Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Automated ligand placement and refinement with a combined force field and shape potential

S. Wlodek,* A. G. Skillman and A. Nicholls

OpenEye Scientific Software, 3600 Cerrillos Road, Santa Fe, NM 87507, USA

Correspondence e-mail: stan@eyesopen.com

Received 19 January 2006
Accepted 1 May 2006

An automated computational procedure for fitting a ligand into its electron density with the use of the MMFF94 force field and a Gaussian shape description has been developed. It employs a series of adiabatic optimizations of gradually increasing shape potential. Starting from a set of energy-relaxed ligand conformations, the final results are structures realistically strained to fit the crystallographic data.

1. Introduction

Although protein crystal structure building is a well established procedure and a number of software tools that perform the task of automated construction of three-dimensional protein coordinates are available to crystallographers [e.g *REFMAC* (Murshudov *et al.*, 1997), *RESOLVE* (Terwilliger, 2003), *MAID* (Levitt, 2001) or *ARP*/*wARP* (Perrakis *et al.*, 1999)], obtaining realistic ligand structures from the corresponding electron density is still not a fully automatic process. This is largely a consequence of the greater complexity of medicinal chemistry and chemical informatics compared with protein chemistry and information science and can be a significant bottleneck in the application of high-throughput crystallography to pharmaceutical lead identification.

Current methods of ligand fitting that are based on either topological analysis of electron density (Menéndez-Velázquez & García-Granda, 2003), global optimization of position and conformation of a ligand in a density blob (Diller *et al.*, 1999), interatomic distance matrices (Koch, 1974; Main & Hull, 1978; Cascarano *et al.*, 1991; Altomare *et al.*, 2002; Zwart *et al.*, 2004) or on varying torsion dihedral angles of shape-matched ligand conformations (Oldfield, 2001) are unable to prevent creation of high-energy, sometimes even chemically unrealistic, ligand models. As a result, there are a number of PDB ligands with unlikely high-energy structures. For example, the PDB structure of an inhibitor of RNA polymerase (PDB code 1nhu) has significant repulsion between the two methylene groups, or consider the adenine dinucleotide PDB code 1xqd which has highly distorted phosphorus–oxygen coordination geometries (see below).

In a recent study of 100 public protein–ligand complexes and a further 50 private structures, Perola & Charifson (2004) determined that modern force fields (OPLS, MMFF) evaluate 90% of total strain energies, as defined relative to the global solvated minima, as less than 38 kJ mol^{-1} and of local strain energies, defined relative to the closest local minima, as less than 21 kJ mol^{-1}. As such, an alternate method of ligand fitting suggests itself: start with an ensemble of low-energy chemically correct conformations and adapt each to the ligand density using a modern force field. Important aspects of this approach would be the adequate sampling of conformational

space, the appropriate combination of forces from the electron density and the force field and the choice of a force field that is general and accurate enough for the tasks faced by industrial crystallography. In this report, we demonstrate such a procedure.

Here, the potential function to be minimized is

$$V = V_{\rm FF} + \lambda V_{\rm shape}. \quad (1)$$

The independent variables here are the ligand coordinates or internal degrees of freedom (*e.g.* torsions); $V_{\rm FF}$ represents the internal energy of the ligand and $V_{\rm shape}$ the (ligand/electron-density) overlap. Details of both terms are presented in §2. λ is a mixing parameter between internal energy and the shape overlap. It represents the degree to which we wish the shape of the density to dominate. A common practice when confronted with such hybrid functions is to find a value of λ that works generally, *i.e.* it is assigned an heuristic value. Such approaches work in crystallographic refinement of proteins, presumably because of the minimal chemical diversity of amino acids compared with general chemistry. However, a single value does not appear to work well for ligand refinement, leading to either under-fitting to the density or over-straining of the ligand. Instead, we have developed an adiabatic approach, as described below, where a series of minimizations are performed with gradually increasing values of λ, each using the previous minimization as a starting point. This allows us to check for and avoid over- and under-fitting on a molecule-by-molecule basis. We show that combining adiabatic optimization with ligand electron-density identification, ligand conformer generation and initial orientation of ligand to electron density, we can generate low-energy, high-quality ligand models in a fully automated manner.

In §2 we describe our potential function (1) and the computational procedures for adiabatic fitting, ligand electron-density identification, conformer generation and initial ligand placement. In §3.1 we tested the technique against a model set of 800 docked ligands represented either as molecular volumes or estimated electron density. These docked structures were minimized against their target protein and so included both local and global strain. Finally, in §3.2 we present the results of fitting ligands into their experimental electron density obtained from X-ray structure determination of the corresponding receptor–ligand complexes. We demonstrate that when starting with only an electron-density map and connectivity record of the ligand, this method can reliably produce low-energy ligand models that are very close in Cartesian space to the ligand models submitted to the PDB.

2. Theory and methods

2.1. Potential function

We have chosen the Merck Molecular Force Field (MMFF94) as the first term, $V_{\rm FF}$, of potential (1),

$$V_{\rm FF} = V_{\rm MMFF94}, \quad (2)$$

where the functional form of $V_{\rm MMFF94}$ is described by Halgren (Halgren, 1994*a*,*b*,*c*,*d*; Halgren & Nachbar, 1996). Halgren designed MMFF94 to cover a wide range of chemical functionalities encountered in medicinal chemistry to an accuracy frequently encountered in *ab initio* quantum mechanics. As such, it appears to be an optimal choice for ligand fitting. In our implementation, we allow the optional removal of Coulomb terms from $V_{\rm MMFF94}$. It is a predicate of this approach not to use the protein structure to guide the ligand positioning, so as not to bias the result with already uncertain information. However, protein electrostatics can dramatically influence the ligand strain, for instance by compensating for internal electrostatic repulsions. As such, a mimic of protein electrostatic compensation is to remove the Coulombic term. Consequences and examples are discussed in §3.2.2.

The term 'shape potential' used in this paper is derived from the work of Grant *et al.* (1996). They have pioneered the use of Gaussian functions to represent molecular volumes and overlaps. Here, the overlap between two molecular Gaussian volume functions representing molecules A and B is

$$V_{AB} = \int \rho_A \rho_B \, {\rm d}\mathbf{r}, \quad (3)$$

where ρ_A and ρ_B are Gaussian shape densities for molecules A and B defined in terms of atomic Gaussian functions g_i, where r_i is the distance from atom i.

$$\rho_A = 1 - \prod_{i \in A} (1 - g_i), \quad (4)$$

$$g_i = p_i \exp(-\alpha_i r_i^2). \quad (5)$$

Owing to the simplicity of the functional form (5), the value of the shape overlap (3) and its derivatives are easy to calculate analytically or from a grid representation. In the above formula, the Gaussian widths α_i have distinct values for every element. The value of α_i for atom i determines its 'Gaussian atomic radii', σ_i,

$$\alpha_i = \kappa_i / \sigma_i^2, \quad (6)$$

where the parameter κ_i depends on the prefactor p_i and is chosen in such a way that the volume integral over Gaussian (5) is equal to the volume of a sphere of radius σ_i (Grant *et al.*, 1996). Assigning a single Gaussian function to an atom and careful selection of a prefactor p_i (2.7 from Grant & Pickup, 1995) can result in very accurate molecular-shape functions, *e.g.* to with 0.1% of the hard-sphere molecular volume. In this report we use the same parameterization but ignore the atom–atom overlaps generated in (4).

When fitting to electron density, further improvement might be expected from molecular representations that use more than one Gaussian atom-centered functions in linear combination, $\sum_{k=1}^{n} p_k \exp(-\alpha_k r_i^2)$, by analogy to the calculation of X-ray scattering factors approximated by a combination of two or five Gaussians (Agarwal, 1978). In the current work, we have used both single and multiple Gaussian representations for validation. In the latter case, we adopted the five prefactors

and widths per atom published in *International Tables for X-ray Crystallography* (1974, Vol. IV).

2.2. Monitoring shape similarity

The ligand density-fitting process requires a monitoring of the progress of the overlap between the molecule and the crystallographic electron density. For this purpose, we use a Tanimoto coefficient, which is a well known similarity measure between objects (Sneath & Sokal, 1973; Willet *et al.*, 1998), calculated at every λ,

$$T(\lambda) = \frac{V_{lt}}{V_l + V_t - V_{lt}}, \qquad (7)$$

where V_l and V_t are self-volume overlaps of the fitted ligand and electron-density target, respectively, and V_{lt} is the overlap volume between the two objects. The fit is terminated when the function $T(\lambda)$ reaches either a maximum or an apparent plateau or the strain energy reaches a predetermined limit. Determining a plateau region can be difficult, especially for low-resolution density where the noise in the data can produce false local mimina in the adiabatic process. To prevent premature termination, the $T(\lambda)$ curve is smoothed with the uniform cubic B-spline curve,

$$T_{\rm ave}(\lambda) = \tfrac{1}{6}[T(\lambda - \Delta\lambda) + 4T(\lambda) + T(\lambda + \Delta\lambda)]. \qquad (8)$$

In these plateau regions, small increases in the Tanimoto coefficient correspond to large jumps in force-field energy $V_{\rm FF}$, typically seen in compressions of bond lengths and angles. This can lead to abnormal strain energies in a minority of cases. These can be fixed by an optional process that relaxes the ligand under MMFF with a flat-bottom harmonic potential of width 0.05 Å imposed on all heavy-atom coordinates and $\lambda = 0$. Although such a relaxation causes a negligible displacement (of the order of 10^{-2} Å in r.m.s.d.), it can sometimes remove significant strain (*e.g.* 80% of the internal energy). That such a procedure is useful was unsurprising. Although the Tanimoto maxima criteria is powerful, it is still a heuristic measure as to the optimal real-space fit and occasionally leads to unnecessary strain.

2.3. Identifying electron-density shapes

A starting point for our method is the definition of an electron-density shape or 'blob'. The source data may be a full density map or a difference map. We apply a blob-finding algorithm to enumerate all closed surfaces produced by isocontouring the electron density. The isocontour value used is the smallest that still produces a discrete contour of volume comparable to the ligand molecule. This approach can fail with low-resolution density maps. The solution we adopted in such cases is to interpolate and smooth the electron-density grid to a finer resolution. We validated this procedure on 11 protein–ligand electron-density maps of $2mF_{\rm o} - DF_{\rm c}$ type downloaded from the Electron Density Server (Kleywegt *et al.*, 2004) in ccp4 format. The results of ligand fitting into resultant electron-density blobs are reported in §3.2.

2.4. Conformations and starting positions

Once the electron-density blob corresponding to the ligand has been determined, an ensemble of ligand conformations are generated. Here, we use the well known conformational expansion tool *Omega* (Open Eye Scientific Software, Santa Fe, NM, USA). One of the advantages of *Omega* is the facility to generate almost exhaustive lists of conformations for a given input of connectivity. The number of conformations generated depends on the flexibility and composition of the molecule, but can range from hundreds to tens of thousands. It would be impractical at this time to apply the adiabatic approach to every such conformer. Instead, each is first rigidly overlaid onto the density blob by alignment of moments of inertia. This produces four starting points for non-symmetric ligands. It is possible that highly symmetric ligands may require more starting orientations, but this has not been observed in practice. Each pose is then rigid-body optimized and sorted according to overlap Tanimoto. The top scoring conformers, typically five to ten, are then flexibly fitted to the electron density.

2.5. Model ligands and model electron density

As a preliminary test, the techniques presented here were validated on a set of 800 small molecules docked into human

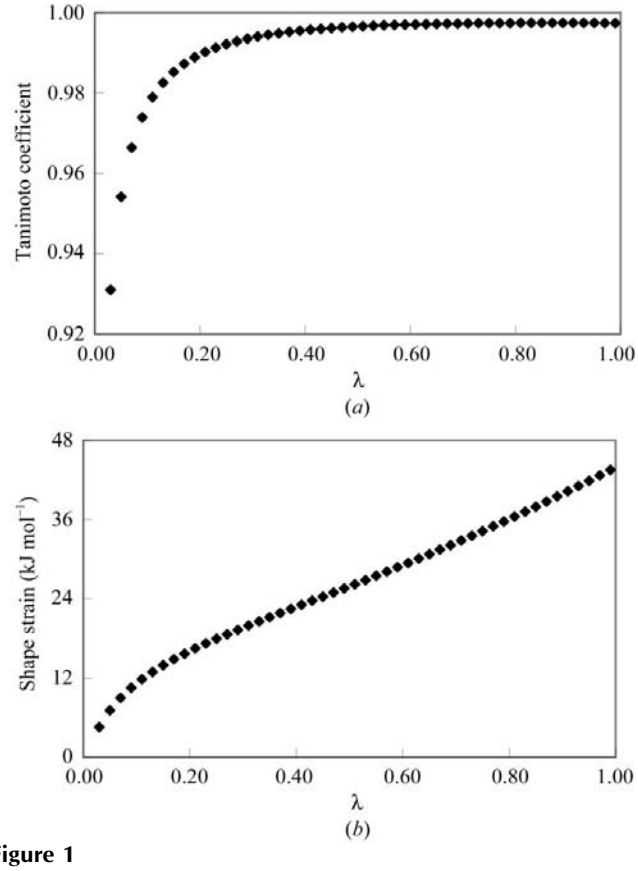

Figure 1
Tanimoto coefficient (*a*) and shape-strain energies (*b*) as functions of shape perturbation parameter λ for fitting molecule 1 onto its model structure. The maximum Tanimoto coefficient occurs at a λ value of about 0.85.

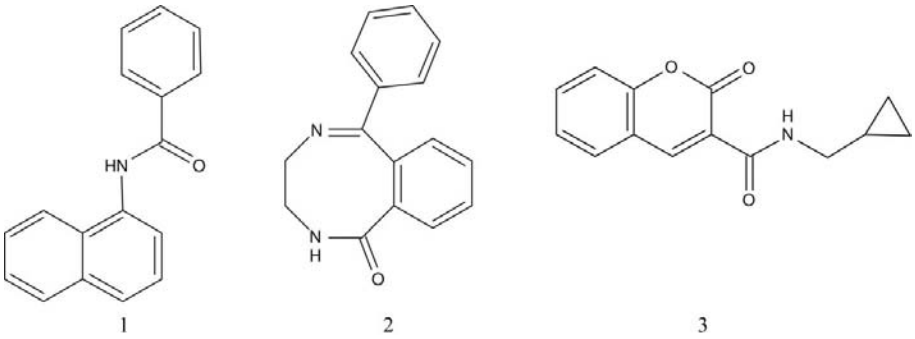

Figure 2
Examples of docked ligands into human p38 MAP kinase.

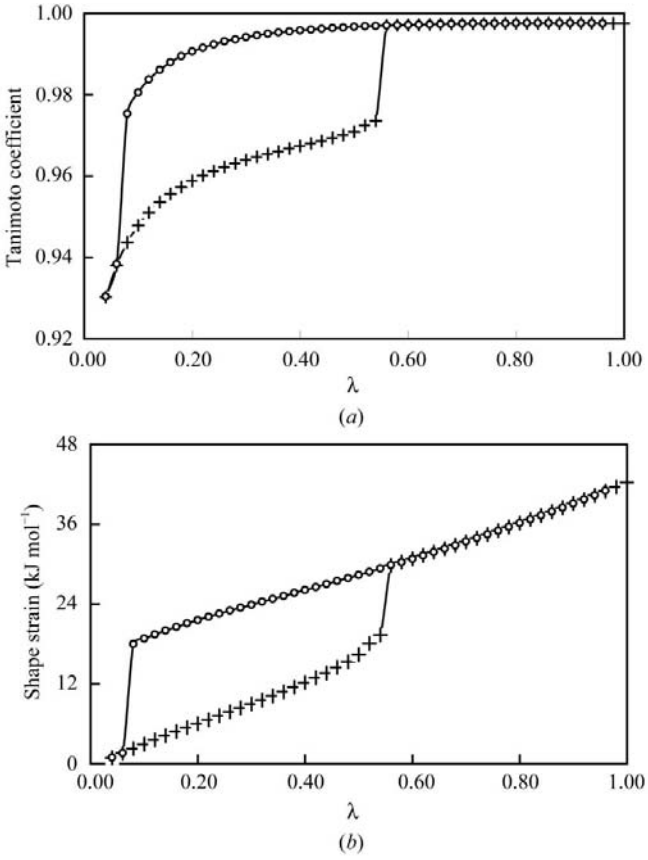

Figure 3
Tanimoto coefficient (*a*) and shape-strain energies (*b*) as functions of shape-perturbation parameter λ for fitting molecule 2 onto its model structure. Circles represent the decrease of both quantities upon backward ramping of λ. The maximum Tanimoto coefficient occurs at a λ value of about 0.89.

p38 MAP kinase (PDB code 1kv2) and minimized with MMFF (P. Charifson, personal communication). In the first test, (3) was used as $V_{\rm shape}$, validating that the method could mix MMFF energies with simple molecular shape to reproduce geometries. Because of the entirely analytic potential used in this series of test calculations, no averaging of the monitored Tanimoto coefficient (8) was necessary. In the second test, electron density for each model ligand was simulated by a linear combination of five Gaussians as described above and used either analytically or *via* a grid representation to validate the approach closer to intended use.

2.6. Fitting to experimental electron density

As a final validation of our methodology, we generated ligand models for 11 ligand–receptor complexes from the PDB. In each case, we started with an electron-density map and a coordinate-free representation of the ligand. The electron-density map was prepared by removing the density arising from the protein model and its crystallographic symmetry partners. The blob-finding algorithm was then used to identify the portion of electron density most likely to represent the ligand density. An ensemble of conformers for the ligand were generated and rigidly oriented and optimized to the ligand density. We then applied the adiabatic fitting process to the best five to ten conformers of each ligand. In several cases, the optimization process was critically influenced by the absence of formal charges present in the protein model. To compensate for this bias, an MMFF force-field without Coulombic terms was primarily used during the adiabatic optimization. Finally, each ligand model was compared by r.m.s.d. with its PDB crystal structure and by evaluation of its strain energy.

3. Results and discussion

3.1. Fitting to the model electron density

Gradual increase of the shape component in potential (1) results in structures with both higher strain and shape similarity to the target. Typical increases of both quantities with λ are shown in Fig. 1 for one of the test compounds, *N*-naphthyl benzamide (compound 1 in Fig. 2). The figures illustrate the importance of determining an appropriate stopping criteria, *i.e.* when the improvement of shape fit is not significant compared with the ever-increasing strain energy. Occasionally, adiabatic forcing fails, *i.e.* there is an abrupt shift in the response of the system, the shape fit, to the change of the adaptive parameter λ. In the theory of adiabatic evolution this corresponds to barrier crossing; the system has changed from one state to another. Here, the new state corresponds to an energy well of a different initial conformer. This can be an advantage, *i.e.* this second conformation may have been missing in the original conformational sampling or it may have been an unstable state in the absence of shape-forcing. However, there are also potential disadvantages. The value of λ for the second conformation may be inappropriately high given it was developed for the original structure. The appropriate course is to perform hysteresis, *i.e.* to perform the reverse experiment on the final structure by slowly ramping λ back down. This is illustrated in the two graphs in Fig. 3 for the adiabatic fitting of molecule 2 (Fig. 2) in which the eight-membered ring undergoes a significant conformational change

and improvement of fit at $\lambda = 0.56$. Clearly, the jump to the second conformation (which is not stable without shape forces, *i.e.* when λ is equal to zero) involves a change to a surface with quite different shape–λ properties, *i.e.* had we been able to start with this second conformation, we might have already terminated the adiabatic process. In our tests against model systems, most examples of over-strain resulted from adiabatic failure. Fortunately, either hysteresis analysis or the constrained end-of-fit minimization described above are efficacious countermeasures. Ideally, the stress applied in shape-fitting should not significantly exceed the strain of the docked ligand. Indeed, as Fig. 4(*a*) shows, this is the general case. It is seen that although the majority of ligands are aligned with strain energy close to the docked ligand energy, a few do not. Closer examination revealed that some anomalies correspond to the above-mentioned barrier crossing between conformations. These typically end with a lower than average shape overlap, *e.g.* $T \leq 0.98$. For example, in Fig. 4(*a*) the ligand (compound 3 in Fig. 2) displays significantly larger (by about 48 kJ mol^{-1}) strain than its corresponding target structure energy of about 46 kJ mol^{-1}. Here, the fitted structure with $T = 0.95$ is trapped in a potential well with the cyclopropane ring rotated with respect to the target conformation. Some 'under-stressed' ligands also show differences between fitted and target conformations, primarily arising from incorrect assessment of the shape–λ maxima. However, the majority of cases are accurately reproduced. Similar data were obtained with the use of analytical gradients of the shape function where five 'structure-factor'-like Gaussian functions per atom were used in the shape component. These are shown in Fig. 4(*b*). Only a tiny improvement in the structure overlap measured with the average r.m.s.d. between fitted and target structures of 0.01 Å was observed upon the usage of five

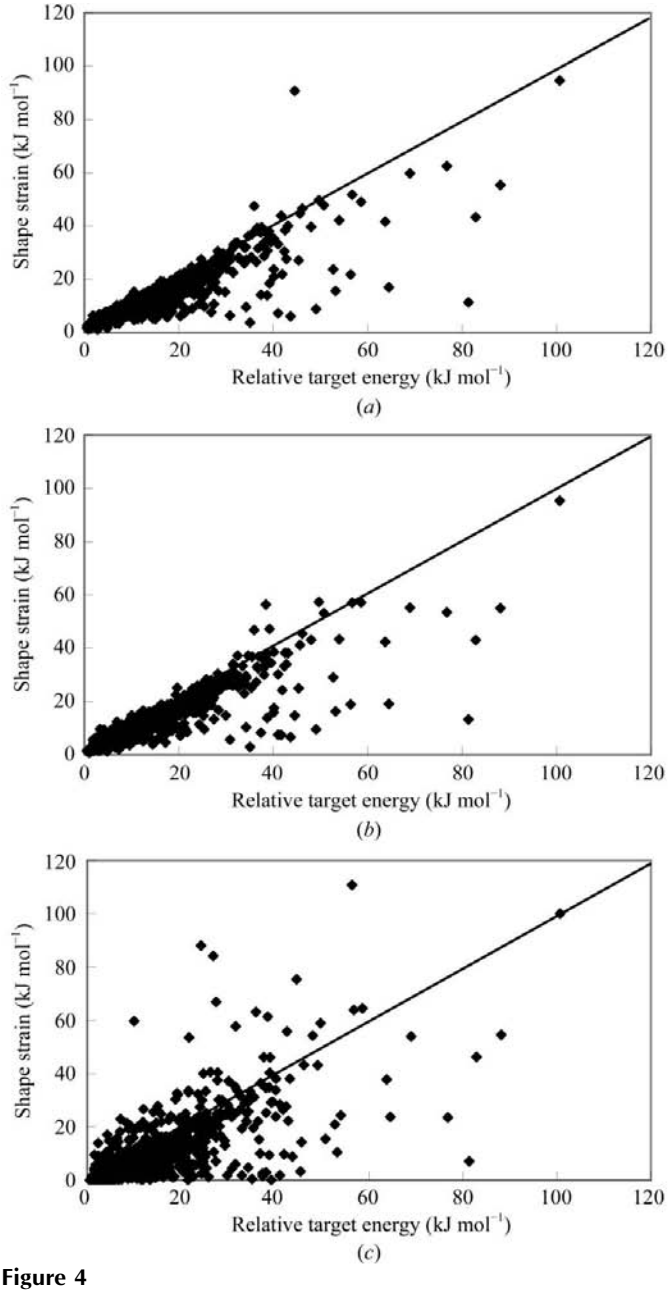

Figure 4
Shape-strain energy required to align fitted and target conformations *versus* energy difference between docked and relaxed ligands. Plots (*a*) and (*b*) were obtained with analytical gradients of potential (1) *via* equation (3) with one and five Gaussians per atom, respectively. The data in plot (*c*) were obtained with a target shape density (7) represented by grid representations at 0.25 Å spacing.

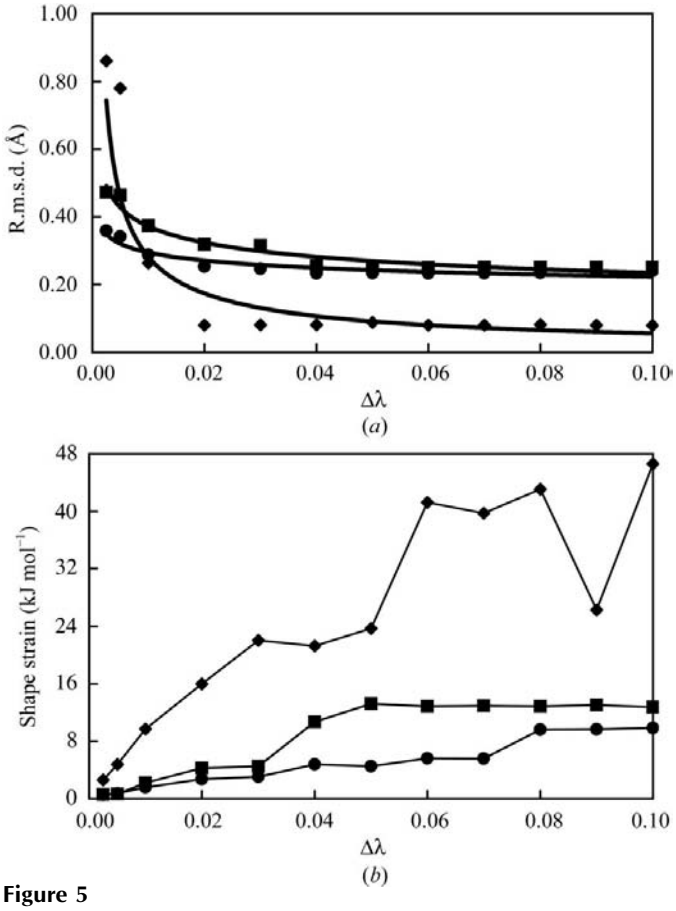

Figure 5
(*a*) R.m.s.d. between ligand structures obtained in the current procedure and their PDB structures and (*b*) shape-strain energies as functions of $\Delta\lambda$ used for ligand fitting. Protein ligands are retinoic acid/transport protein (diamonds), allosteric inhibitor in β-lactamase (squares) and quinazoline/p38 kinase (circles).

Table 1
Protein–ligand systems used for testing shape–force-field refinement.

PDB code	Resolution (Å)	*B*-factor range† (Å^2)	Protein	Ligand
1di9	2.60	24.0–32.3	P38 kinase	Quinazoline
1a28	1.80	19.2–31.8	Progesterone receptor	Progesterone
1xqd	1.80	14.0–21.6	P450NOR	Dinucleotide
1cbs	1.80	9.2–16.2	Transport protein	Retinoic acid
1ld8	1.80	15.3–26.2	Farnesyltransferase	IC49 inhibitor
1obd	1.40	15.8–27.1	Saicar synthase	ATP
1pzp	1.45	27.5–44.2	β-Lactamase	Allosteric inhibitor
1ajx	2.00	14.6–26.8	HIV protease	Cyclic urea
1err	2.60	30.5–50.9	Estrogen receptor	Raloxifene
1b0f	3.00	2.0–27.7	Neutrophil elastase	Peptide mimic‡
1ibw	3.20	35.8–49.3	HIS decarboxylase	HME§

† Ligand *B* factors. ‡ Peptidyl pentafluoroethyl ketone. § Histidine methyl ester.

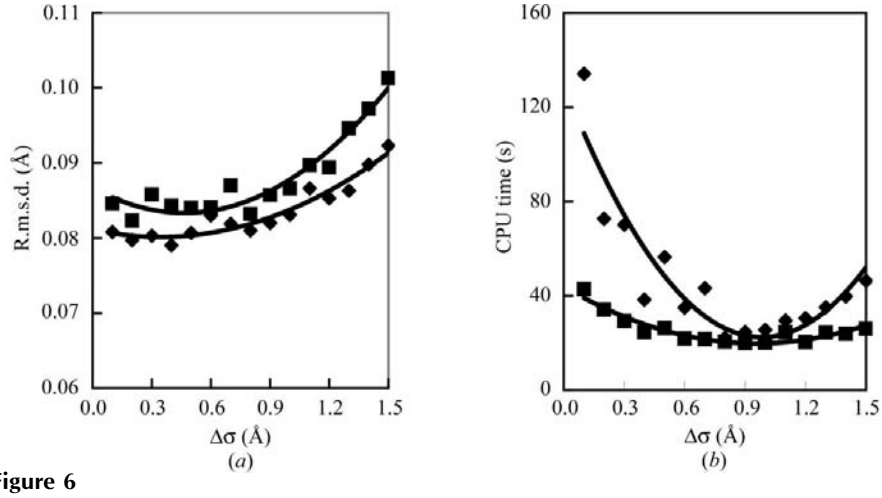

Figure 6
(*a*) R.m.s.d. between fitted structure of a single retinoic acid conformer and its PDB structure with the use of different set of convoluted electron-density grids calculated at 0.25 Å resolution (diamonds) and 0.5 Å resolution subsequently interpolated to 0.25 Å resolution (squares). The $\Delta\sigma$ values on the horizontal axis are increments between Gaussian functions used for convolution of electron density. (*b*) Corresponding CPU times.

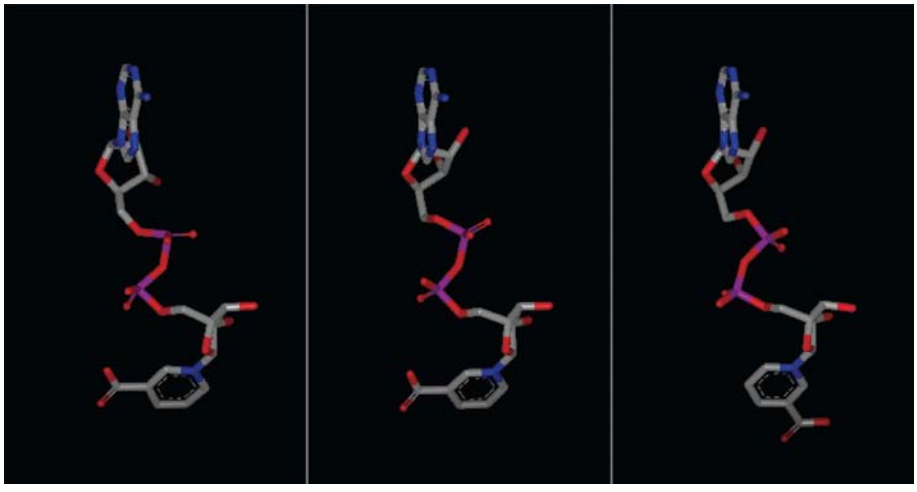

Figure 7
PDB (left) and adiabatically refined structures of the adenine dinucleotide molecule (nicotinic acid adenine dinucleotide) with complete MMFF94 potential (right) and with excluded Coulombic terms (middle). The image of the PDB structure also shows a planar coordination for the three O atoms in the PO_4 group (see text).

Gaussians per atom relative to the results with one Gaussian per atom. In both cases optimized structures differed from the model structures by an average of 0.07 and 0.08 Å r.m.s.d., respectively.

Next, we examined the fitting of ligands into simulated electron density generated on a grid from each model structure according to the protocol described in §2.5. The resultant series of scalar and gradient grids of 0.25 Å resolution was used with the same protocol of variation of λ as in the case of analytical gradients. Fig. 4(*c*) shows the shape strain *versus* model ligand energy for this series of fits. A significantly larger scatter of strain energies with respect to the same experiments with analytical gradients is clearly visible. This is likely to be a consequence of numerical error in gradient interpolation as the scatter reduces with resolution. However, on average all structures were aligned within 0.12 Å r.m.s.d., compared with 0.08 and 0.07 Å for analytical gradient alignments in the case of one and five Gaussians per atom, respectively. Given such a small differences, grid resolutions higher than the 0.25 Å used in the experiment currently seem unjustified.

3.2. Fitting to experimental electron density

3.2.1. Step size, number of Gaussian grids and grid interpolation. Practical usage of potential (1) requires selection of the size of $\Delta\lambda$, *i.e.* the adiabatic increment. If $\Delta\lambda$ is too large the fitting procedure may cross barriers without signature, leading to overstraining. On the other hand, steps that are too small are inefficient and, owing to finite precision in minimization, may lead to premature termination of the adiabatic procedure. Fig. 5 shows that for $\Delta\lambda \leq 0.02$ ligands are indeed underfitted, as is evident from the large r.m.s. deviations from the published structures.

In fact, in one of the examples shown in Fig. 5 (retinoic acid and its transport protein) the decrease of $\Delta\lambda$ from 0.02 to 0.005 results in a tenfold increase in r.m.s.d. from the PDB structure. $\Delta\lambda$ values around 0.05 appear to be a good compromise.

The overlap of a Gaussian function with an electron-density map is a convolution. Efficiency is gained by precalculating convolutions for a series of widths of Gaussians and then interpolating between members of this set

for any particular width. As the calculation of these convolutions is time-intensive and, in particular, memory-intensive, the determination of the number of such grids was of importance to the procedure. In addition, we considered the effect of convoluting on a course grid and interpolating to a finer grid. Fig. 6 illustrates results of this investigation, both in terms of speed and fitting accuracy. These results were used to optimize the efficiency of our ligand-refinement algorithm. The reduction of the number of Gaussians widths used for the calculation of Gaussians overlap with electron density also leads to less accurate forces in quasi-Newton optimization for every λ, leading to more iterations. In practice, a minimum is observed, as shown in Fig. 6(*b*).

3.2.2. Examples of refined ligands. A list of the protein–ligand systems used for testing our refinement method is shown in Table 1. Table 2 contains the shape-strain energies and r.m.s.d. from the corresponding PDB structures for the best shape matches of the ligands.

There are several observations to be made from the results in Tables 1 and 2. While the ligand temperature factors do not seem to be correlated with the differences between structures refined with our method and those from the PDB, poorer experimental resolution (>2.8 Å) can, but does not necessarily, lead to larger deviations (compare 1xqd, 1b0f and 1ibw). The worst resolution structure (1ibw) yields a solution only 0.45 Å from the PDB model and an exceptionally low strain energy (4.19 kJ mol^{-1}). The most striking observation, however, is the improvement of results on removing electrostatic terms from $V_{\rm FF}$. Without electrostatics, the strain on any ligand is at most 21 kJ mol^{-1}, in line with expectations from Charifson's work, and the maximum r.m.s.d. from a crystal structures is 0.82 Å observed for a poor experimental resolution system 1b0f), with an average of 0.31 Å. With electrostatics the maximum strain is almost double, the average r.m.s.d. is 0.56 Å, equal to the maximum r.m.s.d. without electrostatics, and there is a clear 'miss' with 1xqd. In this case the ligand, adenine dinucleotide, is highly charged. We illustrate what happens with and without electrostatics in Fig. 7.

With electrostatics (right), the pyridine ring is rotated to maximize the distance between carboxylic acid group and negatively charged O atoms of the phosphate group, while the X-ray structure (left) has a relatively short distance of 3.7 Å between those two groups, probably as a result of the interaction between Thr234 and Arg174 of the P450NOR protein. The structure from the adiabatic fit without electrostatics (middle) is much closer to the reported structure, although it is still one of our worst results at 0.54 Å. We looked more closely and noticed that one of the PO_4 groups in the X-ray structure is nearly planar rather than pyramidal and hence quite chemically unrealistic. As such, we believe our structure to be the more reasonable of the two. The differences observed with and without an electrostatic component are, of course, reasonable. We are not including protein

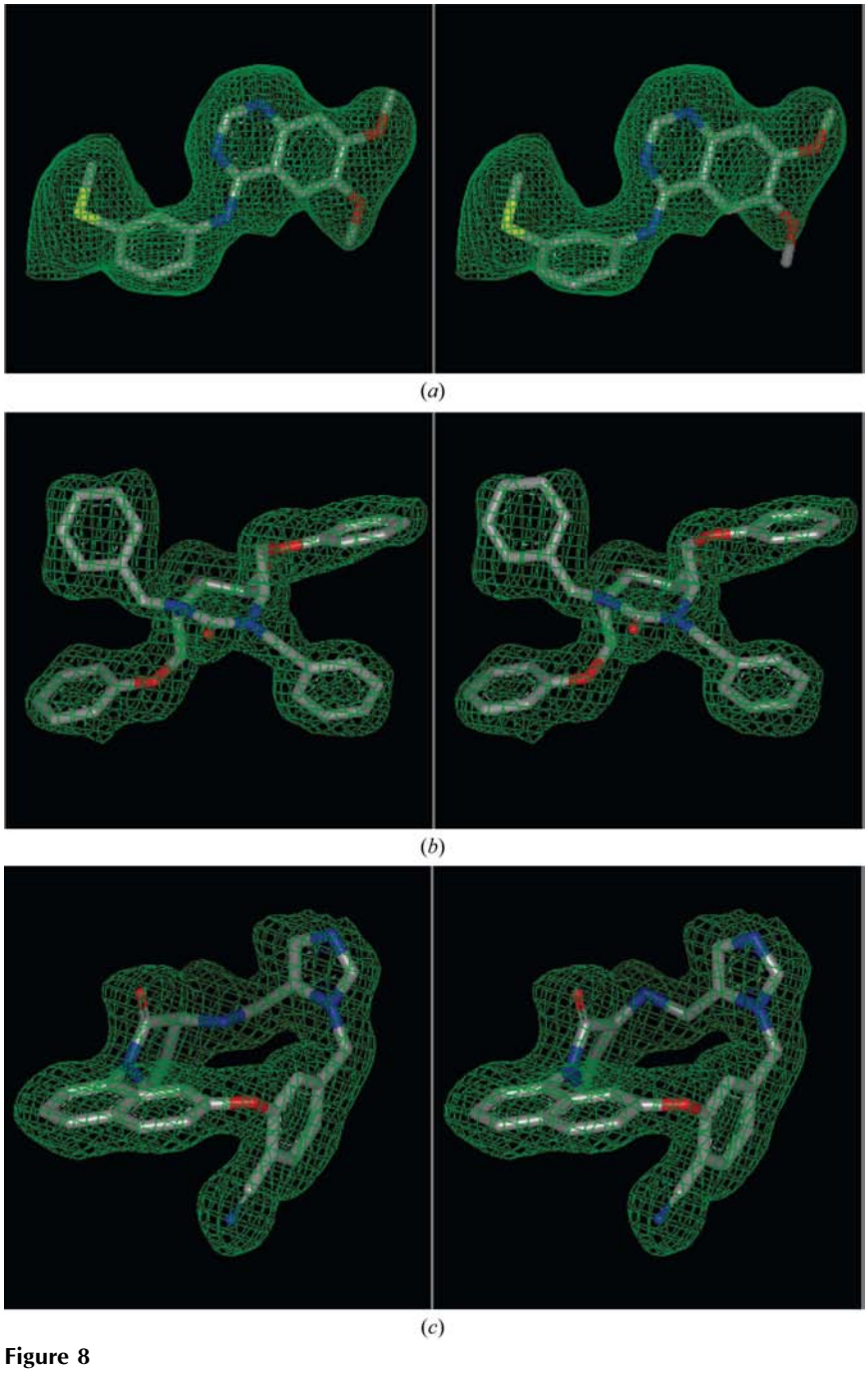

Figure 8
Fitted (left) and PDB structures (right). (*a*) 1di9, quinazoline in p38 kinase. (*b*) 1ajx, cyclic urea in HIV protease. (*c*) 1ld8, IC49 in farnesyltransferase.

Table 2
Shape-strain energies and r.m.s.d. deviations from published X-ray structures of several refined ligands with the use of potential (1).

Force-field–shape fits were performed for the five best rigid conformers; electron-density blob overlaps and the data for the best shape match are reported. $\Delta\lambda$ was 0.04 and the radii increment, $\Delta\sigma$, between Gaussian functions used for electron-density convolution was set at 0.4. Also reported are the number of conformers *Omega* generated initially and the smallest r.m.s.d. from any of these structures to the X-ray structure.

PDB code	No. of conformations	*Omega* r.m.s.d.	Strain† (kJ mol^{-1})	R.m.s.d.† (Å)	CPU time‡ (s)
1di9	59	0.52	3.68 (3.31)	0.24 (0.55)	24.2
1a28	4	0.12	0.63 (0.50)	0.06 (0.06)	43.4
1xqd	4922	1.33	22.4 (38.5)	0.54 (2.19)	69.5
1cbs	34	0.40	10.2 (9.79)	0.09 (0.09)	42.7
1ld8	42	0.56	20.7 (12.6)	0.28 (0.29)	41.7
1obd	3358	0.71	23.9 (39.8)	0.20 (0.75)	34.2
1pzp	163	0.47	6.91 (4.94)	0.27 (0.25)	22.6
1ajx	375	0.75	7.28 (−9.13)	0.16 (0.17)	73.0
1err	427	0.78	21.8 (30.9)	0.27 (0.54)	22.6
1b0f	811	0.88	19.5 (19.7)	0.82 (0.84)	35.4
1ibw	18	0.69	4.22 (7.58)	0.45 (0.48)	14.6
Average		0.65	12.8 (14.4)	0.31 (0.56)	38.5

† Values in parentheses were obtained with Coulomb terms included in the MMFF94 force field. ‡ CPU times were measured on an AMD Athlon-64 3400+ 1GB RAM machine.

information, which includes strong van der Waals and Coulombic forces. Although both these forces can contain attractive and repulsive effects, van der Waals attractive forces are weak compared with Coulombic attractive forces. As such, it should not be surprising that we sometimes fail to account for a reduction in ligand strain from complementary protein electrostatics. This is a drawback in our current approach and might be addressed by adding, for instance, a multipole representation of the protein electrostatic field in the vicinity of the ligand density. This is a current research direction. In structure 1ajx, the ligand is a cyclic urea and the effect of electrostatics is to actually lower the ligand energy from the starting conformation. This occurs owing to a non-adiabatic change to a different ring conformation that was not included in the original conformational sampling. Both the electrostatically strained and unstrained results are remarkably close to the X-ray structure (middle panel, Fig. 8).

Finally, Table 2 also illustrates the importance of a correct starting conformation from *Omega*; the closer the initial conformation, the more exact the eventual match. Programs that generate conformations typically filter ensembles such that structures too similar (in r.m.s.d.) from lower energy conformers are deleted. Although this is a powerful approach for most uses of conformational ensembles, it is a disadvantage in adiabatic fitting because the true conformation might be deleted by a conformation with a lower (unbound) energy. Ongoing work has shown that removing this r.m.s.d. culling improves results with little effect on efficiency.

Examples of fitted ligands are shown on Fig. 8. Small differences between the refined and PDB structures are visible in all three cases: ligands refined with the force-field–shape potential seem to occupy slightly more central positions in their electron-density blobs. This raises the issue as to whether the ligand structures generated by this procedure are potentially better then those generated by traditional methods. In some cases, such as the dinucleotide in 1xqd with the non-pyramidal PO_4 group, the adiabatic solution is clearly better; in others, where the chemistry is correct, this can only be assessed by re-refining the structure with the new ligand coordinates. A comprehensive study is currently under way, with early indications that adiabatic structures are equivalent to the best produced by other approaches, but with far fewer of the mistakes that plague the PDB.

4. Conclusions

We have demonstrated significant advantages to a real-space procedure that uses a Gaussian-based shape function and a modern small-molecule force field to adiabatically fit low-energy conformations to electron density. The protocol can generate high-quality, low-energy models automatically from a ligand-connection table. No protein information, save that in determining the ligand density, is required, although we have noted that the lack of protein electrostatics sometimes requires compensation by the removal of the Coulombic term from the force field. We have shown the procedure is relatively efficient and extensible. Using the optimal values for the parameter steps, grid spacing and convolution frequency reported here, model generation takes about a minute per ligand and can be applied to ligands of considerable flexibility without loss of utility (we have been successful with ligands containing 20 rotatable bonds). Current research directions include adding representations of protein electrostatics and analysis of a large number of previously reported structures by re-refinement.

This work was inspired by the premise that many crystal structures of drug-like ligand complexes show the ligand binding with very little strain energy (Perola & Charifson, 2004). However, there are many examples in the PDB with highly strained ligand conformations (*e.g.* 1nhu). We plan to explore rerefinement of these cases with the method described here to explore the possibility of proposing alternate models that fit the density but have low strain energies.

We would like to thank Tom Peat, Jon Christopher and all those who have contributed to the AFITT project. We would also like to thank Bob Tolbert and Matt Stahl for their insightful comments and advice. Finally, we would like to thank those crystallographers who have made their structure factors available through Uppsala University's Electron Density Server.

References

Agarwal, R. (1978). *Acta Cryst.* A**34**, 791–809.
Altomare, A., Giacovazzo, C., Ianigro, M., Moliterni, A. & Rizzi, R. (2002). *J. Appl. Cryst.* **35**, 21–27.
Cascarano, G., Giacovazzo, C., Camalli, M., Spagna, R. & Watkin, D. (1991). *Acta Cryst.* A**47**, 373–381.
Diller, D., Pohl, E., Redinbo, M., Hovey, B. & Hol, W. (1999). *Proteins*, **36**, 512–525.

Grant, J., Gallardo, M. & Pickup, B. (1996). *J. Comput. Chem.* **36**, 1653–1666.
Grant, J. & Pickup, B. (1995). *J. Phys. Chem.* **99**, 3503–3510.
Halgren, A. (1994*a*). *J. Comput. Chem.* **17**, 490–519.
Halgren, A. (1994*b*). *J. Comput. Chem.* **17**, 520–552.
Halgren, A. (1994*c*). *J. Comput. Chem.* **17**, 553–586.
Halgren, A. (1994*d*). *J. Comput. Chem.* **17**, 616–641.
Halgren, A. & Nachbar, B. (1996). *J. Comput. Chem.* **17**, 587–615.
Kleywegt, G., Harris, M., Zou, J., Taylor, T., Wahlby, A. & Jones, T. (2004). *Acta Cryst.* D**60**, 2240–2249.
Koch, M. H. (1974). *Acta Cryst.* A**30**, 67–70.
Levitt, D. G. (2001). *Acta Cryst.* D**57**, 1013–1019.
Main, P. & Hull, S. (1978). *Acta Cryst.* A**34**, 353–361.
Menéndez-Velázquez, A. & García-Granda, S. (2003). *J. Appl. Cryst.* **36**, 193–205.
Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* D**53**, 240–255.
Oldfield, T. (2001). *Acta Cryst.* D**57**, 696–705.
Perola, E. & Charifson, P. (2004). *J. Med. Chem.* **47**, 2499–2510.
Perrakis, A., Morris, R. & Lamzin, V. (1999). *Nature Struct. Biol.* **6**, 458–463.
Sneath, P. & Sokal, R. (1973). *Numerical Taxonomy*. New York: W. H. Freeman & Co.
Terwilliger, T. (2003). *Acta Cryst.* D**59**, 38–44.
Willet, P., Barnard, J. & Downs, G. (1998). *J. Chem. Inf. Comput. Sci.* **38**, 983–996.
Zwart, P., Langer, G. & Lamzin, V. (2004). *Acta Cryst.* D**60**, 2230–2239.