(83) The two-ring energy minima of proline approximately correspond to a $C_s$ symmetric $C^{\gamma}_{exo}$ puckering and a $C_2$ symmetric $C^{\beta}_{exo}$–$C^{\gamma}_{endo}$ puckering.
(84) The term "flexibility" as used here refers to the narrow range of dihedral angles accessible to the pyrrolidine ring and not to a discrete puckering.
(85) Conformational energy maps for helical homopolymers are not applicable in the calculation of average chain properties.
(86) See ref 14 and 90 for a more complete review of the pre-1972 proline conformational energy maps and their predicted characteristic ratios.
(87) L. Mandelkern and W. Mattice, *Conform. Biol. Mol. Polym., Proc. Int. Symp., 5th* (1973).
(88) The all-trans ($\Psi$ near 160°) polypeptide chain of proline is not an ordered extended helix but is in a statistical conformation due to the breadth of this minimum.[14,15]
(89) H. Strassmair, J. Engel, and S. Knof, *Biopolymers*, **10**, 1759 (1971).
(90) D. A. Torchia, *Macromolecules*, **4**, 440 (1971).
(91) D. A. Torchia, *Macromolecules*, **5**, 566 (1972).
(92) T. F. Koetzle, M. S. Lehmann, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **29**, 231 (1973).
(93) A. E. Tonelli, *J. Am. Chem. Soc.*, **95**, 5946 (1973).
(94) B. P. Roques, C. Garbay-Jaureguiberry, S. Combrisson, and R. Oberlin, *Biopolymers*, **16**, 937 (1977).
(95) B. P. Roques, S. Combrisson, and F. W. Wehrli, *Tetrahedron Lett.*, 1047 (1975).
(96) There are semantic difficulties with the designations form I and form II, which have led to some confusion. These designations refer to the ordered structures found in the solid state, as determined by standard crystallographic methods, for either the all-cis or all-trans chains, respectively. They do not refer to the all-cis or all-trans chain in solution, where even for the predominantly all-trans chain an ordered structure is not observed.[14]
(97) $^1$H NMR proved unsuccessful for this purpose because the cis $\alpha$-proton resonance overlaps the trans resonance in organic solvents.

# Conformational Energy Calculations of Enzyme–Substrate and Enzyme–Inhibitor Complexes of Lysozyme. 2. Calculation of the Structures of Complexes with a Flexible Enzyme[1]

**Matthew R. Pincus[2] and Harold A. Scheraga***

*Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.*
*Received February 12, 1979*

ABSTRACT: Conformational energy calculations were carried out to investigate the most favored binding modes of oligomers of β-D-*N*-acetylglucosamine (GlcNAc) to the active site of lysozyme. Both the substrate and the side chains of the enzyme were allowed to undergo conformational changes and relative motions during energy minimization. It was found that, regardless of whether the side chains of the enzyme were held rigidly or were allowed to move, $(GlcNAc)_6$ with standard geometry had a clear preference for binding to the active site cleft with its last two residues on the "left" side of the cleft region. This region contains such residues as Arg 45, Asn 46, and Thr 47, compared with the "right" side of the cleft which contains such residues as Phe 34 and Arg 114. This result was obtained irrespective of the location of the fourth residue, i.e., irrespective of how deeply it was buried in the active site cleft. Our earlier calculation (with the enzyme held rigidly), starting with the model-built structure (from the literature), with residue 4 buried deeply in the D site on the right side of the cleft and having a half-chair conformation, led to a high-energy structure. However, when the side chains of the enzyme were allowed to undergo changes of conformation, the energy of this structure was lowered significantly because of favorable contacts on the right side of the cleft. Nevertheless, the conformational energy of this structure was still higher (by about 5 kcal/mol) than that of the most stable hexamer having standard geometry (i.e., without distortion of residue 4) and situated on the left side of the cleft. In addition, a hexamer with standard geometry could bind with the same low conformational energy as that of the energy-minimized model-built structure on the right side without distortion of residue 4. However, the conformational energy of this structure could be lowered by ~14 kcal/mol, if distortion of the fourth residue, binding in the D site, were allowed. The conformational energy of this latter structure was the lowest of all those found in the energy search, if the strain energy for the fourth residue were ignored. If it were taken into account, however, the conformational energy of this species would probably increase so that it would become equal to or higher than that for the lowest energy hexamers binding to the left side of the active site. The finding of three low-energy structures, an undistorted mode on the left side of the active site, an undistorted mode of higher conformational energy on the right side, and a distorted mode on the right side, correlates well with recent experimental kinetic data on the binding of $(GlcNAc)_6$ to lysozyme. Finally, the presence of the *N*-acetyl group on GlcNAc provides sufficient interactions to account for the fact that GlcNAc oligomers bind to lysozyme, whereas glucose oligomers bind with much lower affinity.

In a series of publications,[3-5] we have used conformational energy calculations to compute the structures of complexes of oligomers of β-D-*N*-acetylglucosamine (GlcNAc) with the active site of hen egg white lysozyme. The ultimate purpose of this work is to determine the interactions involved in the specific binding of substrates and inhibitors to this enzyme.

Until now, under the assumption that the enzyme has features "built into" its *native* structure that allow for recognition of the substrate, we allowed the substrates to undergo rigid body and internal motion, but we kept the enzyme rigidly fixed. The calculations were carried out in four distinct steps:

1. The low-energy conformations of *isolated* oligomers and polymers of GlcNAc and *N*-acetylmuramic acid (MurNAc) were determined.[3]
2. Using a fragment of a disaccharide molecule to map the conformational space, the low-energy regions for binding $(GlcNAc)_2$ to the *rigid* active site of lysozyme were identified.[4]
3. The energies of a representative set of conformations of $(GlcNAc)_2$ (determined in step 1) in *all* low-energy binding regions (identified in step 2) were minimized,[4] with the conformation of the enzyme held rigidly fixed.
4. The lowest-energy structure of the disaccharide in the complex, obtained in step 3, was lengthened by adding

0024-9297/79/2212-0633$01.00/0 

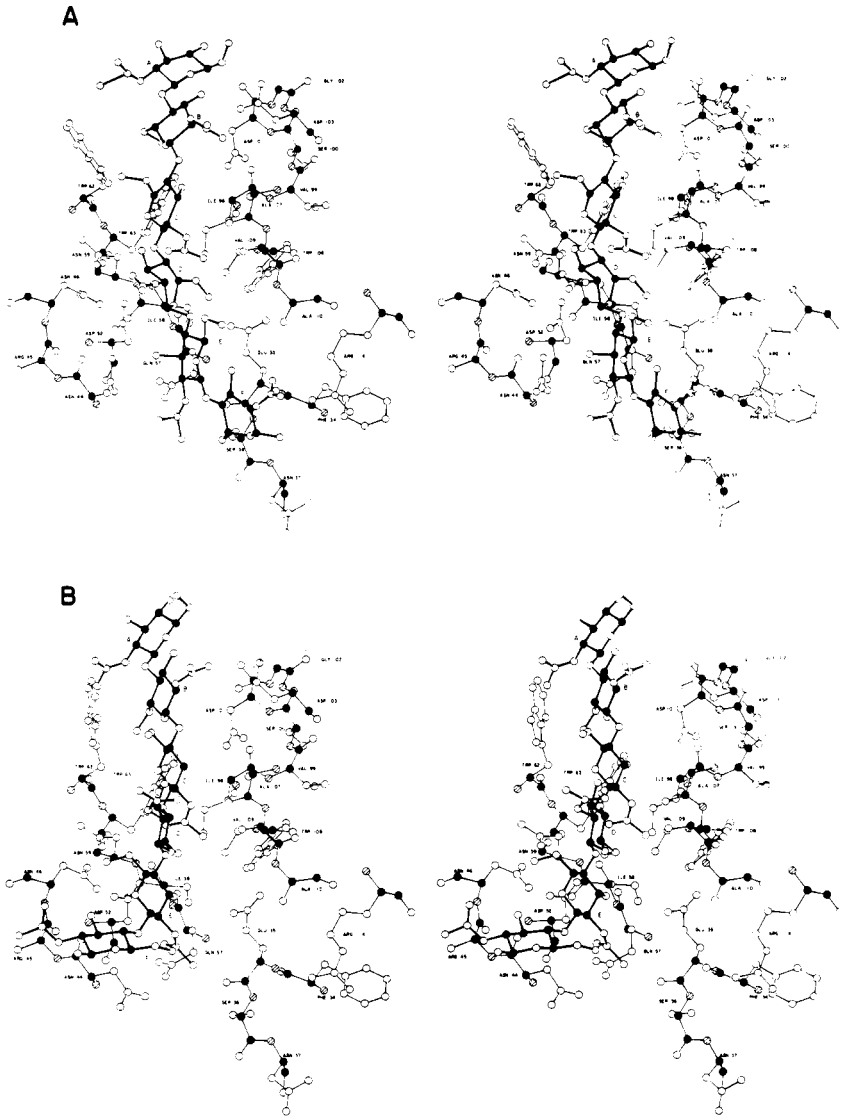


**Figure 1.** Stereoviews of two stable binding modes of $(GlcNAc)_6$ to the active site of lysozyme, one involving the left lower cleft and the other involving the right lower cleft. (A) Energy-minimized model-built structure where the fifth and sixth residues bind to the *right* side of the cleft (conformer 1 of Table II). This structure has nonstandard bond lengths and bond angles and a half-chair conformation for the D residue. (B) Lowest-energy structure of $(GlcNAc)_6$ with standard bond lengths and bond angles (conformer 3 of Table II). The fifth and sixth residues of this conformer make contacts with the lower *left* side of the active site cleft; this binding mode is the same as that described for the rigid enzyme (Figure 1 of ref 5) except that, here, both the substrate and the side chains of the enzyme were allowed to move to optimize the contacts between them to a greater extent.

GlcNAc units to its reducing and nonreducing ends (in conformations found to be favorable in step 1), and the energies of these oligosaccharides in the active site of the rigidly-fixed enzyme were minimized.[4,5] Step 4 was repeated (each time adding a GlcNAc unit to the lowest-energy structure obtained in the previous step) for the tri-, tetra-, penta-, and hexasaccharides. The energy of the hexasaccharide–lysozyme complex could not be lowered any further when we added more GlcNAc units to the hexasaccharide. This result is in good agreement with experimental binding studies which show that lysozyme binds to a maximum of six GlcNAc residues.[6]

The structure of the lowest-energy hexamer–enzyme complex had several distinguishing features, viz.:

1. The binding conformation of the first three residues in the active site was quite close to that of the energy-minimized[4] X-ray structure[7] of the $\delta$-lactone of $(GlcNAc)_4$ in sites A, B, and C.

2. The fourth GlcNAc residue, in the *chair* form, bound with a high affinity to a site somewhat removed from the cleft, between residues Glu 35 and Asp 52. In ref 5, we noted that there were two distinct binding modes for the fourth residue of a tetramer in the chair form (whose first three residues were bound to sites A–C), one in which the fourth residue was bound near the surface of the enzyme and the other in which this residue was "pulled in" closer to but still removed from the groove between Glu 35 and Asp 52. In the latter binding mode, it was possible to add a fifth GlcNAc residue and obtain a stable complex. Both of these calculated binding modes were supported by recent experimental evidence.[8]

3. The fifth and sixth residues of the hexamer bound to a region of the active site on the opposite side of the cleft from the E and F sites proposed[6] on the basis of model-building studies, i.e., near residues Arg 45, Asn 46, and Thr 47 rather than in the vicinity of Phe 34 and Arg 114.

4. Energy minimization[5] of the structure proposed[6] on the basis of model building yielded a conformation of low energy, but one of significantly higher energy than that for our calculated hexamer–enzyme complex.

All of the calculations performed thus far[4,5] were based on structures obtained by lengthening the lowest-energy structure of $(GlcNAc)_2$ in the active site. However, the

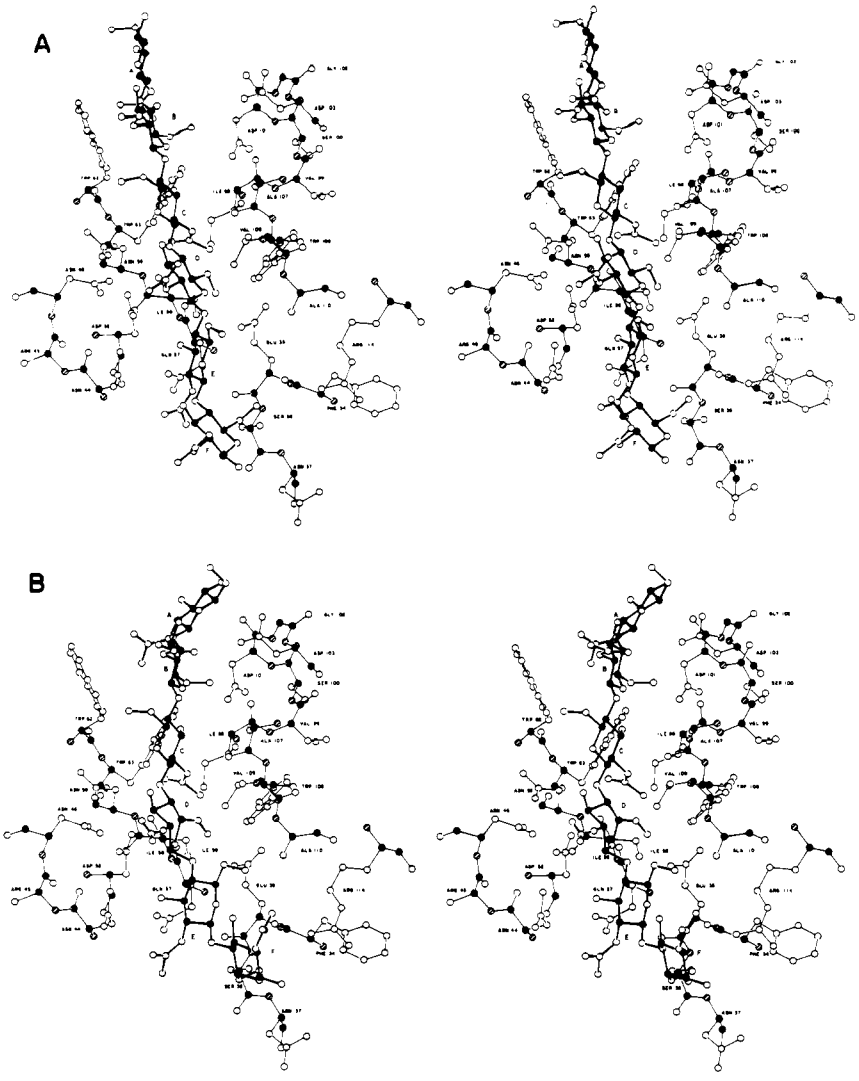


**Figure 2.** Two modes of binding for $(GlcNAc)_6$ to the active site of lysozyme in which the fifth and sixth residues contact the lower right side of the cleft. (A) Energy-minimized structure with standard bond lengths and bond angles and no distortion of the D residue from the chair conformation (conformer 7, Table III). (B) Energy-minimized structure (conformer 9, Table III) with standard bond lengths and bond angles for all residues except residue 4 which was held in a half-chair conformation as in ref 6. The bond lengths and bond angles of this fourth residue are the same as those for the fourth residue of Figure 1A.

possibility existed that other minimum-energy dimer structures (though of significantly higher energy) could be extended to higher oligomers of low energy when bound to other regions of the enzyme. Therefore, in this paper, we now consider oligomers formed from other low-energy structures of the fragment disaccharide in the *rigid* active site.

In addition, we now remove the restriction that the enzyme remain rigid in forming the complex. In this way, it is possible to remove unfavorable contacts between the substrate and the rigid enzyme. Thus, the atoms of the substrate and those of the *side chains* in the active site of the enzyme are allowed to move during energy minimization. With this introduction of flexibility in the enzyme, we re-examine the binding conformations of oligomers of GlcNAc with the enzyme and assess the possibility of distortion of the fourth residue[6] (in the D site) of $(GlcNAc)_6$ bound to lysozyme in order to obtain optimal binding in the E and F sites.

## Methods

**A. Location of Binding Sites of the Rigid Enzyme.** In the previous calculations,[4,5] starting conformations were selected from those fragment–enzyme complexes having energies within 7 kcal/mol of the lowest. To remove the arbitrariness of this cut-off value of the energy, *all* low-energy fragment–enzyme complexes (generated previously[4] by a gridding procedure) were used here as starting points for minimization of the energies of complexes of *complete* dimers in the active site of the *rigid* enzyme. The only constraint on the starting points was that the $(GlcNAc)_2$ molecule lie within the active site region, i.e., $Z < 15$ Å and, for all $Z > 6$ Å, $Y > 2$ Å (see ref 4 for discussion of axes and coordinate system). For the lower values of $Z$ (viz., $Z < 3$ Å), a finer gridding procedure was used with a fragment monomer because no very low-energy dimer minima had been found earlier.[4] Each of the complete monomer conformations, within 10 kcal/mol of the lowest-energy monomer after energy minimization, was extended in the usual way[4] to a dimer.

**B. Minimization of the Energy of Enzyme–Substrate Complexes with a Flexible Enzyme.** The energies of the enzyme–substrate complexes (for di-, tri-, tetra-, penta-, and hexamers) were minimized by allowing for flexibility in *both* the substrate and *side chains* of the enzyme (achieved by rotation about single bonds) and for rigid-body translation and rotation of the substrate relative to the enzyme. The minimization procedure was the same as that described previously[4,5] except that, now, the side chains of the enzyme were allowed to move. All confor-

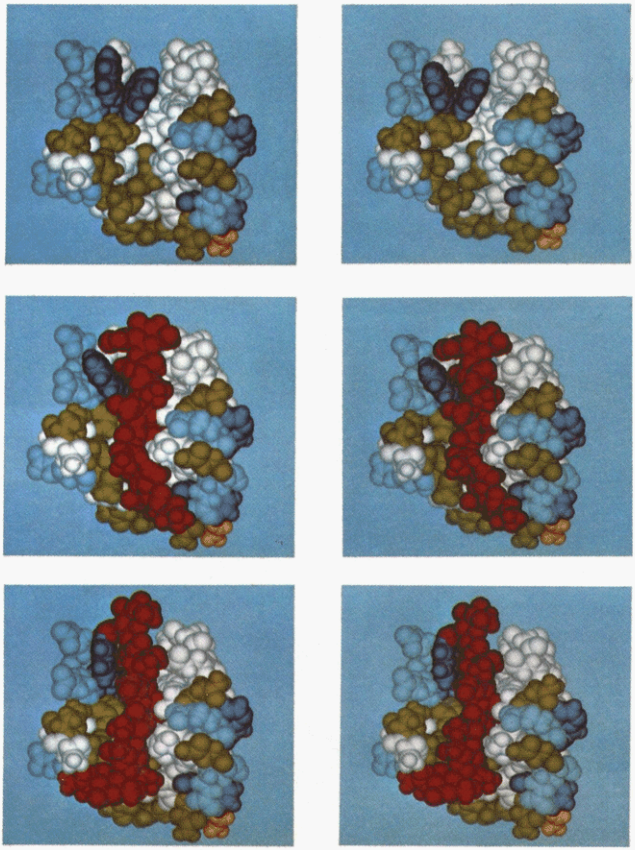

**Figure 3.** Stereoviews of space-filling models of: (top) the active site of native lysozyme; (middle) energy-minimized model-built hexamer bound to the active site (as in Figure 1A); and (bottom) lowest-energy undistorted hexamer bound to the active site with a preference for sites E and F on the left side (as in Figure 1B). The color code employed is as follows: (1) Asn and Gln, green; (2) Trp and Phe, blue; (3) Arg, turquoise; (4) Lys, yellow; (5) all other enzyme residues, white; and (6) substrate, red. The orientation of each view is the same as that for the structures in Figures 1 and 2. In the top figure, that of the native enzyme, Trp 62 and 63 can be seen clearly while, in the lower two figures, the substrate obscures the view of Trp 63 because this residue lies wholly "behind" the substrate. Trp 108 is not well visualized since the indole ring of this residue points directly back into the active site cleft. Color prints were prepared by Dr. Richard J. Feldman, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Md.

mational energies of the enzyme were calculated with the program UNICEPP.[9] The total conformational energy, $E_{TOT}$, for any given conformation of the enzyme–substrate complex was calculated as the sum:

$$E_{TOT} = E_{SUB} + E_{ENZ} + E_{INT} \quad (1)$$

where $E_{SUB}$ and $E_{ENZ}$ are the internal conformational energies of the substrate and enzyme, respectively, and $E_{INT}$ is the interaction energy between the enzyme and substrate. For convenience, we define $E_{CONF}$ as the sum of $E_{SUB}$ and $E_{INT}$; this energy is analogous to $E_{COM}$ in ref 4 and 5. Energy minimization was performed by the method of Powell.[10]

As before,[3-5] the effect of solvent was not introduced explicitly into the calculations. Solvent may not affect the results of the calculations significantly because they are compared primarily with the solid-state (crystal) structure of lysozyme bound to inhibitors [such as $(GlcNAc)_3$]. It is known that oligomeric inhibitors and substrates of lysozyme bind with high affinity to lysozyme in the *crystal*[6] where the effects of solvent are expected to be less influential than they are in bulk solution and where direct contacts between enzyme and substrate would be expected to predominate in determining binding conformations and affinities. Nonetheless, it is possible that, since the crystal of the enzyme is substantially solvated, solvent may still influence the structures of enzyme–substrate complexes.

Since it is possible that the moving side-chain atoms of the active site of the enzyme can make contacts with atoms of residues that lie outside of the active site, it was necessary to include in the energy calculations all residues that "surround" the active site residues. To determine which extra residues were to be included in the calculations, the procedure employed was similar to that described previously.[11] For each active site residue whose side chain was to be moved in the energy minimization procedure, a convenient atom close to the center of mass of the residue was selected. The maximum distance of this reference atom from the most distant side-chain atom of the same residue was then determined. The van der Waals radius

of this most-distant atom was added to this distance to obtain an effective radius for the side chain. For example, the $C^{\beta}$ of a glutamic acid residue is close to the center of mass of this residue. The side-chain atom farthest from it is the H atom of the $\gamma$-COOH group when the side chain is placed in the fully-extended conformation. The effective radius for the glutamic acid side chain is then the sum of the distance from $C^{\beta}$ to the carboxyl H, plus the van der Waals radius of the H atom. Besides the active site residues,[4] additional residues were included if any of their atoms lay within the effective radii of the side chains that were to be allowed to move in the calculations. The following 55 residues were thus included: Ala 31 to Asn 37, Asn 39, Gln 41 to Cys 64, Arg 73, Leu 75, Cys 76, Cys 80, Cys 94, Ala 95, and Ile 98 to Arg 114.

The following procedure was then used for minimizing the energy of complexes in which the enzyme was allowed to move:

1. For each starting conformation of a complex, all unfavorable contacts (i.e., those with energy ≥1 kcal/mol) between enzyme and substrate were determined. These contacts were ordered from highest to lowest in energy.

2. The energy of the complex was minimized by changing first those variables which would move the smallest number of atoms possible while removing the bad contacts. In general, if any of the atoms of a side chain of the enzyme made bad contacts with the side chain and/or backbone atoms of the substrate, the side chain of the enzyme would be moved first; if side-chain atoms of the substrate were also involved, these atoms would be moved next. The variables that moved the atoms of a side chain were changed in an order such that those that moved the smallest number of atoms were changed first, followed by those that moved an increasing number of atoms. The unfavorable contacts were removed in an order from highest to lowest energy contacts.

3. In each minimization, all of the substrate variables were allowed to change as were those of all enzyme side chains with *which the substrate made contacts.* When the energy of the complex converged to within 0.5 to 1.0 kcal/mol, a second step was carried out in which the structure obtained from the first minimization was subjected to further minimization. In this step, *all* of the side chains of the residues at the active site that were held fixed in the first step but which lay within the effective radii of the moving residues were now allowed to move together with all of the atoms moved in the first step. In this second step, no unfavorable contacts existed for any structure, and the variables of the substrate were moved first, followed by those of the enzyme.

For several different starting conformations of $(GlcNAc)_6$ bound to the active site, two sets of test minimizations were performed, one in which the substrate interacted with all 55 residues, i.e., the active-site residues *and* the surrounding residues, and the other in which it interacted only with the original 41 active-site residues[4,5] but *not* with the surrounding residues. Though the energy of the complex, $E_{TOT}$, was lower (because of more interactions) in the former case, i.e., when the substrate interacted with *all* 55 residues, the final conformation of the energy-minimized complex was exactly the same in both cases. Therefore, all subsequent minimizations were performed by allowing the substrate to interact only with the 41 active site residues and not with the surrounding residues.

## Results and Discussion

**Low-Energy Complexes (Rigid Enzyme).** We consider first all possible low-energy complexes of $(GlcNAc)_2$ with the *rigid* enzyme. Previously,[4] we obtained conformer

**Table I**
**Low-Energy Conformations of Oligomers of GlcNAc at the Active Site of the Rigid Enzyme[a]**

| saccharide | ref residue | rigid body variable[b] | | | | | | inter-ring dihedral angles[c] | | binding sites[d] | $E_{SUB}$[e] | $E_{INT}$[e] | $E_{TOT}$[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $X$ | $Y$ | $Z$ | $\alpha$ | $\beta$ | $\gamma$ | $\phi$ | $\psi$ | | | | |
| 1. $(GlcNAc)_2$: lowest energy min, ref 4 | 1 | 7.90 | 6.55 | 6.07 | 44.3 | −32.8 | −3.9 | −64.2 | 111.9 | C–D | −18.7 | −45.5 | −64.2 |
| 2. $(GlcNAc)_2$: new min | 1 | 7.54 | 6.12 | 5.28 | 42.2 | −16.8 | −9.7 | −80.6 | 98.6 | C–D | −18.5 | −43.3 | −61.8 |
| 3. $(GlcNAc)_2$: new min | 1 | 7.52 | 5.99 | 10.51 | −23.3 | 1.8 | 181.7 | −97.4 | 109.8 | B–C′ | −17.7 | −44.6 | −62.3 |
| 4. $(GlcNAc)_2$: new min | 2 | 7.88 | 4.97 | 2.74 | 4.5 | 29.4 | −17.2 | −59.4 | 114.1 | C′–D′ | −15.7 | −48.9 | −64.6 |
| 5. $(GlcNAc)_3$: lowest-energy structure as in ref 4 | 2 | 7.48 | 6.46 | 6.15 | 48.7 | −24.0 | −6.9 | −74.4 | 131.4 | B–C | −24.2 | −63.2 | −87.4 |
| | | | | | | | | −69.4 | 106.2 | C–D | | | |
| 6. $(GlcNAc)_3$: extension of conformer 2, nonreducing end | 2 | 7.63 | 6.40 | 5.88 | 47.2 | −22.2 | −2.0 | −70.4 | 123.2 | B–C | −26.8 | −61.9 | −88.7 |
| | | | | | | | | −88.4 | 94.6 | C–D | | | |
| 7. $(GlcNAc)_3$: extension of conformer 2, reducing end | 1 | 7.58 | 6.18 | 5.18 | 43.7 | −16.8 | −8.5 | −79.3 | 97.6 | C–D | −25.4 | −62.3 | −87.7 |
| | | | | | | | | −100.3 | 100.7 | D–E | | | |
| 8. $(GlcNAc)_4$: extension of conformer 7 | 2 | 7.69 | 5.87 | 5.18 | 43.1 | −22.1 | −7.2 | −99.5 | 70.4 | B–C | −31.9 | −80.4 | −112.3 |
| | | | | | | | | −85.4 | 102.6 | C–D | | | |
| | | | | | | | | −97.4 | 104.8 | D–E | | | |

[a] All terms used in this table are defined in ref 4 except $E_{TOT}$ (=$E_{SUB}$ + $E_{INT}$ here, since $E_{ENZ}$ = 0 because the enzyme is rigid). [b] $X$, $Y$, $Z$ in Å; $\alpha$, $\beta$, $\gamma$ in deg. [c] $\phi$ and $\psi$ in deg. [d] A prime after the site designation indicates a binding site distinct from that given in ref 4 and 5. [e] In kcal/mol.

1 of Table I as the only low-energy dimer in the enzyme–substrate complex. However, by selecting additional (higher-energy) starting points, we obtain three additional low-energy dimers (with energies within 5 kcal/mol of that of conformer 1), viz., conformers 2–4 of Table I.

It may be noted that conformers 1 and 2 are similar in position and orientation, except that conformer 2 is almost 1 Å further displaced down the $Z$ axis. Conformer 4, though having the same energy as that of conformer 1, is buried deeply in the active site in the D region and cannot be extended further down into the E and F regions. Noncatalytic, deep binding of a GlcNAc residue in the D site has been observed for $(GlcNAc)_3$ bound to iodine-oxidized lysozyme.[12]

When dimer conformers 3 and 4 of Table I were extended to trimers by adding a residue either to the reducing or nonreducing end, none of the resulting trimers had as low an energy as that of the lowest-energy trimer (conformer 5) of Table I, the closest in energy being 8 kcal/mol higher in energy.

When conformer 2 was extended, however, two low-energy minima were obtained, one by adding a residue to the nonreducing end and the other by adding a residue to the reducing end. These structures are listed as conformers 6 and 7 in Table I. Conformer 6 is essentially the same as conformer 5, in sites B, C, and D. Conformer 7 is a new trimer structure which extends into the E region.

Because conformer 7 appeared to be a favorable trimer, it was extended to larger structures. When a residue was added to its nonreducing end and the energy was minimized, conformer 8 of Table I (with the added residue in the B site) was obtained. However, addition of a residue to the reducing end of this tetramer resulted in little further stabilization. Thus, of all starting conformations of $(GlcNAc)_2$ examined in our energy search with a rigid enzyme, only one structure (conformer 1 of Table I) was found which had a low energy and which could be extended to a very low-energy hexamer bound to the active site, viz., the hexamer structure reported in ref 5.

**Low-Energy Complexes (Flexible Enzyme). A. Model-Built Structure.**[6] Previously,[5] using a rigid enzyme, we minimized the energy of the structure proposed on the basis of model-building studies[6] and found that the resulting structure had a much higher energy (by 35 kcal/mol) than our lowest-energy enzyme–hexamer complex.[5] Also, the two terminal residues of the hexamer in our lowest-energy structure bound to a region of the active site cleft that was on the opposite side from that proposed on the basis of model building.[6]

Because the starting conformation for the model-built structure contained a number of unfavorable contacts between the substrate and the enzyme (especially between the $N$-acetyl group of residue 5 of GlcNAc and the side chain of Asn 44), it was possible that such contacts could be relieved effectively by allowing the relevant side chains of the enzyme (in addition to the atoms of the substrate) to move. When the minimization was performed with the flexible enzyme, the structure shown in Figure 1A and whose conformation is listed in Table II (conformer 1) was obtained. From Table II, it is apparent that the conformational energy of the complex is now greatly lowered ($E_{CONF}$ = −151.9 kcal/mol) as compared with the result ($E_{CONF}$ = −118.0 kcal/mol) obtained with the rigid enzyme.[5] In fact, $E_{CONF}$ is now comparable with that of our lowest energy complex of a hexamer with a *rigid* enzyme (see conformer 3 of Table 1 of ref 5), provided that one ignores the strain energy involved in distorting the fourth residue to the half-chair conformation.

To obtain conformer 1, major conformational changes of the side chains of the enzyme were required, especially those of Glu 35, Asn 44, Asp 52, and Trp 62. The side chain of this last residue moved toward the substrate upon binding, effectively "closing in" on it in a manner observed directly from Fourier difference maps for the crystal structure of $(GlcNAc)_3$ bound to lysozyme.[6] In addition, besides conformational changes of –OH groups and other side chains of the substrate, there was an overall movement of the energy-minimized model-built structure by 0.5 Å out of the cleft of the active site (compare the Y position of conformer 1 with that of conformer 2, the starting conformation, i.e., model-built structure, in Table II).

The contacts made between enzyme and substrate in the energy-minimized structure involved several good hydrogen bonds. In particular, a good hydrogen bond formed between the NH of the $N$-acetyl group of the GlcNAc residue in site C and the backbone C=O of Ala 107 and between the C=O of the same GlcNAc residue and the backbone NH of Asn 59. These hydrogen-bonding interactions were also present in our lowest energy hexamer. In addition, strong hydrogen bonds formed between the $CH_2OH$ groups of residue 4 of the substrate and the backbone C=O of Gln 57 in the enzyme and between the O3H (H16 in ref 3) of residue 5 of the substrate and the side chain C=O of Gln 57. Other, weaker hydrogen-bonding interactions were observed, such as those between the nonreducing OH group of residue 1 of the substrate and the side-chain carboxyl group of Asp 101, the $CH_2OH$ of residue 6 and the backbone C=O of Phe 34, and the reducing OH of residue 6 and the side-chain terminal nitrogen atoms of Arg 114. A strong hydrogen bond existed between the C=O of residue 6 and a terminal NH of Arg 114. Further, the carboxyl group of Glu 35 came within 2.8 Å of the bridge oxygen between residues 4 and 5 while the carboxyl group of Asp 52 came within 3.2 Å of the Cl atom of residue 4. Both are, therefore, within reasonable distances to effect catalysis. All of these interactions have been either observed directly in the crystal structure of the trimer bound in sites A, B, and C or postulated on the basis of model building.[6]

It should be emphasized that the conformational energy of this structure is in reality much higher than that shown in Table II (conformer 1) because of the large strain energy involved in distorting the ring of the fourth residue to the half-chair conformation. While our calculations did not allow for alteration of bond lengths and bond angles, and hence did not allow for computation of strain energy, it has been estimated that the energy of distortion to form the half-chair is anywhere from 5 to 16 kcal/mol.[13,14] Further, it should be noted that the bond lengths and bond angles of even the "undistorted" residues of the model-built structure departed significantly from standard values based on a large number of crystal structures.[3,15] For example, one inter-ring C–O–C bond angle was 131°, a significant departure from the standard value of 116.5°; this would add additional strain energy to this conformation.

Nonetheless, the energy of the model-built structure was lowered significantly relative to that of the complex involving a rigid enzyme. This result indicated the possible existence of the above mode of binding as an alternative to the one calculated in ref 5, as well as the possible existence of other, different modes of binding of $(GlcNAc)_6$ to lysozyme.

**B. Lowest-Energy Complex.**[5] Since the results in section A indicated the possible existence of binding modes for a hexamer other than the low-energy one computed

**Table II**
**Oligomer Conformations with a Flexible Enzyme[a]**

| saccharide[b] | rigid body variables[c] | | | | | | inter-ring dihedral angles[d] | | binding sites | $E_{SUB}$[e] | $E_{INT}$[e] | $E_{ENZ}$[e] | $E_{TOT}$[e] | $E_{CONF}$[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $X$ | $Y$ | $Z$ | $\alpha$ | $\beta$ | $\gamma$ | $\phi$ | $\psi$ | | | | | | |
| 1. minimized model-built structure | 6.98 | 5.55 | 4.95 | 38.7 | −13.3 | −3.3 | −80.0 | 106.4 | A–B | −42.6 | −109.3 | −1329.1 | −1481.0 | −151.9 |
| | | | | | | | −56.1 | 99.2 | B–C | | | | | |
| | | | | | | | −125.3 | 123.6 | C–D | | | | | |
| | | | | | | | −142.1 | 83.8 | D–E | | | | | |
| | | | | | | | −94.7 | 92.3 | E–F | | | | | |
| 2. starting conformation for 1 | 7.01 | 6.02 | 4.80 | 39.0 | −10.3 | −0.2 | −77.3 | 94.7 | A–B | 48.7 | $3.6 \times 10^3$ | $-1.3 \times 10^3$ | $2.3 \times 10^3$ | $3.6 \times 10^3$ |
| | | | | | | | −63.6 | 83.1 | B–C | | | | | |
| | | | | | | | −126.8 | 124.1 | C–D | | | | | |
| | | | | | | | −154.0 | 83.0 | D–E | | | | | |
| | | | | | | | −87.8 | 102.8 | E–F | | | | | |
| 3. final structure for our lowest-energy $(GlcNAc)_6$ obtained originally with a rigid enzyme | 7.60 | 6.52 | 6.00 | 50.6 | −24.7 | 2.7 | −75.8 | 112.5 | A–B | −52.0 | −105.2 | −1327.4 | −1484.6 | −157.2 |
| | | | | | | | −81.5 | 131.4 | B–C | | | | | |
| | | | | | | | −98.0 | 83.0 | C–D | | | | | |
| | | | | | | | −90.5 | 109.9 | D–E | | | | | |
| | | | | | | | −43.0 | 116.3 | E–F | | | | | |
| 4. continuation of tetramer outside of cleft to a pentamer | 7.63 | 6.36 | 6.12 | 47.3 | −24.2 | −1.5 | −80.0 | 108.5 | A–B | −46.7 | −83.4 | −1332.0 | −1462.1 | −130.1 |
| | | | | | | | −74.2 | 127.2 | B–C | | | | | |
| | | | | | | | −68.4 | 106.7 | C–D | | | | | |
| | | | | | | | −104.2 | 81.2 | D–E | | | | | |
| 5. continuation of 4 to a hexamer | 7.68 | 6.43 | 6.15 | 47.1 | −24.2 | −1.2 | −80.1 | 109.4 | A–B | −55.7 | −98.1 | −1331.5 | −1485.3 | −153.8 |
| | | | | | | | −74.2 | 127.5 | B–C | | | | | |
| | | | | | | | −68.2 | 106.7 | C–D | | | | | |
| | | | | | | | −106.3 | 85.5 | D–E | | | | | |
| | | | | | | | −53.8 | 111.7 | E–F | | | | | |
| 6. minimized tetramer with D residue deeply in cleft | 7.42 | 6.45 | 5.57 | 42.3 | −13.1 | −1.8 | −78.7 | 114.1 | A–B | −40.6 | −72.0 | −1330.7 | −1443.3 | −112.6 |
| | | | | | | | −103.6 | 88.9 | B–C | | | | | |
| | | | | | | | −87.2 | 99.0 | C–D | | | | | |
| 7. lowest energy hexamer from 6 | 7.49 | 6.39 | 5.76 | 49.2 | −21.3 | 2.7 | −55.0 | 121.9 | A–B | −57.4 | −100.7 | −1329.1 | −1487.2 | −158.1 |
| | | | | | | | −108.0 | 85.4 | B–C | | | | | |
| | | | | | | | −90.5 | 93.3 | C–D | | | | | |
| | | | | | | | −100.4 | 105.5 | D–E | | | | | |
| | | | | | | | −46.2 | 111.2 | E–F | | | | | |

[a] All terms used in this table are defined in ref 4, except $E_{CONF}$ ($=E_{SUB} + E_{INT}$) and $E_{TOT}$ ($= E_{SUB} + E_{ENZ} + E_{INT}$). [b] Residue 3 was the reference residue for all structures in this table. [c] $X$, $Y$, $Z$ in Å; $\alpha$, $\beta$, $\gamma$ in deg. [d] $\phi$ and $\psi$ in deg. [e] In kcal/mol.

previously,[5] the allowed conformational space for the $(GlcNAc)_6$–lysozyme complexes was examined further. To search for the low-energy conformations of $(GlcNAc)_6$ bound to the *flexible* enzyme, it was first necessary to minimize the energy of our lowest-energy complex (rigid enzyme) from ref 5 by allowing the enzyme to be flexible. The purpose of this new minimization was to see if the structure might be altered when the conformation of the enzyme was allowed to change and to enable a comparison to be made with the energy-minimized model-built structure.

In the case of our lowest-energy hexamer bound to lysozyme, minimization of the energy of the complex resulted in the structure shown in Figure 1B and listed in Table II, conformer 3. The value of $E_{CONF}$ for the flexible enzyme (−157.2 kcal/mol) is about 4 kcal/mol lower than the value of $E_{COM}$ for the rigid enzyme (−153 kcal/mol), even though the conformation of the substrate is essentially unaltered when the enzyme is allowed to become flexible.

As in the case of the low-energy complex with the rigid enzyme, this "left-sided" structure[5] involved interactions between the Cl atom of residue 4 and the carboxyl side chain of Asp 52 (distance ~3.5 Å), while the carboxyl of Glu 35 was further away (>4 Å) from the bridge oxygen than in the energy-minimized X-ray structure, placing it in a less advantageous catalytic location.

The values of $E_{TOT}$ and $E_{CONF}$ (−1485 and −157 kcal/mol, respectively) are lower for this structure than the corresponding values (−1481 and −152 kcal/mol, respectively) for the energy-minimized model-built structure. However, the value of $E_{INT}$ is lower (by about 4 kcal/mol) for the energy-minimized model-built structure. This extra stabilization energy arises from contacts in the E and F sites.

**C. Other Possible Complexes (Left Side of Cleft).** In calculating the structure of the lowest-energy hexamer bound to the rigid enzyme, the procedure involved the sequential addition of residues, in different conformations, to stable structures obtained by energy minimization.[4,5] It is possible that, in "building" oligomers of increasing chain length, conformations were discarded that may have been stabilized had the side chains of the enzyme been allowed to move. Thus, the next step in searching for *new* low-energy hexamer conformations with a flexible enzyme was to "rebuild" a hexamer from smaller units, while allowing the side chains of the active site to move.

As noted in a previous publication,[5] there are at least two stable conformations for residues in our calculated D site. One of these lies near the surface of the cleft, while the other lies deeper in the cleft though not as deep as the position proposed on the basis of crystallography and model-building studies.[6,7] As a point of departure, therefore, three tetramers were generated, each with its fourth residue in a different conformation in the "D" site. The first tetramer was the lowest-energy minimum for a tetramer in sites A–D, with the fourth residue near the surface of the cleft (see conformer 6 of Table I in ref 4); the second tetramer was a low-energy minimum whose fourth residue lay further inside the cleft and whose residues were generated in a conformation identical with that for the first four residues of our lowest-energy hexamer bound to the rigid enzyme (see conformer 3 of Table I in ref 5); the third tetramer was a structure in which the fourth residue was buried deeply in the cleft and whose conformation was generated using the same values for external and internal variables as for the model-built structure.[6]

To each of the first two (minimum-energy) conformers, a fifth residue was added in one of three conformations, corresponding to left- and right-handed helices and a structure intermediate between these two. The energies of the six resulting complexes were minimized allowing for flexibility of the substrate and for side-chain motion of the enzyme. Three low-energy structures resulted, all of which were similar both in final conformation and in energy. Conformer 4 in Table II illustrates one such structure, in this case obtained from the first structure with residue 4 out of the cleft and residue 5 added in the left-handed conformation. This structure has a similar conformation to the lowest-energy pentamer structure obtained with the rigid enzyme (conformer 2 of Table I, ref 5) and could be readily extended to a hexamer (conformer 5, Table II) in a conformation very close to that for conformer 3 in Table II, one of our lowest-energy hexamers.

Thus, regardless of which low-energy starting conformation for tetramers in sites A–D was used and regardless of how the fifth residue was added, the only low-energy structures that resulted had the same conformation as that calculated with the rigid enzyme. Inspection of Figure 1B provides an explanation for this behavior. Any pentamer or hexamer that binds with its first four residues in sites A–D and with its fourth residue removed from the cleft is "destined" to bind to residues on the left side of the active site as shown in Figure 1B. Binding to positions on the right side of the cleft in the lower active site requires that the D ring bind deeply in the cleft in the region around residues Glu 35 and Asp 52.

In our search procedures, no conformation of a stable tetramer could be discovered in which the D residue was bound deeply in the active-site cleft. Since the possibility existed that movement of the side chains could allow binding in this mode, we performed minimizations of the energies of tetramers in which the fourth residue was allowed to assume such a conformation in the D region. As a starting point, a tetramer conformation, with standard geometry, was selected so that the disposition of the molecule was similar to that of the first four residues of the model-built structure,[6] i.e., the rigid body variables and inter-ring dihedral angles were the same as for the model-built structure.

This starting structure has a number of bad contacts with the enzyme, involving predominantly the second residue of the substrate with Trp 63 and the fourth residue of the substrate with Glu 35, Asn 46, and Asp 52. Energy minimization of this structure with a flexible enzyme resulted in a conformation of low energy, conformer 6 of Table II. Using this structure as a starting point, a fifth residue was added in three different conformations as described above in this section, and the energy of each structure was minimized. To each of the three resulting structures, a sixth residue was added in the same manner, and the energy of the nine resulting hexamers was minimized. Three low-energy hexamers were obtained, all of which had essentially the same basic structure, the lowest-energy conformation of which is conformer 7 in Table II. It is quite interesting that this structure (as well as the other two lowest-energy structures obtained by this procedure) involves quite similar contacts between the substrate and the enzyme as was found in our original lowest-energy hexamer (rigid enzyme), viz., on the left side of the active site in the region of Arg 45, Asn 46, and Thr 47.

Thus, of the 15 hexamer structures tested with the fourth residue in a variety of conformations, the only low-energy structures that resulted were those in which the substrate was bound to the active site on the left side, in the region of Arg 45, Asn 46, and Thr 47.

**D. Other Possible Complexes (Right Side of Cleft).** Since the energy-minimized model-built structure could bind favorably to the opposite (i.e., right) side of the cleft, whereas our lowest-energy structures preferred the left side, the question arose as to whether a hexamer could bind in the "right-side" E and F regions *without* having its fourth residue distorted into the half-chair conformation. To answer this question, three different hexamers were generated with their fifth and sixth residues bound to the right side of the active site. The first such structure, conformer 1, Table III, was a hexamer with standard geometry[3-5,15] generated in a disposition identical with that of the model-built structure.[6] The second conformer, 3 of Table III, was also a hexamer with standard geometry but whose coordinates were fit by least squares as closely as possible (root mean square deviation of ~0.7 Å) to the coordinates of the model-built structure. The third hexamer had a nonstandard geometry. In this case, residues 1–3, 5, and 6 had the geometry of the model-built structure[6] while residue 4 was generated in the chair form with a standard geometry.[3-5,15] The coordinates of this "hybrid" structure were then fit by least squares as closely as possible (root mean square deviation of ~0.5 Å) to those of the model-built structure.[6] The conformation is given in Table III, conformer 5.

Energy minimization of the two structures with standard geometry resulted in high-energy minima, conformers 2 and 4 in Table III. Both of these structures moved relative to the starting structures so that the two terminal residues no longer were present in sites E and F.

Examination of the third structure, conformer 5 of Table III, indicated that the major source of unfavorable contacts with the enzyme was not the fourth residue in the D site but rather the fifth residue in the E site. In particular, highly unfavorable contacts existed between the *N*-acetyl group of the fifth residue and the side-chain atoms of Asn 44 (summing to a total of $>1 \times 10^3$ kcal/mol). It is interesting, though, that the $-CH_2OH$ group of the fourth residue in the D site made only one mildly unfavorable contact (~6 kcal/mol) with the indole ring of Trp 108. This is one of the contacts which has been postulated to cause the fourth residue in the D site to assume a half-chair conformation.[6] The only other unfavorable contact with the enzyme involving the fourth residue in the chair form was between the *N*-acetyl group and the side chain of Asn 46, amounting to ~5 kcal/mol. However, *both* of these unfavorable contacts existed even in the model-built structure[6] (conformer 2 of Table II) in which residue 4 was in the half-chair form. One may expect that both unfavorable contacts, involving only side-chain interactions, may be relieved by simple side-chain rotations rather than by distortion of the D ring.

Energy minimization of conformer 5 in Table III, though it resulted in a high-energy minimum (conformer 6 in Table III), produced a structure which remained bound in the E and F sites and whose binding energy, $E_{INT}$, was low (−104 kcal/mol). This energy is comparable with that for our lowest energy hexamers bound to the opposite side of the active site, conformers 3 and 7 in Table II and Figure 1B. In addition, the unfavorable contacts between the fourth residue in the D site and Trp 108 and Asn 46 of the enzyme were relieved while the chair conformation of residue 4 was maintained. There is no unfavorable steric interaction therefore that forces the fourth residue to assume the half-chair conformation, a conclusion also reached by Levitt[13] in his investigation of the conformational properties of the model-built $(GlcNAc)_6$–lysozyme complex.

Though the energy of interaction, $E_{INT}$, of this energy-minimized structure (conformer 6, Table III) is favorable, the value of $E_{TOT}$ is relatively high when compared with that for conformer 1 of Table II, the energy-minimized model-built structure. It is the internal conformational energy of the substrate, $E_{SUB}$, which is responsible for the destabilization of this conformer. Examination of the interactions responsible for the unfavorable $E_{SUB}$ revealed a number of unfavorable contacts such as the O′ (O13 in ref 3) of residue 6 with the C2 of the same residue. Since the geometry of the molecule is nonstandard even for the undistorted residues and was the cause for most of these bad contacts, a hexamer with standard geometry was fit by least squares to this structure (root mean square deviation of ~0.6 Å) in its favorable binding position that gave a low $E_{INT}$ (conformer 6 in Table III). Minimization of the energy of *this* complex resulted in conformer 7 in Table III whose total conformational energy is now approximately equal to that for the energy-minimized model-built structure, conformer 1 in Table II. This structure is shown in Figure 2A. It may be noted, however, that both structures are higher in energy by about 7 kcal/mol than the best structures bound to the left side of the cleft, conformers 3 and 7, Table II.

It is clear that, with standard geometry, a chair form may be accommodated at the D site with favorable binding of the fifth and sixth residues to the E and F sites on the right side of the cleft. This conclusion was also reached by Levitt,[13] working with the Cartesian coordinates of the model-built structure. However, comparison of the conformational energies of conformers 7 in Table III with that of conformer 1 of Table II (the energy-minimized model-built structure) demonstrates that the latter makes somewhat better contacts with the enzyme while the substrate energy of the former is somewhat more favorable, again in agreement with the findings of Levitt.[13] There was the possibility, however, that the substrate energy of the model-built structure could be improved by using a standard geometry for all residues except residue 4, for which the half-chair conformation would exist.

Therefore, a structure was generated in which residues 1–3, 5, and 6 had standard geometries while residue 4 was held in the model-built (half-chair) conformation. This structure was then fit by least squares (rms deviation of 0.4 Å) to the coordinates of the model-built structure (conformer 8, Table III), and the conformational energy of the resulting structure was minimized at the flexible active site. The result of this minimization is shown in Figure 2B and listed in Table III, conformer 9. This structure is clearly the one of lowest conformational energy obtained thus far, *provided that distortion of the fourth residue is not taken into account.* If it is taken into account, then the energy of this structure would be equal to or higher than the energy of our lowest-energy structures on the left side of the active site, conformers 3 and 7 of Table II.

It was of course possible that a hexamer with standard geometry could bind as favorably to the active site if placed in the same region as that of conformer 9 in Table III. To test this possibility, the coordinates of a hexamer with standard geometry were fit as closely as possible to those of conformer 9 of Table III and Figure 2B. Unlike the latter conformer, the resulting structure with standard geometry made a number of unfavorable contacts with the enzyme, especially between the *N*-acetyl group of residue 5 and the side-chain atoms of Asn 44. Energy minimization of this structure relieved the bad contacts but resulted in a structure of significantly higher energy than

**Table III**
**Hexamers Making Contacts with Enzyme Residues on the Right Side of the Active Site (All Minimizations Performed with Flexible Enzyme)[a]**

| $(GlcNAc)_6$ structure[b] | rigid body variables[c] | | | | | | inter-ring dihedral angles[d] | | binding sites | $E_{SUB}$[e] | $E_{INT}$[e] | $E_{ENZ}$[e] | $E_{TOT}$[e] | $E_{CONF}$[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $X$ | $Y$ | $Z$ | $\alpha$ | $\beta$ | $\gamma$ | $\phi$ | $\psi$ | | | | | | |
| 1. standard geometry, conformation of model-built structure[6] | same as for conformer 2, Table II | | | | | | | | | $1.2 \times 10^5$ | $2.6 \times 10^4$ | $-1.3 \times 10^3$ | $1.5 \times 10^5$ | $1.5 \times 10^5$ |
| 2. energy-minimized conformer 1 | 7.12 | 5.71 | 5.06 | 36.1 | −8.9 | −4.8 | −82.1 | 103.4 | A–B | −53.8 | −89.2 | −1331.5 | −1474.5 | −143.0 |
| | | | | | | | −89.6 | 97.6 | B–C | | | | | |
| | | | | | | | −96.1 | 109.6 | C–D | | | | | |
| | | | | | | | −145.1 | 85.3 | D–E | | | | | |
| | | | | | | | −78.4 | 125.4 | E–F | | | | | |
| 3. standard geometry fit to model-built structure[6] | 7.33 | 5.61 | 5.32 | 30.2 | −18.6 | −0.9 | −52.1 | 95.1 | A–B | $4 \times 10^3$ | $2.1 \times 10^6$ | $-1.3 \times 10^3$ | $2.1 \times 10^6$ | $2.1 \times 10^6$ |
| | | | | | | | −49.1 | 90.2 | B–C | | | | | |
| | | | | | | | −130.5 | 105.0 | C–D | | | | | |
| | | | | | | | −188.5 | 125.3 | D–E | | | | | |
| | | | | | | | −99.0 | 139.1 | E–F | | | | | |
| 4. energy-minimized conformr 3 | 7.13 | 5.36 | 5.18 | 36.6 | −10.6 | −4.4 | −88.3 | 106.6 | A–B | −51.1 | −93.3 | −1331.4 | −1475.8 | −144.4 |
| | | | | | | | −82.8 | 99.1 | B–C | | | | | |
| | | | | | | | −106.6 | 105.0 | C–D | | | | | |
| | | | | | | | −138.7 | 86.3 | D–E | | | | | |
| | | | | | | | −79.6 | 143.7 | E–F | | | | | |
| 5. model-built structure[6] with a chair form for residue 4 | 6.70 | 5.78 | 5.00 | 38.6 | −5.9 | −11.1 | −67.5 | 86.2 | A–B | $2.8 \times 10^3$ | $3.8 \times 10^3$ | $-1.3 \times 10^3$ | $5.1 \times 10^3$ | $6.4 \times 10^3$ |
| | | | | | | | −56.9 | 65.8 | B–C | | | | | |
| | | | | | | | −132.6 | 110.0 | C–D | | | | | |
| | | | | | | | −190.0 | 116.5 | D–E | | | | | |
| | | | | | | | −100.7 | 120.3 | E–F | | | | | |
| 6. energy-minimized conformer 5 | 7.02 | 5.36 | 4.40 | 44.8 | −1.90 | −12.5 | −95.0 | 99.2 | A–B | −34.3 | −103.8 | −1328.5 | −1466.6 | −138.1 |
| | | | | | | | −102.0 | 78.4 | B–C | | | | | |
| | | | | | | | −115.1 | 112.6 | C–D | | | | | |
| | | | | | | | −155.5 | 80.5 | D–E | | | | | |
| | | | | | | | −64.3 | 128.0 | E–F | | | | | |
| 7. standard geometry fit to conformer 6 and energy minimized | 7.08 | 5.30 | 4.70 | 40.1 | −12.2 | −6.3 | −84.7 | 105.6 | A–B | −52.3 | −99.2 | −1329.4 | −1480.9 | −151.5 |
| | | | | | | | −96.8 | 97.6 | B–C | | | | | |
| | | | | | | | −104.0 | 111.4 | C–D | | | | | |
| | | | | | | | −137.1 | 76.2 | D–E | | | | | |
| | | | | | | | −73.5 | 142.3 | E–F | | | | | |
| 8. standard geometry with half-chair form for residue 4 fit to model-built structure[6] | 7.41 | 5.73 | 5.25 | 34.3 | −23.1 | 4.0 | −53.0 | 103.4 | A–B | 56.7 | $5.3 \times 10^4$ | $-1.3 \times 10^3$ | $5.2 \times 10^4$ | $5.3 \times 10^4$ |
| | | | | | | | −60.3 | 101.4 | B–C | | | | | |
| | | | | | | | −117.4 | 124.6 | C–D | | | | | |
| | | | | | | | −156.8 | 97.9 | D–E | | | | | |
| | | | | | | | −88.3 | 122.7 | E–F | | | | | |
| 9. energy-minimized conformer 8 | 7.36 | 5.04 | 4.88 | 35.0 | −25.0 | −6.3 | −77.0 | 130.7 | A–B | −56.6 | −108.8 | −1330.2 | −1495.6 | −165.4 |
| | | | | | | | −84.0 | 95.7 | B–C | | | | | |
| | | | | | | | −114.7 | 130.1 | C–D | | | | | |
| | | | | | | | −142.5 | 78.5 | D–E | | | | | |
| | | | | | | | −65.7 | 109.4 | E–F | | | | | |

[a] All terms used in this table are defined in ref 4 except $E_{CONF}$ ($=E_{SUB} + E_{INT}$) and $E_{TOT}$ ($=E_{SUB} + E_{INT} + E_{ENZ}$). [b] Residue 3 was the reference residue for all structures in this table. [c] $X$, $Y$, $Z$ in Å; $\alpha$, $\beta$, $\gamma$ in deg. [d] $\phi$ and $\psi$ in deg. [e] In kcal/mol.

that of conformer 9 of Table III by about 23 kcal/mol. Thus, even if a hexamer with standard geometry is placed in a conformation very similar to that of conformer 9, Table III, it still cannot bind with the same affinity. If it binds with its two terminal residues in the E and F sites, it must bind in a conformation similar if not identical with that of conformer 7 in Table III (shown in Figure 2A).

**E. Comparison of Complexes.** Comparison of the energies of conformer 7 with those of conformer 9 in Table III suggests that optimization of contacts in sites E and F on the right side of the cleft seems to require distortion of the fourth residue. It is not clear how much distortion is necessary, however. Thus, it may be possible that relatively minor changes in some bond lengths and/or bond angles from the standard values could allow for an equally low-energy conformation without distortion to a half-chair. We are currently investigating this possibility.

It is interesting to speculate about the significance of the three low-energy species calculated at the active site of the flexible enzyme: conformer 7 (or 3) of Table II, the lowest-energy structure obtained with a standard geometry; conformer 7 of Table III, the structure with a standard geometry binding to the E and F sites on the right side of the cleft; and conformer 9 of Table III, a structure with standard geometry for all residues except residue 4 which is maintained in the model-built half-chair form.[6] Recent stopped-flow and relaxation studies on the binding of $(GlcNAc)_6$ to hen egg-white lysozyme[16] indicate that three "productive" species exist in the $\beta$ process (see ref 16 for terminology). The first species formed is the most stable. This form is in equilibrium with a less stable but still energetically favorable form which is in turn in equilibrium with a catalytically active species.[16]

The most favored calculated undistorted binding mode is conformer(s) 3 (or 7) in Table II, on the left side of the active site. This species may be in equilibrium (which favors conformer 3) with conformer 7 of Table III on the right side of the cleft. This conformer, in turn, can undergo a distortion of its fourth residue to produce the catalytic species, conformer 9 of Table III whose distortion is counterbalanced by an increase in favorable contacts in the E and F sites on the right side of the cleft.

In view of these calculations and experimental results that suggest that there may be two different binding regions for hexasaccharides, it becomes important to determine experimentally the binding modes of $(GlcNAc)_6$ to lysozyme especially in the lower active site. The overall features of the two major binding regions are displayed in color stereo in Figure 3 showing the active site of the native enzyme (top) for reference, the energy-minimized model-built structure, "right-sided" complex with half-chair (middle), and the best calculated undistorted binding mode for $(GlcNAc)_6$ on the left side of the active site (bottom). For both major binding conformations (Figure 3, middle and bottom), contacts between enzyme and substrate appear to be quite optimal.

**F. Stabilizing Interactions for the Two Major Binding Modes. Origin of Specificity.** The contribution of each site to the overall energy of interaction, $E_{INT}$, for both major complexes listed in Table II (conformers 3 and 7, "left-sided" complex) and in Table III (conformer 8, "right-sided" complex) is given in Table IV. It is quite interesting that, for the "left-sided" complex, the best interactions occur in subsite C, whereas the most favorable interactions for the complex on the right side occur in subsite E. It is predominantly the E site which stabilizes the "right-sided" complex. The interaction

Table IV
Site Contributions[a] to Lowest Energy Complexes of $(GlcNAc)_6$ with Lysozyme

| site | $E_{INT}$, kcal/mol "left-sided" complex[b] | $E_{INT}$, kcal/mol "right-sided" complex[c] |
|---|---|---|
| A | −5.2 | −10.5 |
| B | −17.3 | −9.0 |
| C | −24.5 | −20.1 |
| D | −21.6 | −28.0[d] |
| E | −19.1 | −29.7 |
| F | −17.5 | −11.6 |

[a] This contribution is the interaction energy between enzyme and substrate and does not include the internal conformational energy of substrate or enzyme. The contribution to the energy of any site is obtained by summing the interactions of the GlcNAc residue in that site with the *whole* protein; in practice, the major contribution comes from residues within ~4 Å of each of the atoms of the GlcNAc residue. [b] Conformer 3 (or 7) in Table II and Figure 1B. [c] Conformer 8 in Table III and Figure 2B. [d] This energy does not include the distortion energy for the binding of a half-chair form for a GlcNAc residue in the D site.

energy for this site, in fact, is the lowest of all and leads to the prediction that a GlcNAc monomer will bind with high affinity to this site. Though the X-ray crystal structure[6] shows that both $\alpha$ and $\beta$ anomers of GlcNAc bind in or near site C, it is known that both sites E and F are blocked in the crystal, thus making comparison impossible.[6] Recent X-ray studies on turkey egg-white lysozyme indicate, however, that in addition to a well-defined binding region in the C site there is a strong binding region for a GlcNAc residue in the E site on the right side of the cleft.[17]

Comparison of the values of $E_{INT}$ for the two complexes in Table IV reveals almost equal energies for sites A–D (−68.6 kcal/mol for the undistorted mode and −67.6 kcal/mol for the distorted mode), ignoring, of course, the strain energy that occurs in the distorted binding mode. When the contributions for sites E and F are compared, the right site is favored (−41.3 kcal/mol vs. −36.6 kcal/mol). Apparently, to reach this site, though, the fourth residue from the nonreducing end in the D site must undergo distortion.

Further inspection of Table IV shows that, for the "right-sided" complex, binding to sites E and F is as favorable as binding to sites A–C. This result leads to the conclusion that $(GlcNAc)_2$ and $(GlcNAc)_3$ should both have the same affinities for lysozyme, a conclusion that contradicts experimental data showing a significantly higher affinity for the trimer.[6] Binding of a trimer to sites A–C for the "left-sided" complex in Table IV, however, is favored over the E and F sites of both "left-sided" and "right-sided" complexes. The C site in the binding mode for the "left-sided" complex has the greatest affinity for a GlcNAc residue.

It should be borne in mind that all of the above conclusions concerning relative affinities of oligomers of GlcNAc for various binding sites are tentative because the possibility of differential solvent effects has not been considered here.

It is of interest to examine the interactions in the various sites that lead to strong affinities between enzyme and substrate. Analysis of contributions to binding energies in each site reveals that, for the "left-sided" complex, approximately one-third of the total $E_{INT}$ arises from interactions of the six *N*-acetyl groups with the enzyme, in good agreement with experiment.[6] In sites B and C, the

*N*-acetyl groups contribute the most, accounting for one-half of the interaction energy. A significant fraction (~20%) of this energy arises from hydrogen bonding as occurs, for example, between the –NH of the *N*-acetyl group and the C=O of Ala 107, as discussed earlier in the Discussion section. The bulk of the interaction energy, though, between enzyme and substrate arises from the attractive nonbonded interactions. Thus, the *N*-acetyl group confers binding specificity, which is lacking in glucose.[6]

In the case of the "left-sided" complex in Table IV, a number of hydrogen bonds occur which confer great stability on this structure. Thus, in addition to the *N*-acetyl interactions in site C, highly favorable hydrogen bonds exist between the O6H (H19 in ref 3) of residue 4 and the terminal C=O of Gln 57, between the O3H (H17 in ref 3) of residue 5 and the carboxyl group of Asp 52, and between the corresponding OH of residue 6 and the backbone C=O of Arg 45. In the latter case, it would be of great interest to determine if the removal of the 3-OH group from a dimer acceptor (i.e., GlcNAc-$\beta$-1,4-3-deoxyGlcNAc) affects its ability to serve as an acceptor.

Consideration of the "right-sided" complex in Table IV reveals a somewhat different set of stabilizing interactions, viz., the *N*-acetyl groups in this case contributing significantly less than one-third of the total energy. The important interactions have already been discussed earlier in the Discussion section. A major contribution to the stability of the binding of a GlcNAc residue to the E site on the right side of the cleft is a strong hydrogen bond between the O3H (H17 in ref 3) and the side chain C=O of Gln 57.

## Summary and Conclusions

Energy searches for low-energy binding regions for $(GlcNAc)_6$ at the active site of the *rigid* enzyme reveal a unique binding mode in which residues 5 and 6 associate with enzyme residues on the "left" side of the active site involving such residues as Arg 45, Asn 46, and Thr 47. Stable binding regions in which residues 5 and 6 are bound to residues such as Phe 34 and Arg 114 (on the "right" side) cannot be achieved unless the side chains of the enzyme are allowed to move. When the enzyme side chains are allowed to move, a hexamer with standard geometry can bind with good affinity to the "right" side of the cleft though not as favorably as on the "left" side. A dramatic increase in the affinity of the hexamer with standard geometry can be achieved by introducing the model-built (strained D ring) structure for the fourth GlcNAc residue. Thus, though the structure with standard geometry can make favorable contacts in the "right" region of the active site, distortion of the fourth residue can result in optimizing the contacts. How much distortion is needed to make such an accommodation is an unsettled question. Certainly, distortion of residue D to the model-built structure will raise the energy, and this additional (strain) energy may or may not be compensated for by new and favorable interactions with the enzyme.

**Acknowledgment.** We wish to thank Drs. G. Némethy and S. Rackovsky for stimulating discussions, Mrs. S. Rumsey for expert computing assistance, and Dr. Richard J. Feldman of the Division of Computer Research and Technology, N.I.H., for the preparation of the color stereo representations presented in this paper.

## References and Notes

(1) This work was supported by research grants from the National Science Foundation (PCM 75-08691) and from the National Institute of General Medical Sciences, National Institutes of Health, U.S. Public Health Service (GM-14312).
(2) NIH Postdoctoral Fellow, 1975–76. Todd Fellow, 1978–79.
(3) M. R. Pincus, A. W. Burgess, and H. A. Scheraga, *Biopolymers*, **15**, 2485 (1976).
(4) M. R. Pincus, S. S. Zimmerman, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 4261 (1976).
(5) M. R. Pincus, S. S. Zimmerman, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 2629 (1977).
(6) T. Imoto, L. N. Johnson, A. C. T. North, D. C. Phillips, and J. A. Rupley, "The Enzymes", 3rd ed., Vol. VII, P. D. Boyer, Ed., Academic Press, New York, 1972, p 665.
(7) L. O. Ford, L. N. Johnson, P. A. Machin, D. C. Phillips, and R. Tjian, *J. Mol. Biol.*, **88**, 349 (1974).
(8) M. Schindler, Y. Assaf, N. Sharon, and D. M. Chipman, *Biochemistry*, **16**, 423 (1977).
(9) L. G. Dunfield, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **82**, 2609 (1978).
(10) M. J. D. Powell, *Comput. J.*, **7**, 155 (1964).
(11) M. R. Pincus and H. A. Scheraga, *J. Phys. Chem.*, **81**, 1579 (1977).
(12) C. R. Beddell, C. C. F. Blake, and S. J. Oatley, *J. Mol. Biol.*, **97**, 643 (1975).
(13) M. Levitt in "Peptides, Polypeptides and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Eds., Wiley, New York, 1972, p 99.
(14) M. Schindler, Ph.D. Thesis, The Weizmann Institute of Science, Rehovot, Israel, 1977.
(15) S. Arnott and W. E. Scott, *J. Chem. Soc., Perkin Trans. 2*, **3**, 324 (1972).
(16) E. Holler, J. A. Rupley, and G. P. Hess, *Biochemistry*, **14**, 2377 (1975).
(17) R. Sarma and R. Bott, *J. Mol. Biol.*, **113**, 555 (1977).