

Ligand Electron Density Shape Recognition Using 3D Zernike Descriptors

Prasad Gunasekaran^{1,*}, Scott Grandison^{1,*}, Kevin Cowtan², Lora Mak³,
David M. Lawson⁴, and Richard J. Morris¹

¹ Department of Computational & Systems Biology, John Innes Centre,
Norwich Research Park, Colney Lane, NR4 7UH Norwich, UK

² Structural Biology Laboratory, Department of Chemistry, University of York,
Heslington, York, UK YO10 5DD

³ Theoretical Systems Biology, Institute of Food Research, Norwich Research Park,
Colney Lane, NR4 7UA Norwich, UK

⁴ Department of Biological Chemistry, John Innes Centre, Norwich Research Park,
Colney Lane, NR4 7UH Norwich, UK

prasad.gunasekaran@bbsrc.ac.uk, scott.grandison@bbsrc.ac.uk,
cowtan@ysbl.york.ac.uk, lora.mak@bbsrc.ac.uk,
david.lawson@bbsrc.ac.uk, richard.morris@bbsrc.ac.uk

Abstract. We present a novel approach to crystallographic ligand density interpretation based on Zernike shape descriptors. Electron density for a bound ligand is expanded in an orthogonal polynomial series (3D Zernike polynomials) and the coefficients from this expansion are employed to construct rotation-invariant descriptors. These descriptors can be compared highly efficiently against large databases of descriptors computed from other molecules. In this manuscript we describe this process and show initial results from an electron density interpretation study on a dataset containing over a hundred OMIT maps. We could identify the correct ligand as the first hit in about 30 % of the cases, within the top five in a further 30 % of the cases, and giving rise to an 80 % probability of getting the correct ligand within the top ten matches. In all but a few examples, the top hit was highly similar to the correct ligand in both shape and chemistry. Further extensions and intrinsic limitations of the method are discussed.

Keywords: pattern recognition, structural bioinformatics, electron density, protein crystallography, 3D Zernike moments.

1 Introduction

With the success of structural genomics worldwide and the recognised importance of 3D information in unravelling protein function and enzymatic mechanism, computational methods are increasingly serving a vital role in structural biology. From target selection (using sequence analysis, phylogeny, fold recognition and homology modeling), registration, laboratory information management

* Equal contributing first author.

systems, crystal optimisation, data collection and analysis, the solution of the phase problem (often using molecular replacement techniques with databases of fragments or homology models), to complete and validated models and the prediction of protein function, computational biology has a major role to play.

Automated model building into crystallographic electron density has come a long way in the last decade. Although many highly sophisticated algorithms and graphical approaches existed to aid the model building process, prior to the release of packages such as ARP/wARP [1], RESOLVE [2], PHENIX [3], and BUCCANEER [4], the construction of a macromolecular model into electron density often required many months of expert crystallographer time. Electron density interpretation has always been driven by pattern recognition [5]. Initially this process was carried out by the human brain based on structural insights and experience, but alternative approaches based on algorithmic pattern recognition developments such as skeletonization [6] quickly reduced the demand on human effort. Automated methods [7,4,8,9,10,11,12,13,14] contributed greatly to and benefitted hugely from structural genomics initiatives and recent developments [15,16,17,18,4] have reached an impressive level of sophistication, providing an intelligent expert system approach to model building for protein structures.

Methods for determining structures from small molecule crystallographic data are completely automated and very well advanced. The success of these techniques relies strongly on direct methods, Patterson approaches and sufficient data to warrant atomicity [19,20], thus allowing identified atoms to be connected and the molecule to be built in a rather straightforward manner. These methods break down with decreasing resolution. Whereas proteins have regular repeating elements that can be distinguished already at lower resolution, ligands exhibit less pronounced and regular patterns, making their interpretation problematic. Thus, automated building of protein ligands into macromolecular structures is by comparison less well developed than for proteins, and DNA/RNA molecules seem to have either been ignored or successfully resisted automation. However, a number of original, diverse, and interesting approaches have been developed for the identification and modeling of bound ligands.

Zwart *et al.* [21] describe methods for automated identification and building of protein-bound ligands in electron-density maps. Their approach relies on a graph-based search for geometrical features which match stereochemical expectations and an atomic labeling algorithm which explores the combinatorics of the problem. This approach has recently been assessed [22] and has shown to be most successful for high resolution data.

Based on his powerful methods for density modification using structural motifs [23], Terwilliger successfully enhanced and extended the approach to ligand fitting [24]. This technique relies on the placement of core fragments into electron density and then uses a novel procedure to extend and build the remainder of the ligand. The authors performed extensive testing over a large dataset and found their methods functioned well, placing the ligand within 2 Å r.m.s.d of the coordinates in the original structure. Recently, a new development [25] has been described which allows for the identification of ligands bound in crystal structures of

macromolecules. This method uses density and fingerprint correlations. The density correlation is computed from the density after optimization of each entry of a test set of ligands and the ligand density. The fingerprint correlations are lists of computed model density features between the ligands.

An interesting development geared towards structure-based drug design is presented in [26]. Their approach aims to automate protein-ligand crystallography for drug design. Ligands are positioned by docking directly into electron density, whilst taking care of protein-ligand interactions. In addition to the identification of full ligands, the authors show the potential for analysing fragment-library screening experiments. A force-field approach can be included to provide a better fit whilst maintaining good geometry [27]. An alternative approach captures the central axis of the electron density isosurface with a graph. This graph is matched against a graph of the molecular model [28]. This approach was reported to work as well as density peak picking methods and showed promise for the extension to lower resolution [28]. In a novel development using rotation-invariant features derived from moments of various orders, Lamzin and co-workers [29] developed a powerful method for accurately locating planar objects in electron density. This approach was tested on protein and DNA/RNA crystal structures and was able to place the plane centres within 0.5 \AA of the correct position. Although not specifically designed for ligand density interpretation, this approach seems a promising strategy and has much in common with our methodology.

In this manuscript we present a method for ligand identification based on global pattern recognition using 3D Zernike descriptors. Zernike polynomial techniques have recently been applied with success in the biomolecular sciences [30,31,32]. The advantages of this approach include accurate feature representation and speed of comparison, meaning that large libraries can be scanned efficiently. Although for the current application the top hits already identified the correct ligand in many cases, a more promising set up might be to combine this technique with other methods such as those mentioned above to build a hierarchical search system which filters quickly through large databases and progressively invests more time only in relatively few promising candidates.

2 Methodology

An overview of the ligand density identification method described in this manuscript is shown in Figure 1. The individual steps are explained in detail below. To evaluate the success of this approach, we chose datasets of experimental crystallographic density which we attempted to interpret. The success was judged by how many times we predicted the correct ligand and, in cases where we got the wrong ligand, how distant our prediction was from the correct solution.

2.1 Density Extraction

Clipper library routines [33] were used to 1) compute crystallographic OMIT maps (OMIT maps show the difference between the density computed from the

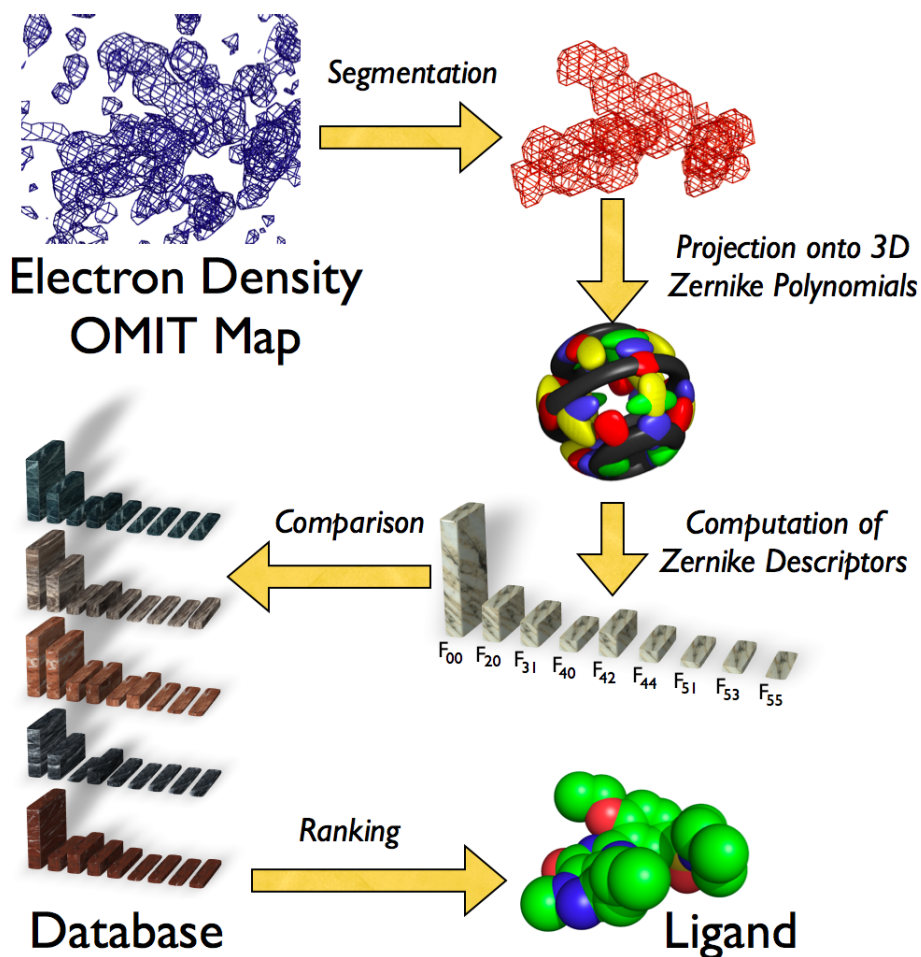


Fig. 1. Flowchart of the 3D Zernike descriptor method for ligand density interpretation. Crystallographic difference density was computed using Clipper routines and passed through our region growing algorithm. The resulting density is projected into Zernike space (the individual colours in the multi-colour sphere representation correspond to different Zernike polynomials). The Zernike descriptors are scanned against a database of pre-computed descriptors from existing ligands and the best hit is returned as the most probable ligand.

observed crystallographic structure factors and the theoretical density from a given arrangement of atoms and can be used to spot those parts of the density that have not been modeled or contain errors) by leaving out all the ligands from the model; 2) compute a crystallographic R-factor to check the density was reasonable; 3) to extract an orthogonal grid of difference density around each ligand. Such electron density maps are shown in the left column in Figures 2 and 3. Using OMIT maps based on models that have 'seen' the ligand during

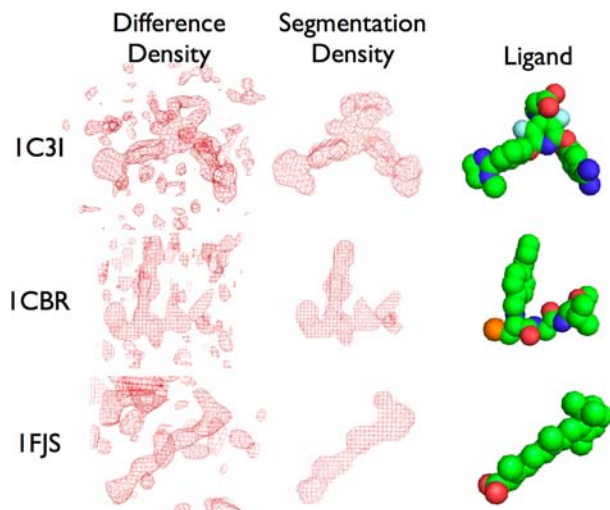


Fig. 2. Density, segmentation and ligands. A selection of density maps for which the 3D Zernike descriptor ligand identification gave the correct answer.

refinement introduces a bias towards improved ligand density and yet these maps are by nature often noisy and contain errors from missing data and/or poorly modelled parts of the structure. To address this, we implemented a region-growing algorithm following the approach described in [34]. This method allows for the automated segmentation of 3D images. The algorithm starts from a seed point and progresses to visit neighbouring pixels, marks them, and either adds them to the current region list, or not, depending on the defined growth criteria. As the seed point we chose the highest point in the OMIT map and as growth criteria we defined that at least two neighbouring points to be above a given threshold. The choice of threshold was map dependent and our selected values ranged from 1.2 to 2.0 σ (standard deviation of the electron density values). The threshold was varied to produce the most ligand-like shape as judged by the human eye. The average threshold value was about 1.4 σ , so this value was employed for automated density extraction. Images of automatically extracted difference density are shown in the second column of Figure 2 and 3.

2.2 Zernike Moments

The success of 3D Zernike moments for object recognition in computer science has been documented in the number of publications, outperforming many other well-established techniques in the area. Following developments by [35] and [36] for 3D shape retrieval, we have employed 3D Zernike polynomials as a basis set for molecular shape comparison [30]. This approach was initially used to describe and compare binary objects, but the underlying mathematics are sufficiently general to allow any function in 3D to be described within the same framework.

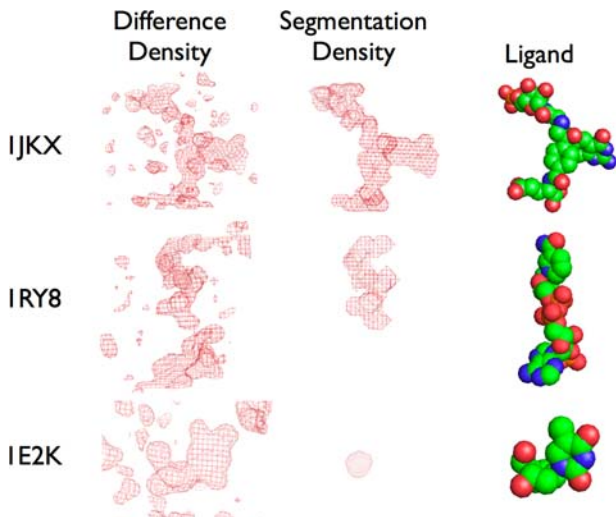


Fig. 3. Crystallographic difference densities, extracted ligand densities and the correct ligands. A selection of density maps for which the correct answer was not obtained. For 1JKX another conformation was ranked top, for 1RY8 the density segmentation picked up only half the ligand density because of a density break, and for 1E2K the region growing algorithm missed the actual ligand density.

As demonstrated in [31] and [37], the Zernike descriptor approach is well-suited for molecular comparisons and can be adapted to a range of physicochemical properties. We recently demonstrated the application of 3D Zernike moments to capture model uncertainty and protein flexibility [32]. A detailed description of the method for matching shapes using Zernike moments has been presented in [35,36,30]. We summarise some of the main points below.

Any square-integrable function on the unit ball, f , can be represented as

$$f(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{l=0}^n \sum_{m=-l}^l c_{nlm} Z_{nlm}(\mathbf{r}), \quad (1)$$

in which $Z_{nlm}(\mathbf{r})$ are the 3D Zernike polynomials given here as functions of the spherical coordinates, $\mathbf{r} = (r, \theta, \phi)$. The 3D Zernike polynomials are basis functions on the unit ball consisting of a radial term, $R_{nl}(r)$, and an angular term, $Y_{lm}(\theta, \phi)$. The angular functions, $Y_{lm}(\mathbf{r})$, are known as spherical harmonics and can themselves be used for molecular shape recognition purposes [38,39]. The function indices are denoted by nlm , where n ranges from 0 to the maximum expansion order, $l \leq n$ and $-l \leq m \leq l$. The radial part of the Zernike polynomials is given by

$$R_{nl}(r) = \sum_{k=0}^{(n-l)/2} N_{nlk} r^{n-2k} \quad (2)$$

for $n - l$ even, otherwise zero. N_{nlk} is a normalisation constant. See [40,41] for further details and cartesian transformations. The expansion coefficients, c_{nlm} , are called 3D Zernike moments. The determination of the 3D Zernike moments requires the integration of the function of interest, $f(\mathbf{r})$, multiplied by the complex conjugate Zernike polynomials, $Z_{nlm}^*(\mathbf{r})$, over the unit ball,

$$c_{nlm} = \int_0^1 \int_0^{2\pi} \int_0^\pi Z_{nlm}^*(\mathbf{r}) f(\mathbf{r}) r^2 \sin \theta \, d\theta \, d\phi \, dr. \quad (3)$$

The symmetry properties of the spherical harmonics induce a symmetry in the coefficients between positive and negative m indices, $c_{nl,-m} = (-1)^m c_{nlm}^*$. The symbol * denotes the complex conjugate.

Rotationally invariant descriptors may be constructed from the moments by defining $(2l + 1)$ -dimensional vectors for each index l with all the m indexed coefficients, c_{nlm} , making up each subspace,

$$F_{nl} = \begin{pmatrix} c_{nl,-l} \\ c_{nl,-l+1} \\ \vdots \\ c_{nl,l} \end{pmatrix}. \quad (4)$$

These rotationally invariant coefficients, F_{nl} , provide an efficient way of comparing shapes, a and b , using a standard L_2 metric,

$$d = \sqrt{\sum_{n=0}^{n_{\max}} \sum_{l=0}^n (F_{nl}^a - F_{nl}^b)^2}, \quad (5)$$

for a maximum expansion order equal to n_{\max} .

Thus, in terms of ligand identification the Zernike descriptors can be employed to find promising molecules of similar shape to the crystallographic density without the need for translational or rotational searches and optimisation. Once potential candidates have been identified a more sensitive method could be used, such as [8], and coupled with refinement.

3 Datasets

3.1 Electron Density Maps

We downloaded a selection of PDB structures with their crystallographic structure factors. In total we selected 110 structures in a resolution range of 1.5 Å to 2.8 Å, which gave rise to 586 ligand density maps. Only maps corresponding to ligands with more than 10 non-hydrogen atoms were selected, leaving us with 164 maps in total. The size of these ligands varies from 10 to 111 atoms, with the average being 22.7 non-hydrogen atoms.

3.2 Ligand Database

We downloaded the latest version, release 12.1, of the HIC-Up database [42]. This version contains 7870 small molecules extracted from the PDB. The 3D Zernike moments were computed following the procedure outlined in Grandison *et al.* [32]. In short, each molecule was projected onto the unit ball such that the largest distance from the centre of geometry was scaled to 0.6. The atoms were projected onto a 64^3 grid using a Gaussian function with a variance defined by the van der Waals radius. For each voxel the geometric moments were computed from which the Zernike moments were determined. The Zernike moments were used to compute rotationally-invariant descriptors, following Equation 4.

4 Results

For each density map we applied the noise reduction algorithm described in the methods section to choose a continuous high density ($> 1.5\sigma$) region. This density was expanded in 3D Zernike polynomials to an order of 30 and the rotationally invariant shape descriptors were computed. These shape descriptors were scanned against a database consisting of 7870 pre-computed small molecule shape descriptors and the 164 ligand shape descriptors from the PDB files used for the density maps. Surprisingly, the density shape was not able to recognise its own ligand in a large number of cases. There are a number of reasons for this. First, although the region-growing algorithm performed well overall, it failed to extract the correct region in almost 30 % of the cases, see Figure 3, due to poor density. There are numerous ways to address this problem but for the

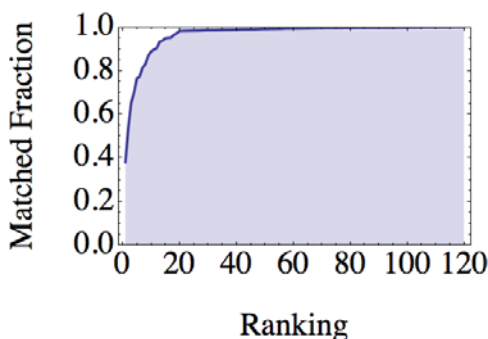


Fig. 4. This plot shows the frequency of correctly identified ligands (y -axis) as a function of the number of selected molecules (x -axis) over the entire data set. A perfect method would give 1.0 already for the number of molecules equal to one (this is the case for our approach using computed ligand densities from the ligand model). Poor density and insufficiently powerful segmentation and image enhancement algorithms limit the performance in our test cases. For real applications the conformational flexibility of ligands will reduce the performance further.

present analysis we simply excluded those cases in order to evaluate the shape matching procedure rather than the region growing algorithm. Second, in cases where the region-growing procedure performed well, the density was still poor in places and resulted in parts being missed (and other ligands matching better) or moved (resulting in correct ligands in different conformations being the best hit). Another surprising finding is that a higher expansion order often decreases the performance. In [38] we found that for shape matching purposes using spherical harmonics a maximum expansion order of about 6 was sufficient and the performance increase for higher order was negligible. Similarly for Zernike descriptors [30], we found an order of about 10 to be sufficient for molecule classification and that the performance increased only marginally overall with higher orders, dropping slightly in only a few cases for protein comparisons. Here, the higher orders are capturing more noise in the density, whereas the lower orders are making good approximations to the overall shape and thereby performing noise reduction. We found the ligand interpretation performance peaked at a maximum expansion order of 8.

5 Discussion

We have described a novel method for the interpretation of electron density which can be used for the identification of ligands in crystallographic electronic difference density. We have performed an initial analysis on a realistic test dataset which covers a range of ligands from experimental crystallographic data of varying resolution and map quality. We have developed methods to automatically extract the difference density maps, place them on an orthogonal grid, expand them in 3D Zernike polynomials, and to efficiently compare their rotationally invariant moment against other shape databases. Overall, the performance was fair but not outstanding. We could identify the correct ligand in 30 % cases as the top hit and in a further 30 % cases within the top five matches.

Despite encouraging results, several points became apparent during this analysis that require further investigation before the method can be properly automated. The first point is the density extraction itself. Advanced image processing tools, more specifically better techniques for image segmentation such a sophisticated watershed transform, may provide improve this step. Another option would be to allow the user to manually select and modify the density (via the placement of atoms) in existing model building graphics software. However, even with the perfect ligand density this method has limitations in that it performs static global shape matching, although multi-conformer searching presents a tractable solution given the speed of comparison. With the current test set, the conformational diversity was limited and the full effect of this limitation is therefore reduced. This conformational sampling issue could represent a problem, although initial analyses suggest that the number of bound conformations of flexible ligands is rather small [43]. We stress the point that the use of OMIT maps based on models that have 'seen' the ligand during refinement implies that the current results will be better than in a more realistic scenario with poorer ligand density.

Overall, the performance in selecting the correct ligand as the top hit did not meet our expectations. For this dataset the correct match was usually within the top twenty and with 80 % frequency within the top ten hits. In 30 % of the cases, we could identify the correct ligand with the first hit. However, this approach does act as an impressive filter and delivers a small set of potential candidates. More elaborate methods could then be employed to select the most likely ligand from this set. We therefore suggest that this approach would be best-suited as an efficient pre-screening method, prior to launching more computationally costly techniques.

Acknowledgments

The idea to this piece of work is based on discussions with Dr Victor S Lamzin (EMBL Hamburg) about using spherical harmonics to interpret ligand density shapes. We thank Abdullah Kahraman for discussion. This project was funded in part by a BBSRC Tools and Resources grant - grant number CA340H10B.

References

1. Perrakis, A., Morris, R., Lamzin, V.S.: Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6(5), 458–463 (1999)
2. Terwilliger, T.C.: Solve and resolve: automated structure solution and density modification. *Methods Enzymol.* 374, 22–37 (2003)
3. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., Terwilliger, T.C.: Phenix: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58(Pt 11), 1948–1954 (2002)
4. Cowtan, K.: The buccaneer software for automated model building. 1. tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* 62(Pt 9), 1002–1011 (2006)
5. Morris, R.J.: Statistical pattern recognition for macromolecular crystallographers. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2133–2143 (2004)
6. Greer, J.: Computer skeletonization and automatic electron density map analysis. *Methods Enzymol.* 115, 206–224 (1985)
7. Cowtan, K.: Fast fourier feature recognition. *Acta Crystallogr. D Biol. Crystallogr.* 57(Pt 10), 1435–1444 (2001)
8. Cowtan, K.: Fitting molecular fragments into electron density. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 1), 83–89 (2008)
9. Morris, R.J., Perrakis, A., Lamzin, V.S.: ARP/wARP model-building algorithms. i. the main chain. *Acta Crystallogr. D Biol. Crystallogr.* 58(Pt 6 Pt 2), 968–975 (2002)
10. Morris, R.J., Perrakis, A., Lamzin, V.S.: ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* 374, 229–244 (2003)
11. Morris, R.J., Zwart, P.H., Cohen, S., Fernandez, F.J., Kakaris, M., Kirillova, O., Vornrhein, C., Perrakis, A., Lamzin, V.S.: Breaking good resolutions with ARP/wARP. *J. Synchrotron. Radiat.* 11(Pt 1), 56–59 (2004)

12. Adams, P.D., Gopal, K., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., Sacchettini, J.C., Sauter, N.K., Storoni, L.C., Terwilliger, T.C.: Recent developments in the phenix software for automated crystallographic structure determination. *J. Synchrotron. Radiat.* 11(Pt 1), 53–55 (2004)
13. Terwilliger, T.C.: Automated structure solution, density modification and model building. *Acta Crystallogr. D Biol. Crystallogr.* 58(Pt 11), 1937–1940 (2002)
14. Terwilliger, T.: Solve and resolve: automated structure solution, density modification and model building. *J. Synchrotron. Radiat.* 11(Pt 1), 49–52 (2004)
15. Cohen, S.X., Morris, R.J., Fernandez, F.J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V.S., Kleywegt, G.J., Perrakis, A.: Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2222–2229 (2004)
16. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J., Adams, P.D.: Iterative model building, structure refinement and density modification with the phenix autobuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 1), 61–69 (2008)
17. Joosten, K., Cohen, S.X., Emsley, P., Mooij, W., Lamzin, V.S., Perrakis, A.: A knowledge-driven approach for crystallographic protein model completion. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 4), 416–424 (2008)
18. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Adams, P.D., Read, R.J., Zwart, P.H., Hung, L.W.: Iterative-build omit maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. D Biol. Crystallogr.* 64(Pt 5), 515–524 (2008)
19. Morris, R.J., Bricogne, G.: Sheldrick's 1.2 Å rule and beyond. *Acta Crystallogr. D Biol. Crystallogr.* 59(Pt 3), 615–617 (2003)
20. Morris, R.J., Blanc, E., Bricogne, G.: On the interpretation and use of $\langle |E|^2 \rangle$ (d^*) profiles. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 2), 227–240 (2004)
21. Zwart, P.H., Langer, G.G., Lamzin, V.S.: Modelling bound ligands in protein crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2230–2239 (2004)
22. Evrard, G.X., Langer, G.G., Perrakis, A., Lamzin, V.S.: Assessment of automatic ligand building in arp/warp. *Acta Crystallogr. D Biol. Crystallogr.* 63(Pt 1), 108–117 (2007)
23. Terwilliger, T.C.: Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta Crystallogr. D Biol. Crystallogr.* 57(Pt 12), 1755–1762 (2001)
24. Terwilliger, T.C., Klei, H., Adams, P.D., Moriarty, N.W., Cohn, J.D.: Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallogr. D Biol. Crystallogr.* 62(Pt 8), 915–922 (2006)
25. Terwilliger, T.C., Adams, P.D., Moriarty, N.W., Cohn, J.D.: Ligand identification using electron-density map correlations. *Acta Crystallogr. D Biol. Crystallogr.* 63(Pt 1), 101–107 (2007)
26. Mooij, W.T.M., Hartshorn, M.J., Tickle, I.J., Sharff, A.J., Verdonk, M.L., Jhoti, H.: Automated protein-ligand crystallography for structure-based drug design. *Chem. Med. Chem.* 1(8), 827–838 (2006)
27. Wlodek, S., Skillman, A.G., Nicholls, A.: Automated ligand placement and refinement with a combined force field and shape potential. *Acta Crystallogr. D Biol. Crystallogr.* 62(Pt 7), 741–749 (2006)

28. Aishima, J., Russel, D.S., Guibas, L.J., Adams, P.D., Brunger, A.T.: Automated crystallographic ligand building using the medial axis transform of an electron-density isosurface. *Acta Crystallogr. D Biol. Crystallogr.* 61(Pt 10), 1354–1363 (2005)
29. Hattne, J., Lamzin, V.S.: Pattern-recognition-based detection of planar objects in three-dimensional electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* D64(Pt 8), 834–842 (2008)
30. Mak, L., Grandison, S., Morris, R.J.: An extension of spherical harmonics to region-based rotationally invariant descriptors for molecular shape description and comparison. *J. Mol. Graph. Model* 26(7), 1035–1045 (2008)
31. Sael, L., Li, B., La, D., Fang, Y., Ramani, K., Rustamov, R., Kihara, D.: Fast protein tertiary structure retrieval based on global surface shape similarity. *Proteins* (2008)
32. Grandison, S., Roberts, C., Morris, R.J.: The application of 3d zernike moments for the description of “model-free” molecular structure, functional motion, and structural reliability. *J. Comput. Biol.* 16(3), 487–500 (2009)
33. Cowtan, K.: The clipper C++ libraries for x-ray crystallography. *IUCr Computing Commission Newsletter* 2, 4–9 (2003)
34. Revol-Muller, C., Peyrin, F., Carrillon, Y., Odet, C.: Automated 3d region growing algorithm based on an assessment function. *Pattern Recogn. Lett.* 23(1-3), 137–150 (2002)
35. Canterakis, N.: 3D zernike moments and zernike affine invariants for 3D image analysis and recognition. In: *Scandinavian Conference on Image Analysis* (1999)
36. Novotni, M., Klein, R.: Shape retrieval using 3d zernike descriptors. *Computer-Aided Design* 36(11), 1047–1062 (2004)
37. Sael, L., La, D., Li, B., Rustamov, R., Kihara, D.: Rapid comparison of properties on protein surface. *Proteins* 73(1), 1–10 (2008)
38. Morris, R.J., Najmanovich, R.J., Kahraman, A., Thornton, J.M.: Real spherical harmonic expansion coefficients as 3d shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics* 21(10), 2347–2355 (2005)
39. Morris, R.J.: An evaluation of spherical designs for molecular-like surfaces. *J. Mol. Graph. Model* 24(5), 356–361 (2006)
40. Mathar, R.J.: Third order newton’s method for zernike polynomial zeros, arXiv: math.NA/0705.1329 (2007)
41. Mathar, R.J.: Zernike basis to cartesian transformations, arXiv: 0809.2368 math-ph (2008)
42. Kleywegt, G.J.: Crystallographic refinement of ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 63(Pt 1), 94–100 (2007)
43. Funkhouser, T., Glaser, F., Laskowski, R.A., Morris, R.J., Najmanovich, R., Stockwell, G., Thornton, J.M.: Shape-based classification of bound ligands. In: Barber, S., Baxter, P.D., Mardia, K.V., Wells, R.E. (eds.) *Quantitative Biology Shape Analysis and Wavelets*, pp. 39–42 (2005)