Molecular Modelling of the Interaction of Cyanoacrylate Inhibitors with Photosystem II
Part 1. The Effect of Hydrophobicity of Inhibitor Binding

Simon P. Mackay and Patrick J. O’Malley
Department of Chemistry, Sackville Street, UMIST, Manchester, M60 1 QD, U.K.
Z. Naturforsch. 48c, 773–781 (1993); received February 22/June 3, 1993

Cyanocrylate Inhibitors, Photosystem II, Hydrophobicity

The secondary quinone binding site of photosystem II is also the binding site for many different herbicides. The 2-cyanoacrylate inhibitors are a potent class of electron transfer inhibitors which bind at this site and are extremely sensitive to minor structural variation. In order to understand their mode of binding, we have studied the interaction between the inhibitors and receptor in the D1 protein binding region (residues Leu 210 to Val 280) in terms of non-bonded intermolecular forces. The intermolecular energy was calculated by van der Waals and electrostatic interactions after energy minimization of the combined structures to reduce inter and intramolecular strain. We have identified specific amino acid residues within the binding protein which are instrumental in binding the herbicide and have shown that the spatial arrangement of the herbicide functional groups within the binding site rather than their lipophilicity is the determining factor in binding efficiency.

Introduction

The photosynthetic electron transport (PET) inhibitors are believed to act by displacing the secondary quinone from its binding niche in the photosynthetic reaction centre thus interrupting electron flow [1–3]. In the photosynthetic purple bacterium *Rhodopseudomonas viridis*, X-ray crystallographic analysis of the reaction centre complexed with the triazine PET inhibitor terbutryn has allowed characterization of the binding site within the L protein [4]. So far, it has not been possible to analyse the photosystem II (PS II) reaction centre by X-ray crystallography, but a variety of chemical and biochemical studies have shown that the D1 peptide, one of the several peptides associated with the PS II reaction centre, contains the inhibitor receptor site [3, 5]. Competitive displacement studies have indicated that a variety of the PET inhibitors, including the ureas, triazines and cyanoacrylates have the same receptor region [1, 2, 6]. It appears that for each class of PS II herbicide there are multiple binding sites which accounts for the differential sensitivity and an overlap of some of these sites between the different PET inhibitors accounts for their competitive displacement [7–9].

Comparison between the L protein of the bacterial systems and the D1 protein of PS II suggest distinct structural and functional similarities. In particular, the secondary quinone binding region of both proteins possess significant homology in the amino acid sequences [10–13]. The precise structure of the D1 protein is subject for speculation and models of the herbicide binding niche have been based upon comparisons with the L protein [1, 14] and analysis of the positions of mutations which affect inhibitor binding [15–24]. A number of studies concerning the inhibition of photosynthetic electron flow in PS II by a series of cyanoacrylate inhibitors have shown that they are extremely sensitive to minor structural variation [25–28]. For example, introduction of an ether linkage in the ester side chain [26], the steric arrangement of an alkyl substituent (R2) at the β-carbon [29, 30] and the number of methylene groups separating a phenyl from the amino group (R1) [30, 31] all have marked influences on inhibitory activity. Consequently, this class of inhibitors have proved useful probes for determining the nature of the receptor topography. Modelling of their interaction with the binding site through calculation of nonbonded interactions between amino acid residues and functional groups of the inhibitors can help identify regions of the receptor responsible for structural selectivity. Agreement between experimentally determined binding data and calculated intermolecular energies between the
ligand and protein will further confirm our structural model of the interactions in this region.

It is recognized that the conformation of both the ligand and the receptor may change on binding to adopt higher energy structures to match the steric requirements of the site. However, the conformational lability of the cyanoacrylate molecule means an X-ray determined structure of the compound in a crystal lattice provides a useful starting point for modelling such interactions. We have used the crystal structure of the highly active compound (Z)-ethoxyethyl-3-(4-chlorobenzylamino)-2-cyano-4-methylpentanoate 1 as the basis for modelling the interaction of a number of cyanoacrylate derivatives with a model of the herbicide binding domain from the D1 protein of *Pisum sativum*.

```
      R1NH
         |   C==C
         |   COOR3
      R2

Compound   R1               R2
1           p-C6H4CH3       CH3CHCH3
2           p-C6H4CH2       CH3CH3
3           p-C6H4CH2       CH3CHCH3
4           C6H5            CH3CH2
5           C6H5(CHO)3      CH3CH2
6           C6H5(CHO)3      CH3CH2
7           C6H5(CHO)3      CH3CH2
8           C6H5(CHO)3      CH3CH2
```

Methods

The source of the parameters used to calculate nonbonded energies is the forcefield which is the empirical fit to the potential energy surface of the molecules involved. It defines the coordinates used, the mathematical form of the equations involving the coordinates and the parameters adjusted in the empirical fit of the potential energy surface [33, 34]. The forcefield employs a combination of internal coordinates (bond distances, angles and torsions) to describe the bonded part of the potential energy surface, and interatomic distances to describe the van der Waals and electrostatic interactions between atoms. For the purpose of calculating the intermolecular energies between the atoms of two molecules, we are interested in the expressions which determine the nonbonded interaction [35, 36].

The nonbonded van der Waals interaction are represented by the first two terms in Eqn. (1) where $A_{ij}$ and $B_{ij}$ are parameters with units of kcal mol$^{-1}$ angstrom$^{-12}$ and kcal mol$^{-1}$ angstrom$^{-6}$ respectively and $R_{ij}$ is the distance between the atoms i and j in angstroms. The second component of the nonbonded intermolecular energy is the electrostatic energy, which is represented by the third expression in Eqn. (1) where $q_i$ and $q_j$ are the charges on atoms i and j and $D$ is the dielectric constant. The intermolecular energy is computed by summing the energy contributions between atoms of the two molecules. The contribution between atoms interacting with atoms in the same molecule is ignored.

$$E_{interaction} = \sum_i \sum_j \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{D R_{ij}}. \quad (1)$$

Enclosure analysis focuses on a smaller region of a molecular system in order to generate intermolecular energies between a ligand and individual amino acid residues within its binding site. In doing so, Eqn. (1) is used to calculate the interaction energy between the atoms of the ligand and the atoms of the residues which fall into a defined sphere of a given radius around that ligand. This allows the identification of the main residue-ligand nonbonded interactions which make up the intermolecular energy between the two molecules.

Enger et al. [37] have stated, when investigating the binding of triazine herbicides to bacterial reaction centres, that parameters within a molecular mechanics/dynamics program may have to be adjusted to achieve a correlation with biological and calculated data. Only when this correlation has been found can the impact of different substituents be compared and new compounds designed. Our study demonstrates that such adjustments are not required when examining cyanoacrylate binding to PS II, there being an apparent correlation between biological and calculated data using the standard forcefield parameters. It is important to point out that the assumptions of Eqn. (1) *i.e.* modelling the interaction energy in terms of only the electrostatic and van der Waals terms is a severe approximation. It does, for example, ignore completely polarization effects. The widespread use of Eqn. (1) to calculate intermolecular interactions depends critically on the parameterization method. In essence, by taking a certain combination of values for the
electrostatic and van der Waals parameters one can model the total interaction energy. The combination of electrostatic and van der Waals parameters supplied in the Discover program have been shown to be suitable for this purpose. Modification solely of the electrostatic component is therefore likely to lead to a breakdown of the approximations inherent in Eqn. (1).

The model used was the D1 protein from the PS II photosynthetic reaction centre of Pisum sativum (coordinates supplied by J. Nugent, Department of Biology, University College London). The Qb binding domain was represented by residues Leu 210 to Val 280 and all other residues were deleted from the model for simplification. All hydrogen atoms, polar and non-polar were included. The model of the cyanoacrylate was built from the atomic coordinates for the non-hydrogen atoms determined by X-ray crystallography [32] using the crystal building facility in the CERIUS molecular graphics program (Version 3.1, Molecular Simulations Ltd., Cambridge, U.K.). The hydrogen atoms were subsequently added and minimized accordingly. Atom partial charges and potentials for both protein and herbicide models were assigned according to the parameters defined within the Consistent Valence Force Field (CVFF) used by the Discover (Version 2.8.0) molecular simulation program to be used in conjunction with the Insight II (Version 2.1.0) molecular graphics modelling program (Biosym Technologies, San Diego, California).

The herbicide was orientated within the Qb binding site employing the three dimensional stereo viewing facility of the Insight II modelling program until a minimum intermolecular energy was achieved. Energy minimization of the combined structures involved constraining the backbone of the protein to relieve unfavourable interactions between the amino acid side chains and herbicide hydrogen atoms whilst maintaining the coordinates of the heavy atoms determined from the crystal structure [32]. This was performed using steepest descents and conjugate gradients algorithms successively until the average first derivative was less than 0.005 kcal mol\(^{-1}\) angstrom\(^{-1}\). The cancellation of the nonbonded interactions between atoms after a specified cutoff distance was not carried out during minimization in order to achieve a more accurate final structure. A dielectric constant of one was employed throughout the study. A sphere of 8 angstrom radius around each functional group of the cyanoacrylate was used to calculate the nonbonded interaction energy between the herbicide and individual amino acid residues of the binding site.

The minimum energy structure obtained for the cyanoacrylate derivative in the binding site was modified according to the functional group substitutions in compounds 2–8. The heavy atoms of the structure were fixed whilst the new functional groups were orientated within the Qb binding site until a minimum intermolecular energy was achieved. Energy minimization of the combined structures was performed as before.

**Results and Discussion**

Crystal data for compound 1 [32] (Fig. 1) reveals an effectively planar central core involving a delocalized amino-cyano-enoate \(\pi\) electron system. A six-membered ring structure stabilized by the formation of a hydrogen bond between the benzylamino-NH and the ester carbonyl oxygen is also evident. The iso-propyl \(\beta\)-alkyl substituent has the two methyl groups orientated away from the benzylamino function. When modelling the interaction of this structure with the protein, any deviation from this low energy structure must be compensated by a net improvement in intermolecular energy (an increase in negativity). We have found that a favourable interactive model of compound 1 with the protein can be achieved without deviating from the crystal structure with an intermolecular energy of \(-23.8\) kcal mol\(^{-1}\) (Fig. 2). The cyanoacrylate occupies the same binding niche as the secondary quinone and other PET inhibitors in the

![Fig. 1. Conformation of compound 1 according to the coordinates of the heavy atoms determined by X-ray crystallography.](image-url)
Fig. 2. Stereo plot of the binding of the cyanoacrylate derivative 1 to the D1 protein model of *Pisum sativum*. The iso-propyl β-substituent is in the crystal conformation and in conflict with the Ala 251 and Asn 267 residues. The residue numbers of the D1 protein are prefixed with the number 1. For viewing stereo plots a reflecting stereoscope (available from Aldrich) permits three dimensional perception.

region of the D1 protein linking the forth and fifth transmembrane helices. It appears that the ether substituent contributes by forming a predominantly electrostatic interaction with the Ser 268 residue (−3.6 kcal mol⁻¹) and probably involves a hydrogen bond with the hydroxyl side chain of the amino acid. This confirms the observation that the introduction of an ether function in the ester group enhances binding activity [26]. The same residue also interacts favourably with the carboxylate (−1.4 kcal mol⁻¹) and cyano group (−1.0 kcal mol⁻¹).

The phenyl substituent, in common with aromatic substituents of other herbicides such as DCMU and the triazines [38–41], interacts through mainly van der Waals ring stacking interactions with the phenyl side chain of the Phe 255 residue (−5.1 kcal mol⁻¹). It occupies a hydrophobic pocket comprising of the aromatic residues Phe 211 (−1.9 kcal mol⁻¹), Tyr 262 (−2.7 kcal mol⁻¹) and Phe 274 (−1.0 kcal mol⁻¹). In all reported experimental studies of cyanocrylate inhibition of the Hill reaction, activity was increased by substitution at the β-carbon, reaching a maximum when a β-ethyl substituent was present and declining as the chain length of the group was further increased [29–31]. The large and consistent amplification of potency by the presence of a favourable β-substituent in aralkyl-substituted compounds (R₂) implies a high degree of spatial specificity in the interaction of the aryl ring with the binding domain. A well fitting β-alkyl substituent appears to play an important role in determining the orientation of the aralkyl group in the receptor in order to maximize interaction with a specific binding region within the hydrophobic domain. In addition, it appears that a branch at the α-carbon of the β-substituent is a particularly favourable structural feature [30], for example the β-iso-propyl derivative is considerably more active than the n-propyl isomer. Whilst it is recognized that an α-methyl substituent is critical for high affinity binding, it has not been established if this is due to hydrophobic interaction with a highly constrained pocket which can accommodate a methyl group. There are significant repulsive interaction energies between the iso-propyl β-substituent and the Ala 251 (+27.5 kcal mol⁻¹) and Asn 267 (+7.5 kcal mol⁻¹) at the top of the binding niche which form part of the constrained pocket. We have examined the intermolecular energy profile by rotating this group within the pocket with respect to the enoate plain and found the lowest interaction with the group to be at an angle of −100° from the crystal structure.
Fig. 3. The relative positioning of the β-substituent with respect to the enoate plain. The first conformation of the iso-propyl group is according to the crystal structure. In the second conformation, the iso-propyl is rotated −100° relative to the crystal structure and gives the optimum intermolecular energy with the protein. Note that the α-methyl group is proximal to the benzyl group. The third conformation shows a β-ethyl substituent with the α-methyl again proximal to the benzyl group. The fourth conformation has the β-n-propyl group away from the benzyl group as a result of repulsion by the His 252 residue in the β-substituent binding pocket.

(Fig. 3). The repulsive energies with the Ala 251 and Asn 267 residues were reduced significantly to −1.7 and −1.3 kcal mol⁻¹ respectively which resulted in an overall intermolecular energy between the cyanoacrylate and the protein of −61.0 kcal mol⁻¹, a marked improvement on the crystal structure. This energy also reflects an improved binding affinity compared to DCMU [42] (−19.7 kcal mol⁻¹) which experimental studies indicate [6]. The resultant repulsive energy between the iso-propyl group and the methylene hydrogens of the benzyl group of +11.3 kcal mol⁻¹ is compensated by this corresponding favourable interaction of the compound with the protein. We would argue that the role of the α-methyl substituent in high affinity binding is more important for determining the configuration of the aralkyl group within the binding domain to achieve maximum hydrophobic interactions with the receptor rather than forming good hydrophobic interactions itself, although the two factors are inextricably linked. In order to bind tightly to the receptor, the constraints of the pocket which accommodate the β-iso-propyl group necessitate its rotation by approximately 100° in order to overcome repulsive forces with the Ala 251 and Asn 267 residues which form part of the pocket. As a result, there is a favourable binding interaction between the iso-propyl group and the protein of −11.8 kcal mol⁻¹. In the rotated conformation, the α-methyl group of the β-substituent sterically repulsed the benzyl group, forcing it downwards into the hydrophobic pocket to maximize interactions. Similarly, a β-ethyl substituent (Fig. 3, compound 2) has a binding interaction of −8.4 kcal mol⁻¹ with the protein when the methyl group is rotated proximal to the benzyl group by −100° as before. This again would result in the alignment of the benzyl group in its hydrophobic pocket to achieve an intermolecular energy of −55.9 kcal mol⁻¹ between the herbicide and protein.

In contrast, a β-n-propyl substituent (compound 3) is too large to be accommodated proximal to the benzyl group after rotation, being repulsed by the His 252 residue (+14.7 kcal mol⁻¹) within the β-group binding pocket. The n-propyl group must be rotated away from this residue (Fig. 4) in order to fit into the highly constrained pocket which means in the absence of an α-methyl in the β-substituent there is no repulsion of the benzyl group into its binding domain and is therefore less tightly bound (−37.4 kcal mol⁻¹).

It appears therefore that the area of the receptor which interacts with the β-substituent is a highly constrained pocket. In order to achieve a tightly bound configuration, the constraints of this pocket force a rotation of the β-substituent to minimize repulsion with the Ala 251 and Asn 267 residues. In the presence of an α-methyl group, the rotation ensures that the aralkyl group of the inhibitor is contained within a hydrophobic pocket in the binding niche thus maximizing van der Waals interactions.

Studies with aryl and aralkyl derivatives with varying chain length 4–8 suggest that the proposed interaction of the hydrophobic group with an unconstrained lipophilic area in the thylakoid membrane is more complicated [30, 31], with the hydrophobic interactions being modified by steric effects [29, 43]. The phenylamino derivative 4 is a weak inhibitor [31], but inclusion of a methylene group between the phenyl and amino function 5 produces a large increase in activity. Further lengthening of the carbon chain produces a stepwise increase in potency as the chain length increases. This has been explained in terms of crystal packing phenomena [44] whereby the greater binding affinity with the protein is a result of the phen-
yl group having optimum hydrophobic interactions, the orientation being dictated by the chain length.

We have demonstrated that a similar trend exists between the intermolecular energy calculated for the inhibitor-binding site interaction and the number of carbon atoms between the phenyl and amino functions (Table I).

Table I. Intermolecular energies between D1 protein and cyanoacrylate derivatives.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Intermolecular Energy [kcal mol(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Van der Waals</td>
</tr>
<tr>
<td>4</td>
<td>-2.8</td>
</tr>
<tr>
<td>5</td>
<td>-49.4</td>
</tr>
<tr>
<td>6</td>
<td>-40.7</td>
</tr>
<tr>
<td>7</td>
<td>-58.8</td>
</tr>
<tr>
<td>8</td>
<td>-49.4</td>
</tr>
</tbody>
</table>

Calculated according to Eqn. (1).

A weak binding interaction exists between compound 4 and the protein (Table I) which reflects the low inhibitory activity. The \(\beta\)-ethyl substituent is unable to occupy its optimum position within the binding pocket due to repulsion from the phenyl group which is directly adjacent to the amino function. It is forced into direct conflict with the Asn 267 residue resulting in a repulsion energy of +7.3 kcal mol\(^{-1}\) (Fig. 5). The phenyl substituent is able to ring stack with the Phe 255 residue resulting in an attractive energy of -4.6 kcal mol\(^{-1}\) (Table II) but is unable to interact with other residues in the hydrophobic pocket as it is too far removed. Rotation of the phenyl group also minimizes repulsion between the aromatic hydrogen atoms and the amino hydrogen, which is also rotated out of the enoate plain to the same effect. This results in the hydrogen bond between the carbonyl and amino groups breaking due to the loss of planarity. A major repulsive interaction of +5.7 kcal mol\(^{-1}\) is

Table II. Comparison of the hydrophobic interactions between the phenyl substituent and amino acid residues in the hydrophobic pocket of the binding site.

<table>
<thead>
<tr>
<th>Residue in D1 protein</th>
<th>Number of carbons in alkyl chain</th>
<th>Intermolecular energy in [kcal mol(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe 211</td>
<td>1</td>
<td>-0.2 - 1.2 - 3.0 - 2.7 - 2.0</td>
</tr>
<tr>
<td>Met 214</td>
<td>1</td>
<td>-0.1 - 1.0 + 3.9 - 1.4 - 1.6</td>
</tr>
<tr>
<td>His 215</td>
<td>1</td>
<td>-0.2 - 1.4 - 1.0 - 0.3 - 0.1</td>
</tr>
<tr>
<td>Phe 255</td>
<td>1</td>
<td>-4.6 - 5.2 - 4.0 - 3.0 - 0.3</td>
</tr>
<tr>
<td>Tyr 262</td>
<td>1</td>
<td>-0.1 - 2.7 - 2.9 - 4.0 + 1.8</td>
</tr>
<tr>
<td>Ser 264</td>
<td>1</td>
<td>+5.7 - 0.4 - 0.2 - 0.2 - 0.1</td>
</tr>
<tr>
<td>Phe 265</td>
<td>1</td>
<td>-1.0 - 1.3 - 0.4 - 1.3 - 1.0</td>
</tr>
<tr>
<td>Leu 271</td>
<td>1</td>
<td>-0.5 - 0.1 - 1.0 - 2.0 - 1.9</td>
</tr>
<tr>
<td>Phe 274</td>
<td>1</td>
<td>-0.2 - 1.6 - 1.3 - 3.6 - 3.3</td>
</tr>
</tbody>
</table>

Calculated according to Eqn. (1).
Fig. 5. Stereo plot of the binding of cyanoacrylate derivative 4 to the D1 protein model of *Pisum sativum*. The β-ethyl substituent is forced into direct conflict with the Asn 267 residue due to repulsion from the phenyl group. The residue numbers of the D1 protein are prefixed with the number 1.

evident between the phenyl group and the Ser 264 residue. The overall effect is to produce a less favourable binding interaction than those compounds which have an alkyl chain separating the amino and phenyl functions. Table I shows the changing interaction energies between the different phenyl substituents and the amino acid residues within the hydrophobic pocket. The interaction with the Phe 255 residue decreases as the phenyl group moves away whereas a stronger interaction with the Phe 274 residue is established which is further down the transmembrane helix down which the aralkyl group is projected. Superimposition of compounds 5–8 upon the plastoquinone in its binding site shows the aralkyl substituents following the same path through the protein as the isoprenoid side chain (Fig. 6) as do aliphatic alkyl chain substituents. The improved binding energies of 5 and 7 with the protein appear to be due to a combination of favourable interactions with the aromatic residues of the binding pocket and the absence of repulsive energies which are evident for compounds 6 and 8 (Met 214 and Tyr 262 respectively). The length of the alkyl chain determines the orientation of the phenyl group with respect to the residues within the pocket, and a chain length of three carbons appears to maximize the favourable interactions, for example, ring stacking can take place between the phenyl substituent and the side chain of Tyr 262 (Fig. 7).

Fig. 6. Stereo plot of the binding of cyanoacrylate derivative 8 to the D1 protein model of *Pisum sativum* superimposed onto the binding orientation of the secondary plastoquinone.
Our results confirm that changes in activity as a function of hydrophobicity cannot be explained simply in terms of longer carbon chains partitioning more effectively into the phospholipid phase of the membrane [30, 31], for example the phenyl group of compound 8 with the longest carbon chain is within the protein complex of the reaction centre according to our model and does not protrude into the membrane space (Fig. 8). However, we do not discount lipophilicity playing a role as regards transport and distribution of the herbicide to its site of action. The interaction with the recep-
tor depends more importantly on the phenyl substituent adopting a favourable orientation with respect to residues in its vicinity. This is determined by two factors. In the shorter chain derivatives \(n = 0, 1\), the \(\beta\)-substituent is instrumental in determining the orientation of the phenyl group and thus binding affinity. As the chain increases, the tetrahedral geometry of the methylene groups in the chain determines the position of the phenyl group within the binding pocket.

The limitations of theoretical methods must be kept in perspective. Much progress has been made in the development of approaches to represent as accurately as possible the energy of complex molecules. Quantum mechanical calculations are not yet feasible for calculations involving biomolecules because of the computational expense for dealing with such large numbers of particles. The approximations of the potential energy function and mechanics employed in modelling determines the reliability of the results which reflects a balance of the requirements of computational efficiency and meaningful accuracy.

Acknowledgements

We thank the Agriculture and Food Research Council and the Royal Society for their support of this work. We acknowledge S.V. Ruffle and J.H.A. Nugent for the coordinates of PS II.

References