcal/mol (*c.f.*, 0.8 kcal/mol difference from free energy of transfer experiments). Clearly, this is offset by enthalpic terms on complex formation. We also used the minimum perturbation method of Shi *et al.*[14] to calculate the binding energy differences. The use of this method with only the side chain torsion angles of the Gln and His as variables suggested that His would be favored over Gln by *ca.* 10 kcal/mol. Extending the search to include all side chains within a 10Å sphere of the Cα of residue 121 did not give significantly different results. Since the probable errors in the atomic positions exceed 0.5Å for the D1.3/HEL complex at 2.8Å resolution, molecular mechanics studies with all degrees of freedom were required to establish the origin of the sequence specificity of D1.3.

The SIMMIN method has been successfully used to establish the basis of much higher affinity of the antibody D1.3 for lysozymes with Gln at position 121 over lysozymes with His at 121. Our recent studies on the binding of two inhibitors, methotrexate and trimethoprim, to vertebrate (Mouse) and bacterial (*L. Casei*) DHFR's has established the SIMMIN technique to be of much wider applicability than reported here. The results described in this article are encouraging, and suggest that molecular mechanics studies can be used to study a wide range of molecular recognition processes, and that such studies can be used to make useful predictions about the effects of site-directed mutations, and also to determine the relative binding energies between a number of similar drugs to the same protein molecule.

ACKNOWLEDGEMENTS

This work was funded by a grant from the EEC. The author is grateful for the coordinates of the D1.3/HEL complex from Dr. S. E. V. Phillips (University of Leeds) and Dr. R. J. Poljak (Institute Pasteur, Paris).

REFERENCES

1 Winter, G. and Fersht, A. R. *Trends Biotech.* 1984, **2**, 115
2 Blundell, T. L. and Sternberg, M. J. E. *Trends Biotech.* 1985, **3**, 228
3 Amit, A. G., Mariuzza, R. A. Phillips, S. E. V. and Poljak, R. J. *Science* 1986, **233**, 747
4 Boyd, D. B. and Lipkowitz, K. B. *J. Chem. Education* 1982, **59**, 269
5 Sibanda, B. L. and Thornton, J. M. *Nature* 1985, **316**, 170
6 Matsumura, M., Becktel, W. J. and Mathews, B. W. *Nature* 1988, **334**, 406
7 Janin, J., Wodak, S., Levitt, M. and Maigret, B. *J. Mol. Biol.* 1978, **125**, 357
8 White, D. N. J. and Morrow, C. *Computers and Chemistry* 1979, **3**, 33
9 Nelder, J. A. and Mead, R. *Comp J.* 1965, **7**, 308
10 Somers, W. S. and Swainson, A. G. University of Leeds, 1987, unpublished results
11 Parsons, M. R. University of Leeds, 1988, PhD Thesis
12 Creighton, T. E., In *Proteins*, Freeman and Co., New York, NY, 1986
13 Eisenberg, D. R. and McLachlan, A. D. *Nature* 1986, **319**, 199
14 Shi, H.-L., Brady, J. and Karplus, M. *Proc. Nat. Acad. Sci. USA* 1985, **82**, 1697