Substrate-Assisted Catalysis in Sialic Acid Aldolase

Brian J. Smith,* Michael C. Lawrence, and João A. R. G. Barbosa†

Biomolecular Research Institute, Royal Parade, Parkville, Victoria 3052, Australia

Received September 28, 1998

Sialic acid aldolase catalyses the reversible aldol condensation of pyruvate and *N*-acetylmannosamine with an apparent lack of stereospecificity. Consistent with this, modeling of Schiff base and enamine intermediates in the active site of this enzyme yields two conformations, corresponding to *si*- and *re*-face attack in the aldol condensation reaction. The acceptor−aldehyde group is found on different sides of the enamine in the two conformations, but with the remainder of the substrate having very similar geometries in the protein. No histidine residue previously speculated to function as a general base in the mechanism is found near the enzyme active site. In the absence of functionally active groups in the active site, the carboxylate of the substrate is proposed to function as the general acid/base. Molecular orbital calculations indicate that the barrier to aldol cleavage via this mechanism in the gas phase of the related system, 4-hydroxy-2-methyiminopentanoic acid, is 74 kJ mol^{-1}.

1. Introduction

Sialic acid aldolase (EC 4.1.3.3, *N*-acetylneuraminate lyase, NAL) catalyzes the cleavage of sialic acid (*N*-acetylneuraminic acid) to produce pyruvate and *N*-acetyl-D-mannosamine (Scheme 1). The enzyme is a member of the class I aldolases that form a Schiff base between a lysil group of the enzyme and the acetal carbonyl of the substrate. Following Schiff base formation, proton abstraction from the C_4 hydroxyl brings about cleavage of the C_3-C_4 bond and formation of an enamine (atoms of the sialic acid substrate are labeled with a subscript (C_x) and atoms of the acceptor as C-*x*). Beyond this level of understanding, details of the mechanism have not been forthcoming. The X-ray crystallographic structure[1] has enabled identification of the lysine responsible for Schiff base formation; however, the absence of functionally active groups in the active site (other than the Schiff base forming lysine) has obfuscated any attempts to reveal further details of the mechanism.

A histidine in the active site had been presumed[2] to function as the base in the proton abstraction, with precedences in fructose-1,6-diphosphate-aldolase (FDP)[3] and transaldolase.[4] Unfortunately, the structure of both transaldolase B[5] and sialic acid aldolase[1,6] have shown there is no histidine in the active site of either enzyme. Also, the histidine in the active site of FDP is involved only in maintaining the structural integrity of the active site.[7] Despite this, histidine continues to be implicated in the mechanism of NAL.[8,9] The role of histidine as the general base in aldolase activity is, however, fallacious.

NAL is specific to the α-anomer of sialic acid[10] and must, therefore, facilitate ring opening to reveal the acetal carbonyl. NAL also catalyzes the reverse aldol condensation reaction, usually induced by an excess of pyruvate (Scheme 2).[11] The enzyme therefore provides a means to synthesize many sialic acid derivatives.[11] In this capacity, the enzyme is specific to pyruvate as a donor and yet tolerates a wide variety of hexoses, pentoses, and tetroses of both D and L configuration as the acceptor substrate. Unlike most aldolases, NAL appears to lack stereochemical specificity; the outcome of the condensation reaction depends on the acceptor substrate. The reaction appears to be under thermodynamic control, although *re*-face attack (that leads to the *R* configuration) is kinetically disfavored. Thus, while *re*-face attack occurs (predominantly or exclusively) where the product is the thermodynamically more stable species, the reaction occurs more slowly.[11] The orientation of the hydroxyl at C-3 of the parent butanal acceptor is critical for the enzyme's stereoselectivity. Acceptors with an *S* configuration at C-3 (hydroxyl in the C position in Scheme 2) confer condensation through *si*-face attack, whereas an *R* configuration at C-3 (hydroxyl in the D position in Scheme 2) confers either partial or exclusive *re*-face attack.[11] This would tend to indicate that the reaction can proceed from a substrate bound to the enzyme in more than one conformation.

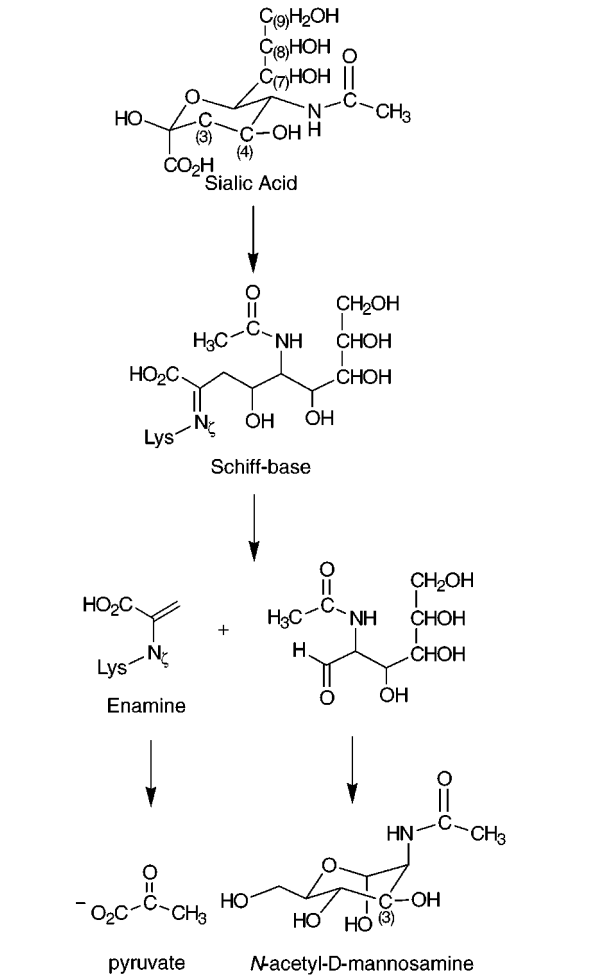
Several explanations for the stereoselectivity of NAL have been proposed.[9,12,13] In particular, Wong et al.[9] have suggested that the acceptor binds the enzyme in a

† Present address: Crystallography Group, IFSC, USP, Caixa Postal 369, Sao Carlos, SP, Brazil 13560-250.

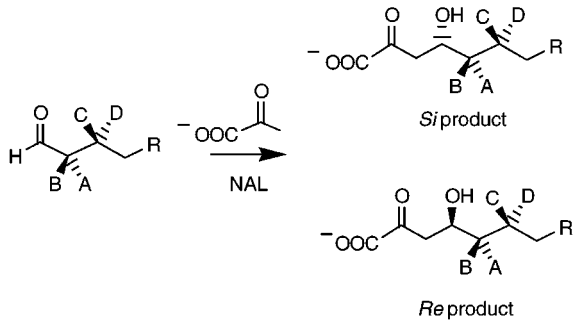
(1) Lawrence, M. C.; Barbosa, J. A. R. G.; Smith, B. J.; Hall, N. E.; Pilling, P. A.; Ooi, H. C.; Marcuccio, S. M. *J. Mol. Biol.* **1997**, *266*, 381.
(2) Schauer, R.; Wember, M. *Hoppe-Seyler's Z. Physiol. Chem.* **1971**, *352*, 1517.
(3) Brand, K.; Tsolas, O.; Horeker, B. L. *Arch. Biochem. Biophys.* **1969**, *130*, 521.
(4) Hoffee, P.; Lai, C. Y.; Pugh, E. L.; Horecker, B. L. *Proc. Nat. Acad. Sci. U.S.A.* **1967**, *57*, 107.
(5) Jia, J.; Schörken, U.; Lindqvist, Y.; Sprenger, G. A.; Schneider, G. *Prot. Sci.* **1997**, *6*, 119.
(6) Izard, T.; Lawrence, M. C.; Malby, R. L.; Lilley G. G.; Colman, P. M. *Structure* **1994**, *2*, 361.
(7) Gamblin, S. J.; Davies, G. J.; Grimes, J. M.; Jackson, R. M.; Littlefield, J. A.; Watson, H. C. *J. Mol. Biol.* **1991**, *219*, 573.
(8) Lin, C.-C.; Lin, C.-H.; Wong, C.-H. *Tetrahedron Lett.* **1997**, *38*, 2649.
(9) Fitz, W.; Schwark, J.-R.; Wong, C.-H. *J. Org. Chem.* **1995**, *60*, 3663.
(10) Deijl, C. m.; Vliegenthart, J. F. G. *Biochem. Biophys. Res. Commun.* **1983**, *111*, 668.
(11) Gijsen, H. J. M.; Qiao, L.; Fitz, W.; Wong, C.-H. *Chem. Rev.* **1996**, *96*, 443.
(12) Schauer, R.; Stoll, S.; Zbiral, E.; Schreiner, E.; Brandsetter, H. H.; Vasella, A.; Baumberger, F. *Glycoconjugate J.* **1987**, *4*, 361.
(13) Kragl, U.; Gödde, A.; Wandrey, C.; Lubin, N.; Augé, C. *J. Chem. Soc., Perkin Trans. 1* **1994**, 119.

10.1021/jo981960v CCC: $18.00
Published on Web 01/20/1999

Scheme 1. Schematic of the Reaction Catalyzed by Sialic Acid Aldolase (NAL) through Schiff Base and Enamine Intermediates



Scheme 2



boatlike conformation when *re*-face attack occurs and in a chairlike conformation for *si*-face attack. This enables attack of the acceptor aldehyde carbonyl from the same face of the enamine but with different faces of the aldehyde group.

We report here results of modeling studies of the Schiff base and enamine intermediates in the active site of sialic acid aldolase. These are based on recently determined X-ray crystallographic structures of the enzyme cocrystallized with two inhibitors, 3-hydroxypyruvic acid,[6] and the alditol resulting from the sodium borotetrahydride reduction of *N*-acetylneuraminic acid.[14] From these, we propose a mechanism of aldol cleavage/condensation in NAL that accounts for the observed stereochemical outcome and the apparent paucity of catalytic machinery in the active site.

2. Methods

Schiff-base and enamine intermediates of *Haemophilus influenzae* NAL were constructed on the basis of the X-ray crystallographic structure including inhibitors.[6,14] These structures were then minimized using the Discover program with the CVFF force field.[15] Residues lying beyond a 12.5 Å radius of the lysil nitrogen were removed to reduce the time of calculation. All non-hydrogen atoms of the protein, with the exception of the lysine side-chain atoms, were constrained to their positions found in the inhibitor-free structure[14] (only minor differences were observed between the native enzyme and those with covalently bound inhibitors).[14] Standard ab initio molecular orbital calculations[16] were performed at the G2(MP2,SVP) level[17] using the Gaussian94 program.[18] Tables of optimized geometries and calculated energies are provided as Supporting Information.

3. Results

3.1. Protein Simulations. The Schiff-base and enamine intermediates could be accommodated in the active site in two conformations that differed in the configuration at C_4 of the Schiff base or in the side of the enamine upon which the aldehyde was located. These two conformations correspond to *si*- and *re*-face attack of the carbonyl. The structures of the substrate portion of the intermediates of these two conformations are presented in Figure 1 for the Schiff base (top) and enamine (bottom). There is a remarkable level of coincidence of many of the atoms, especially those beyond C_4 of the Schiff base.

The interactions of the two Schiff base intermediates with the protein are illustrated in Figure 2. The Schiff base is formed with the lysil nitrogen of Lys165. In both conformations, the carboxylate of the substrate forms hydrogen bonds with the hydroxyl groups of Thr48 and Tyr137 and the backbone amide groups of Ser47 and Thr48, as was found in the crystal structure of the enzyme complexed with the 3-hydroxypyruvic acid.[6] The remainder of both configurations lie in an extended conformation similar to that found for the alditol complexed to the enzyme.[14] At the distal end of the molecule, Glu192 forms hydrogen bonds with the hydroxyl groups at the C_8 and C_9 positions. These interactions had been anticipated from the inhibitor-complexed structures.[6] The C_4 hydroxyl forms an intramolecular hydrogen bond to the carboxylate of the substrate in both conformations. The hydroxyl at C_7 is hydrogen bonded to the hydroxyl

(14) Barbosa, J. A. R. G.; Lawrence, M. C. Manuscript in preparation.

(15) Discover 2.9.7, Biosym/MSI, October 1995. Minimizations were performed with conjugate gradients until the root mean square derivative was less than 0.001 kcal mol^{-1} $Å^{-1}$.

(16) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(17) (a) Smith, B. J.; Radom, L. *J. Phys. Chem.* **1995**, *99*, 6468. (b) Curtiss, L. A.; Redfern, P. C.; Smith, B. J.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 5148.

(18) Frisch, M. J.; Trucks, G.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94, Revision C.3, Gaussian, Inc., Pittsburgh, PA, 1995.

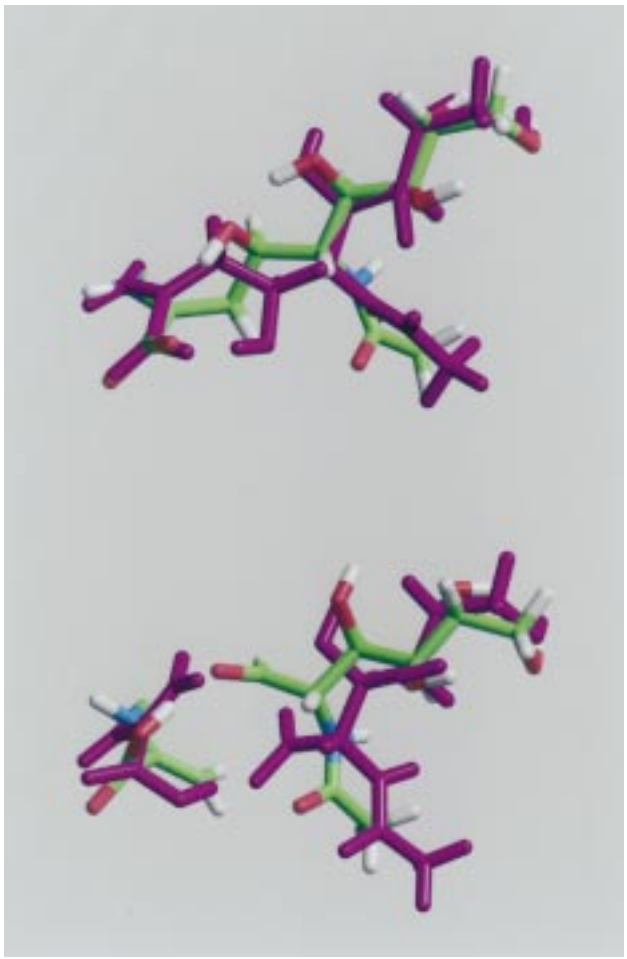


Figure 1. Structures of two configurations of Schiff base (top) and enamine (bottom) intermediates in the active site of sialic acid lyase (NAL): *si* configuration in mauve, *re*-configuration colored according to atom type (hydrogen, white; carbon, green; nitrogen, blue; oxygen, red).

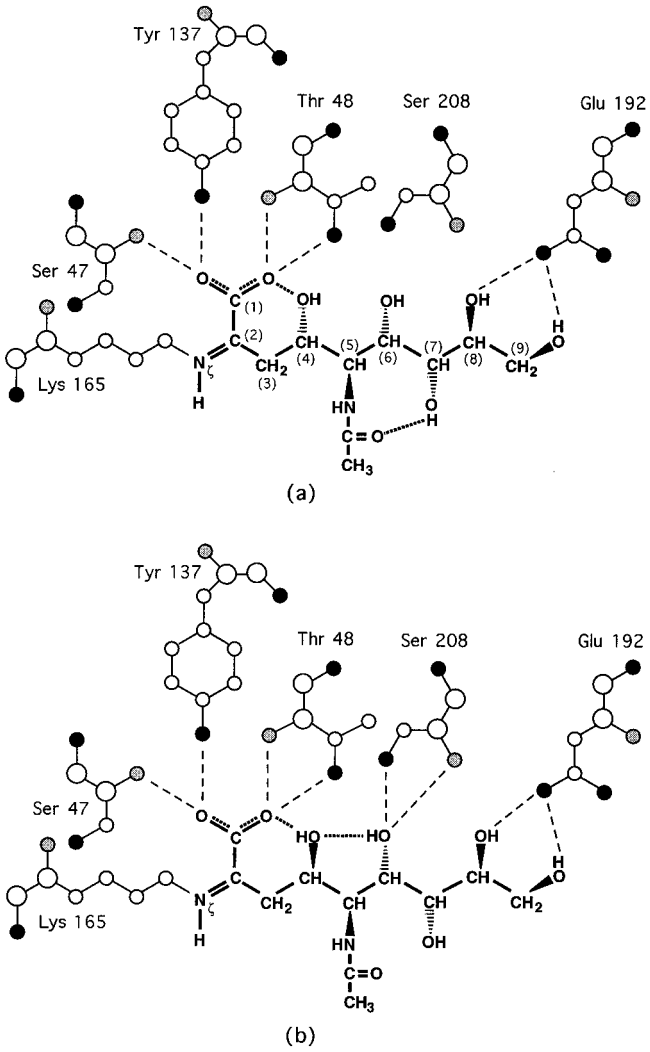


Figure 2. Illustration of interactions between the protein and Schiff base intermediate: (a) *si* configuration (numbering of the sialic acid substrate carbon atoms in parentheses), (b) *re* configuration (open circles, carbon; larger open circles, protein backbone carbon; shaded circles, nitrogen; full circles, oxygen).

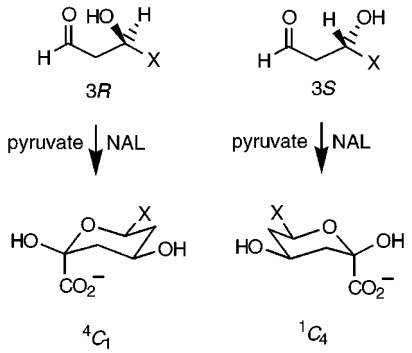
of Ser208 (not shown) and forms an intramolecular hydrogen bond with the carbonyl oxygen of the *N*-acetyl group in the conformation leading to *si*-face attack. In the configuration leading to *re*-face attack, the C_7 hydroxyl makes no specific interactions, while the hydroxyl at C_6 forms a hydrogen bond with the hydroxyl of Ser208 and with the hydroxyl at C_4. In both conformations, the *N*-acetyl group is directed outward from the active site and makes no direct contact with any of the protein atoms, which results in a lack of specificity in its alignment. This is consistent with the enzyme's ability to accommodate a wide variety of N-substituted mannosamines.[8] The general features described above are consistent with the points of recognition of the substrate previously anticipated on the basis of kinetic data.[19]

Apart from the carboxylate of the substrate there are no acidic or basic groups in the active site that might be involved in protonation or proton abstraction. It is also clear that both the *N*-acetyl and glycerol groups of the substrate are far removed from the lysine and cannot be directly involved in the mechanism. A tyrosine residue found in the active site of NAL (Tyr137), and conserved in the amino acid sequence of all other enzymes in this subfamily,[6] does not appear to be able to participate in the aldol reaction as had been postulated earlier.[20] In the configuration leading to *si*-face attack, the C_4 hydoxyl lies on the side of the substrate opposite to the tyrosine. The function of the tyrosine is likely to be related to binding the carboxylate of the substrate and stabilizing the side chain of the Schiff base forming lysine and the bound substrate.[6] We propose, therefore, that proton abstraction from the C_4 hydroxyl, resulting in cleavage of the C_3-C_4 bond, occurs through the carboxylate of the substrate. This mechanism of condensation differs significantly from that proposed by Wong et al.,[9] in that the substrate adopts almost the same conformation for both *si*- and *re*-face attack. The interaction between enamine and aldehyde occurs through different faces of *both* the enamine and aldehyde, with the aldehyde in the same orientation in both conformations. In transaldolase, a similar mechanism involving a protein aspartate (Asp17) has been suggested as the agent responsible for proton abstraction leading to C−C bond cleavage.[5]

(19) Zbiral, E. In *Carbohydrates—Synthetic Methods and Applications in Medicinal Chemistry*; Ogura, H., Hasegawa, A., Suami, T., Eds.; Kodansha VCH: Tokyo, 1992; pp 304−339.

(20) Fesner, W.-D.; Walter, C. *Top. Curr. Chem.* **1996**, *184*, 97−194.

Scheme 3



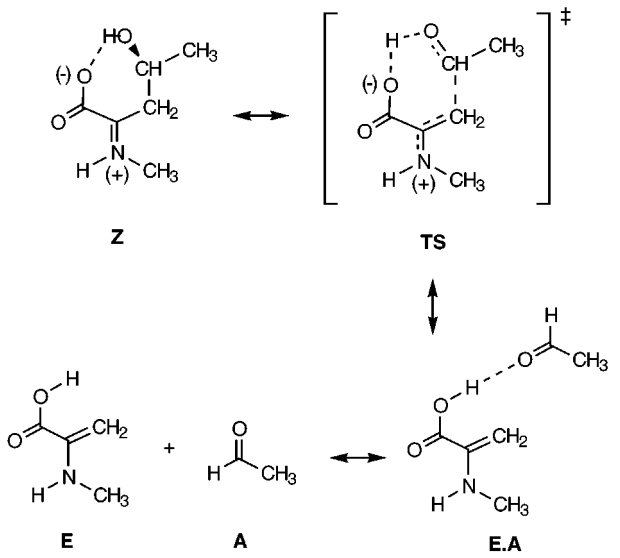
Since the reaction appears to be under thermodynamic control, we propose the enzyme forms intermediates corresponding to both *si*- and *re*-face attack. Thermodynamically favored products are produced preferentially, while kinetically favored products (lower reaction energies) are produced more rapidly. Attack occurs at the *si*-face to yield the product in the $^1C_4$ chair conformation, and at the *re*-face to yield the $^4C_1$ chair, when the configuration at C-3 of the acceptor is *R* and *S*, respectively. This presents the acceptor substituents at C-1 and C-3 in the equatorial position in the condensation product (at C_4 and C_6), the thermodynamically preferred conformation (Scheme 3).

si-Face attack of the acceptor with the *S* configuration at C-3, or *re*-face attack of the acceptor with *R* configuration at C-3, yields chair conformations (either $^1C_4$ or $^4C_1$) in which one of the substituents at C_4 and C_6 must occupy an axial position. In these cases the most stable conformation will be determined by which substituent can be accommodated in the axial position. All of the acceptors studied to date with an *R* configuration at C-3 undergo *si*-face attack[11] leading to a product with ring substituents at the C_4 and C_6 positions in an equatorial position, as expected. The only exceptions to this are those acceptors with an *S* configuration at C-2 ($A = D =$ H in Table 2 of ref 11), which do not undergo condensation.[21]

3.2. Molecular Orbital Calculations. We have obtained an estimate of the gas-phase barrier through consideration of the aldol cleavage of the zwitterion of 4-hydroxy-2-(methyimino)pentanoic acid (**Z**) to form acetaldehyde (**A**) and the enamine 2-methylaminoacrylic acid (**E**) (Scheme 4). The transition state (**TS**) has a C_3-C_4 separation of 1.939 Å, and the proton substantially transferred to the carboxyl oxygen (O−H separations of 1.103 and 1.356 Å to the carboxyl and hydroxyl oxygens, respectively).

The energy difference between **Z** and the complex **E.A** (ΔH_{298}) is just 5.8 kJ mol^{-1}. The barrier to intramolecular aldol cleavage ($\Delta H^{\ddagger}_{298}$) is calculated to be 74.1 kJ mol^{-1}, while for the aldol condensation from the complex **E.A** the barrier is 68.4 kJ mol^{-1}. Dissociation of this complex requires 28.7 kJ mol^{-1}; the barrier, therefore, lies 39.7 kJ mol^{-1} above the energy of the dissociation products. At the AM1 level the barrier to aldol cleavage is 138.2 kJ mol^{-1}, while the condensation barrier is 172.6 kJ mol^{-1}. The PM3 method performs equally poorly with barriers of 121.7 and 210.0 kJ mol^{-1} for the cleavage and condensation barriers, respectively. Thus, unless this system is a pathalogical case for the G2(MP2,SVP) method, the semiempirical methods cannot be used to describe this system satisfactorily.

Scheme 4



The intermediates obtained from the protein simulations (section 3.1) all have geometries in the protein that differ significantly from those in the gas phase obtained at the MP2/6-31G(d) level (Figure 1), in particular, rotation of the carboxylate group (θ) and the dihedral angle involving the *N*-methyl group (τ). We have examined the effect on the energies of **Z**, **TS**, and **E.A** by performing minimizations of **Z** and **E.A** in which these angles are constrained to those found in the protein. For the transition state, values of θ and τ that were the average of those in **Z** and **E.A** were used (a constrained molecular mechanics minimzation of the enamine with the C_3-C_4 separation fixed at 1.94 Å yielded values of θ and τ in close agreement with the average values). MP2/6-31G(d) electronic energies for all the stationary points are roughly 20 kJ mol^{-1} higher than for the unconstrained geometries. Thus, distortion of the substrate does not appear to affect the potential energy profile in either direction (condensation or cleavage) or either conformation (*re*- or *si*-face attack).

One other possibility for the mechanism involves proton transfer between the hydroxyl oxygen and the imonium nitrogen. At the G2(MP2,SVP) level, the transition state for this process in 4-hydroxy-2-(methyimino)-pentanoic acid lies 170.8 kJ mol^{-1} higher than **Z**. Although the influence of the environment of the enzyme can have a substantial impact on the reaction energies, we favor as the more likely mode of reaction in NAL, proton transfer involving the carboxylate since the barrier to cleavage is considerably smaller (∼100 kJ mol^{-1}) in this model system.

4. Discussion

NAL has recently found utility in the catalysis of aldol condensation reactions. Among the growing number of aldolases being used in this way,[22] NAL presents several

(21) Uchida, Y.; Tsukada, Y.; Sugimori, T. *J. Biochem.* **1984**, *96*, 507.

(22) Wong, C.-H.; Halcomb, R. L.; Ichikawa, Y.; Kajimoto, T. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 412.

interesting features that include its apparent lack of stereoselective control, its ability to employ a wide variety of aldehyde acceptors, and its specificity for pyruvate as the donor.[23] The former of these aspects highlight an underlying indiscriminate nature of the active site. The capacity to bind the substrate in more than one configuration leads to the formation of prodcts in both the *R* and *S* configurations. Beyond the aldehyde substituent, the acceptor makes contact with the protein through just one residue side chain (Glu192), accommodating many acceptors, including those with large and bulky substituents.

NAL is specific for pyruvate as the donor.[23] The X-ray structure of the inhibitors bound to the enzyme showed[1] that the carboxylate is involved in several interactions with the protein. This suggests that replacement of the carboxylate with an alternative anionic group may be restricted by size. The mechanism we propose here develops further this restriction by requiring a carboxylate group to bring about the aldol reaction.

Recently, it was shown that sialic acid aldolase is just one member of a subfamily of enzymes, including dihydrodipicolinate synthetase, *trans-o*-hydroxybenzylidenepyruvate hydratase-lyase, and D-4-deoxy-5-oxoglucarate dehydratase. While acting on very different substrates, they are all believed to have similar three-dimensional structures (particularly across the $(\beta/\alpha)_8$ barrel region) and also function in a similar manner.[1] We therefore expect that in these enzymes, too, substrate-assisted catalysis will operate. In the sequence alignment of this family[1] there are no histidines within four residues of any NAL active-site equivalent residue. Moreover, no residue in NAL that aligns with a histidine in any of the other enzymes lies within 10 Å of the active-site lysine.

Substrate-assisted catalysis has been postulated for a wide variety of enzymes.[24] Several enzymes have been engineered such that catalytic group has been removed causing inactivation, which is then partially restored by substrates containing the missing catalytic functional group.[25] It has been speculated that the phosphate of FDP may function in a manner similar to that of the carboxylate of sialic acid in the mechanism of FDP aldolase.[26]

5. Conclusions

Modeling of the Schiff base and enamine intermediates in sialic acid aldolase yields two conformations that correspond to *si*- and *re*-face attack in the aldol condensation reaction and can account for the apparent lack of stereospecificity demonstrated by this enzyme. In both Schiff base configurations, the proton of the hydroxyl group at C_4 is hydrogen bonded to the carboxylate of the substrate. The lack of any other functionalities in the active site capable of abstracting a proton leads us to propose that aldol cleavage is brought about by the substrate carboxylate.

Supporting Information Available: Tables listing calculated G2(MP2,SVP) energies and Cartesian coordinates of optimized geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

JO981960V

(23) Kim, M. J.; Hennon, W. J.; Sweers, H. M.; Wong, C.-H. *J. Am. Chem. Soc.* **1988**, *110*, 6481.

(24) (a) Jacob-Dubuisson, F.; Lamotte-Brasseur, J.; Dideberg, O.; Joris, B.; Frère, J.-M. *Prot. Eng.* **1991**, *4*, 811. (b) Krueger, J. K.; Stock, J.; Schutt, C. E. *Biochim. Biophys. Acta* **1992**, *1119*, 322. (c) Jeltsch, A.; Alves, J.; Wolfes, H.; Maass, G.; Pingoulg, A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8499. (d) Bullock, T. L.; Branchaud, B.; Remington, S. J. *Biochemistry* **1994**, *33*, 11127. (e) Jeltsch, A.; Pleckaityte, M.; Selent, U.; Wolfes, H.; Siksnys, V.; Pingould, A. *Gene* **1995**, *157*, 157. (f) Wang, Q.; Withers, S. G. *J. Am. Chem. Soc.* **1995**, *117*, 10137. (g) Schweins, T.; Geyer, M.; Scheffzek, K.; Warshel, A.; Kalbitzer, H. R.; Wittinghofer, A. *Struct. Biol.* **1995**, *2*, 36. (h) Scheltinga, A. C. T. v.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W. *Biochemistry* **1995**, *34*, 15619. (i) Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. *Nat. Struct. Biol.* **1996**, *3*, 638. (j) Cupp-Vickery, J. R.; Han, O.; Hutchinson, C. R.; Poulos, T. L. *Nat. Struct. Biol.* **1996**, *3*, 632. (k) Thunnissen, M. M. G. M.; Taddei, N.; Liguri, G.; Ramponi, G.; Nordlund, P. *Structure* **1997**, *5*, 69. (l) Crane, B. R.; Arvai, A. S.; Gachhui, R.; Wu, C.; Ghosh, D. K.; Getzoff, E. D.; Stuehr, D. J.; Tainer, J. A. *Science* **1997**, *278*, 425. (m) Reion, M. D.; Stoeckler, J. D.; Guida, W. C.; Walter, R. L.; Ealick, S. E. *Biochemistry* **1997**, *36*, 11735. (n) Pingould, A.; Jeltsch, A. *Eur. J. Biochem.* **1997**, *246*, 1.

(25) (a) Carter, P.; Wells, J. A. *Science* **1987**, *237*, 394. (b) Carter, P.; Abrahmsén, L.; Wells, J. A. *Biochemistry* **1991**, *30*, 6142. (c) Corey, D. R.; Willet, W. S.; Coombs, G. S.; Craik, C. S. *Biochemistry* **1995**, *34*, 11521. (d) Matsumura, I.; Kirsch, J. F. *Biochemistry* **1996**, *35*, 1881.

(26) Gefflaut, T.; Blonski, C.; Perie, J.; Willson, M. *Prog. Biophys. Mol. Biol.* **1995**, *63*, 301.