Published in final edited form as:

J Struct Biol. 2007 May ; 158(2): 188–195.

Density-based score for selecting near-native atomic models of unknown structures

Eran Shacham, Brian Sheehan, and Niels Volkmann*

The Burnham Institute for Medical Research, Bioinformatics and Systems Biology Program

Abstract

We present a low-resolution density-based scoring scheme for selecting high quality models from a large pool of lesser quality models. We use pre-configured decoy data sets that contain large numbers of models with different degrees of correctness to evaluate the performance of the strategy. We find that the scoring scheme consistently identifies one of the highest quality models for a wide variety of target structures, resolution ranges, and noise models. Tests with experimental data yield similar results.

Keywords

Electron microscopy; structure prediction; small angle X-ray scattering; protein decoys; threading; docking; fitting; three-dimensional reconstruction; scoring

Introduction

High-resolution structures of macromoles are needed for studying their assembly, function and regulation. Despite the progress in crystallography and NMR for determining these structures and their components, only a small number compared to the thousands of different proteins in a typical cell are available. One possible way for generating high-resolution information on a structure is the combination of homology modeling and density-based docking into intermediate-resolution maps from electron microscopy (Topf and Sali, 2005). Consequently, this combination is becoming increasingly common (see for example Baker et al., 2002; Fotin et al., 2004; Gao et al., 2003; Liu et al., 2004; Sengupta et al., 2004; Topf et al., 2005; Topf et al., 2006; Volkmann et al., 2001). A recent study indicated that fitting a homology model based on a remotely related template is generally better than fitting the template itself and that the most accurate models can often be identified by the density docking score, even at 15 Å resolution (Topf et al., 2005). Here, we show that this concept can be extended for selecting models from arbitrary modeling sources and that, in many cases, density information at 20 Å resolution is sufficient to select high-quality structures from a set of alternative models with lower quality.

To evaluate performance, we used structures from the Decoys 'R' Us database (Samudrala and Levitt, 2000). Decoys are artificial conformations of protein sequences that possess some

*Correspondence should be addressed to Niels Volkmann, The Burnham Institute for Medical Research, Bioinformatics and Systems Biology Program, 10901 North Torrey Pines Road, La Jolla, CA 92037, Phone: 858 646 3187, Fax: 858 646 3195, e-mail: niels@burnham.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

characteristics of native proteins but are not actually correct. The database contains over 120 crystal structures where a range of conformations with different root-mean-square deviations (RMSD) were generated using various structure prediction algorithms including homology modeling and ab-initio blind predictions. The database is specifically designed to provide a representative and comprehensive set of decoys for the evaluation of new scoring algorithms. In this context, multiple decoy sets are essential for testing the ability of a scoring function to succeed in many different settings. If only one type of decoys is used for evaluation, discrimination may be achieved by exploiting some specific artifact of the respective decoys, such as lack of compactness or systematic distortions (Samudrala and Levitt, 2000). Using a pre-configured database ensures that a wide range of well tested targets are used for score evaluation.

The lower size limit for structure determination at 1-2nm resolution by electron microscopy (EM) is currently at ~200 kDa. The density for smaller proteins or domains can only be obtained as part of larger complexes and needs to be computationally extracted from the density of the larger entity. Possibilities for doing that include difference mapping using two EM reconstructions with one being a substructure of the other (see for example Hanein et al., 1998), discrepancy mapping using an EM reconstruction and a docked atomic model of a substructure (Volkman et al., 2000), or segmentation of the EM reconstruction into self-consistent density segments using only the density information from the EM data (Volkman, 2002). All of these methods may introduce distortions in the extracted density of the protein or domain in question and may hamper the applicability of our methodology to this type of data. In order to validate the applicability of our methodology in such a scenario we employed the structure of human rhinovirus complexed with Fab fragments. This structure was solved by EM to ~28 Å resolution (Smith et al., 1993). Later, the same structure was also solved by crystallography (Smith et al., 1996), allowing atomic level comparison of candidate models with the structure imaged by EM. Our analysis using the experimental density of the Fab fragment, extracted from the rhinovirus-Fab complex EM reconstruction by a variety of techniques, verifies that our methodology can indeed be useful for model selection in a real-life scenario.

Methods

Synthetic data

To emulate the existence of low-resolution density information, we calculated density maps of all target crystal structures at resolutions of 8, 10, 15 and 20 Å. In order to investigate the influence of random noise on the scoring performance, we also generated two additional maps for each of the calculated maps by perturbing them with either Gaussian or Laplacian impulse random noise at 0.5 signal-to-noise ratio. Thus, for each target structure, a total of 12 density maps were used for evaluation. For each of the decoy structures, we evaluated density scores with each of the corresponding densities using the CoAn system (Volkman and Hanein, 1999; Volkman and Hanein, 2003). CoAn uses the real-space correlation coefficient between a target map and a search model as a scoring function for determining the best orientation of the search model in respect to the target map. This is then followed by a statistical analysis of search space to determine confidence intervals for the positioning and derivation of other parameters such as the probability of residues to participate in interactions.

Here, CoAn is used to provide a single score for each model to allow model selection. Briefly, the center of mass of the respective decoy was superimposed with that of the noise-free target map. This was followed by a coarse global rotation search at the lowest resolution (20 Å) with step size determined by the target map size to ensure proper coverage of search space. As a final step, a six-dimensional refinement (rotation and translation) of the top-scoring solution from the previous step was performed. For the higher resolution targets and the noisy maps,

only a refinement for the pre-positioned decoy was performed to speed up the process. We run a complete global docking protocol that includes full rotation/translation search with refinement for all three noise models at all resolutions on a representative subset of structures (~5,000) to test whether the strategy outlined above may lead to serious errors in estimating the best possible density score for any given target. For all cases tested, the sped-up strategy and the full global protocol gave essentially the same scores. With the speed up, docking a single decoy into a target map took about 25s in average on a single processor Linux machine. We employed 70 processors on a Linux cluster to perform the roughly 480,000 docking experiments for this study, taking a total of approximately 48 hours.

For evaluation of performance, we estimated the probability of the top ranked solution to have the lowest RMSD (P^1), be within the lowest two RMSDs (P^2), lowest five RMSDs (P^5) and the lowest ten RMSDs (P^{10}) of the decoy set. We also estimated the probability of the top 10 ranking solution to contain the lowest RMSD structure ($P^{1/10}$) and the probability of one of the lowest 10 RMSD structures being present in the top ten ranking solutions ($P^{10/10}$). In addition, we calculated the correlation coefficient (CC) between the density score and the mean deviation of the decoy with the target structure. Probabilities were directly estimated from the respective frequency distributions.

Experimental data

In order to test performance in a real-life scenario, we used an experimental 28-Å resolution map of human rhinovirus complexed with Fab fragments (Smith et al., 1993). The crystal structure of exactly the same construct was solved to atomic resolution a few years later (Smith et al., 1996) giving a one-to-one correspondence between the low-resolution density obtained by EM and the corresponding atomic structure. This gives us the opportunity to test our model selection procedure on an experimental map with known content at atomic resolution. For the tests performed here, we pretended that only the EM reconstruction, the structure of the uncomplexed virion (Rossmann et al., 1985), and homology models for the Fab fragment are available. The task would be to select a good model for the bound Fab fragment from the pool of homology models.

To isolate the Fab density from the reconstruction, we went through a number of steps. First, we docked the crystal structure of the uncomplexed virion into the reconstruction using CoAn. Then, we removed the contribution of the docked model using discrepancy mapping (Volkman and Hanein, 2003). One of the symmetry-related densities corresponding to a single Fab fragment was boxed out from the discrepancy map for further analysis. The fragment used to decorate the virion consists of two domains, clearly visible in the EM reconstruction. Application of watershed segmentation (Volkman, 2002) to the discrepancy map readily yields two segments. Only the domain directly attached to the virion is ordered in the crystal structure. This fragment was also solved by crystallography as an isolated entity (Liu et al., 1994) and was used for generating homology models for the Decoys'R'Us database (id: 1for). Thus, we restricted our analysis to the density segment corresponding to that domain. The correlation of the extracted density with the crystal structure of the virion-bound Fab fragment is 0.987. For comparison purposes we also generated a synthetic EM density from the bound Fab fragment at 28 Å resolution and with added Gaussian noise. RMSDs with homology models were calculated in respect to the virion-bound fragment after least-squares fitting of the α -carbons.

Results

The decoy database is divided into a number of subsets. These subsets contain decoys generated at different levels of accuracy using different structure prediction methods. The subsets and the results of the respective density scoring experiments are summarized below. In addition,

the results for experimental data are described. Individual score evaluations for each target structure in all the decoy set used are provided as supplementary material.

Four-state lattice model test set

This set (denoted 4state_reduced in the database) contains decoys for seven proteins. The α -carbon positions for these decoys were generated by exhaustively enumerating ten selectively chosen residues in each protein using a four-state off-lattice model (Park and Levitt, 1995). All other residues positions were based on the best fit of a four-state model to the native chain. The resulting conformations were ranked using steric and compactness criteria as well as a variety of scoring functions (Park and Levitt, 1996). The top ~665 solutions were kept in the decoy set for each target. The RMSD between the decoys and their targets were in average between 1.15 and 8.80 Å with a mean value of 5.21 Å. At 8 Å resolution and in the absence of noise, the average RMSD of the top ranked solution was 1.32 Å, P^5 was 0.86 and the $P^{1/10}$ was 1.00. The CC was 0.82. Lowering the resolution and adding noise to the target maps generally degraded all these numbers but still maintained reasonable performance (see Table 1). For all resolutions and noise models $P^{10/10}$ was 1.00 with one exception (0.86), indicating the high quality of density ranking for this decoy set.

Local minima decoy set

This set (denoted lmds_v2 in the database) contains decoys for ten small proteins that belong to diverse structural classes. Each decoy is derived as a local minimum of an energy function based on the ENCAD force field (Levitt, 1974; Levitt and Lifson, 1969). Initial conformations for each decoy were generated by randomizing the torsion angles of the loop regions. The set contains the 500 lowest local energy minima for each protein. The RMSD between the decoys and their targets were in average between 4.04 and 11.58 Å with a mean value of 7.83 Å. At 8 Å resolution and in the absence of noise, the RMSDs of the top ranking solutions are in average 5.43 Å and P^5 and $P^{1/10}$ are both 0.60. The CC was 0.30 and $P^{10/10}$ was 1.00 over all resolutions and noise models except for two (both 0.90).

Fragment insertion simulated annealing test set

This set (denoted fisa in the database) contains decoys for four α -helical proteins. For each of the proteins, 500 decoys were generated using a fragment insertion simulated annealing procedure to assemble native-like structures from fragments of unrelated protein structures with similar local sequences using Bayesian scoring functions (Simons et al., 1997). The RMSD for the decoys is in average between 3.73 and 12.56 Å with a mean value of 7.51 Å. At 8 Å resolution and in the absence of noise, the RMSD of the top ranking solution are in average 5.01 Å. P^5 and $P^{1/10}$ are 0.25 and 0.50 respectively. The CC was 0.35 and $P^{10/10}$ was 1.00. At resolutions better than 15 Å, $P^{10/10}$ was 1.00 or 0.75 for all noise models but drops to 0.25 at 20 Å resolution.

CASP3 blind prediction decoy set

This set (denoted fisa_casp3 in the database) contains six targets of the CASP3 (Critical Assessment of methods for Structure Prediction) blind prediction experiment (Zemla et al., 1999). The predictions were generated using fragment insertion simulated annealing (Simons et al., 1997). For each of the targets, an average of 1432 decoys were provided. The RMSD between the decoys and their target structures are between 6.08 and 18.63 Å with a mean value of 12.9 Å. At 8 Å resolution and in the absence of noise, the RMSDs of the top ranking solutions are in average 7.67 Å. P^5 and $P^{1/10}$ are both 0.25. The CC was 0.41 and $P^{10/10}$ was 0.50 for all noise models and to a resolution down to 15 Å. At 20 Å resolution $P^{10/10}$ is 0.25.

Ab-initio methods

This set (denoted `lattice_ssfit` in the database) contains conformations for eight proteins generated by an ab-initio prediction methods (Samudrala et al., 1999). Briefly, the conformational space for a sequence is exhaustively enumerated and scored on a tetrahedral lattice (Hinds and Levitt, 1994). The predicted secondary structures (Rost et al., 1994) are then fitted to these conformations using the four-state model (Park and Levitt, 1996). The 2000 best scoring conformations for each protein are used for this set. The RMSD for the decoys is between 5.68 and 13.23 Å with a mean value of 9.93 Å. The CC is close to random with 0.03 and for none of the tests any of the probabilities was different from 0.00 except for $P^{10/10}$ which was 0.13 in 3 out of 12 conditions. For this decoy set, the density score does not work well as a ranking method.

Homology modeling

The decoy set contains 90 homology modeling experiments. This includes decoys for 29 globins (`hg_structal` in the database) where each globin has been built by homology modeling using 29 other globins. It further contains decoys for 61 immunoglobulins (`ig_structal` in the database) that have been built by homology modeling using all the other immunoglobulins as templates. The range of RMSD was between 1.39 and 8.56 Å RMSD with a mean value of 3.25 Å. At 8 Å resolution, the RMSD of the best-scoring solution is in average 1.48 Å, the CC is 0.83 and P^2 is 0.63. $P^{10/10}$ is between 0.95 and 1.00 for all noise and resolution conditions.

Experimental data

The 60 homology models built on target 1 for from the immunoglobulin decoy set were used with the experimental and synthetic densities of 1 for at 28 Å resolution. Both analyses do not place the best model in the 10 highest ranking solutions ($P^{1/10} = 0$) and both place the fourth best homology model at rank 3 ($P^{10/10} = 1$). The CC is 0.44 and 0.52, the RMSD of the highest ranking model 1.82 Å and 1.33 Å for the experimental and synthetic densities respectively.

Conclusions

We evaluated the utility of a density-based scoring function for selecting near-native atomic models from a pool of less accurate structures using a large decoy data set that spans a representative set of prediction methods and structure types. We find that the general relationship between the model quality and the density score follows the same behavior, independent of the particular structure type, size and prediction method used (Fig. 1 and Table 1). Decoys with RMSDs below a certain “magic” cutoff (usually ~3 Å) generally fall into a funnel-like area of the model-quality density-score space (Fig. 1L, pink) that exhibits high correlation between the two variables, usually larger than 0.7. The density score is an extremely efficient model selector within that funnel zone, virtually guaranteeing that one of the lowest RMSD structures is selected within the top scoring solutions. The funnel area is overlaid with a “dead zone”, where a large number of decoys with variable RMSDs (but larger than the magic cutoff) exhibit similar density scores. The addition of random Gaussian or impulse noise only reorders decoys within the funnel and dead-zone areas without significantly changing the shape and characteristics of these areas or the probability of selecting high-quality models (see Fig. 1A-C and Table 1). Lowering the resolution of the underlying density maps shifts the front of the dead zone towards the funnel area until it eventually overlaps the funnel completely so that model selection is incapacitated (Fig. 1D-F). The exact funnel width, magic cutoff and incapacitating resolution appears to depend on the individual target structure rather than the prediction method (Fig. 1J-K). We did not find any correlation with structure size or type (i.e. helical versus β -barrel). In most cases, 20 Å resolution is still sufficient for selecting high-quality models, provided the funnel is sufficiently populated (Table 1). If the funnel is not populated at all (as is the case for the ab-initio decoy set), the density score is insufficient for

selecting the higher-quality models and does not work at all. However, if at least a few decoys are located in the funnel area, even among prediction schemes that predominantly populate the dead zone (see for example Fig. 1G-H), the probability of picking one of the lowest few RMSD structures based on the density score is close to 1.0.

Most of the structures in the decoy set are too small to be amenable for direct structure analysis by electron microscopy in order to obtain density information. However, the observed independence of our findings from the size of the molecule indicates that our conclusions will also hold for larger molecules that are directly amenable to electron cryomicroscopy. Also, the density of smaller proteins and domains can be often be extracted from EM reconstructions of larger complexes. In fact, several of the target structures in the decoy set were solved as parts of larger assemblies. One example is the c-terminal domain of ribosomal protein L7/L12 from *Escherichia coli* (1ctf, Leijonmarck and Liljas, 1987) which is part of the four-state lattice model, ab-initio and local minima decoy sets. This structure was solved by electron cryomicroscopy as part of the *Escherichia coli* 70S ribosome (Gao et al., 2003). Another example is the Fab fragment that neutralizes human rhinovirus (1for, Liu et al., 1994) from the immunoglobulin homology decoy set that was solved by electron cryomicroscopy as part of a human rhinovirus-Fab complex (Smith et al., 1993). For the latter, an atomic resolution structure (Smith et al., 1996) is available as well, facilitating a one-to-one comparison between the model-selection performance of experimental and synthetic. Despite the distortions introduced by the extraction of the experimental density of the Fab fragment from the larger complex, there is little difference in performance as compared to synthetic density at the same resolution (Fig. 1 N-O). Both sample the same area of RMSD/density-score space, clearly reminiscent of the funnel/dead-zone shape detected by the decoy-set analyses. The additional distortions introduced by the density extraction process does not appear to alter performance significantly. The robustness of the scoring in terms of additive noise, resolution, and distortions of the target density suggests that alternative, lower-resolution techniques applicable to smaller structures such as negativestain electron microscopy and X-ray solution scattering can also be used within the same density-scoring structure-prediction framework to select high-quality models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Drs. Tom Smith and Tim Baker for providing the reconstruction of human rhinovirus complexed with Fab fragments. We would like to thank Jonathan Sexton for initial test calculations and Martin Fleming for the initial setup of the Linux cluster used in this study. This work was supported by NIH grant GM076503 to NV.

References

- Baker ML, Serysheva S II, Sencer Y, Wu SJ, Ludtke W, Jiang SL, Hamilton, Chiu W. The skeletal muscle Ca²⁺ release channel has an oxidoreductase-like domain. *Proc Natl Acad Sci U S A* 2002;99:12155–60. [PubMed: 12218169]
- de Beer T, Carter RE, Lobel-Rice KE, Sorkin A, Overduin M. Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science* 1998;281:1357–60. [PubMed: 9721102]
- Fotin A, Cheng Y, Sliz P, Grigorieff N, Harrison SC, Kirchhausen T, Walz T. Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 2004;432:573–9. [PubMed: 15502812]
- Gao H, Sengupta J, Valle M, Korostelev A, Eswar N, Stagg SM, Van Roey P, Agrawal RK, Harvey SC, Sali A, Chapman MS, Frank J. Study of the structural dynamics of the *E coli* 70S ribosome using real-space refinement. *Cell* 2003;113:789–801. [PubMed: 12809609]

- Hanein D, Volkman N, Goldsmith S, Michon AM, Lehman W, Craig R, DeRosier D, Almo S, Matsudaira P. An atomic model of fimbrin binding to F-actin and its implications for filament crosslinking and regulation. *Nat Struct Biol* 1998;5:787–92. [PubMed: 9731773]
- Hinds DA, Levitt M. Exploring conformational space with a simple lattice model for protein structure. *J Mol Biol* 1994;243:668–82. [PubMed: 7966290]
- Kissinger CR, Liu BS, Martin-Blanco E, Kornberg TB, Pabo CO. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 1990;63:579–90. [PubMed: 1977522]
- Leijonmarck M, Liljas A. Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. *J Mol Biol* 1987;195:555–79. [PubMed: 3309338]
- Levitt M. Energy refinement of hen egg-white lysozyme. *J Mol Biol* 1974;82:393–420. [PubMed: 4856348]
- Levitt M, Lifson S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J Mol Biol* 1969;46:269–79. [PubMed: 5360040]
- Liu H, Smith TJ, Lee WM, Mosser AG, Rueckert RR, Olson NH, Cheng RH, Baker TS. Structure determination of an Fab fragment that neutralizes human rhinovirus 14 and analysis of the Fab-virus complex. *J Mol Biol* 1994;240:127–37. [PubMed: 8027997]
- Liu J, Taylor DW, Taylor KA. A 3-D reconstruction of smooth muscle alpha-actinin by CryoEm reveals two different conformations at the actin-binding region. *J Mol Biol* 2004;338:115–25. [PubMed: 15050827]
- Mondragon A, Wolberger C, Harrison SC. Structure of phage 434 Cro protein at 2.35 Å resolution. *J Mol Biol* 1989;205:179–88. [PubMed: 2647998]
- Park B, Levitt M. Energy functions that discriminate X-ray and near native folds from wellconstructed decoys. *J Mol Biol* 1996;258:367–92. [PubMed: 8627632]
- Park BH, Levitt M. The complexity and accuracy of discrete state models of protein structure. *J Mol Biol* 1995;249:493–507. [PubMed: 7783205]
- Rhee S, Martin RG, Rosner JL, Davies DR. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc Natl Acad Sci U S A* 1998;95:10413–8. [PubMed: 9724717]
- Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ, Johnson JE, Kamer G, Luo M, Mosser AG, Rueckert RR, Sherry B, Vriend G. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 1985;317:145–53. [PubMed: 2993920]
- Rost B, Sander C, Schneider R. PHD--an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci* 1994;10:53–60. [PubMed: 8193956]
- Samudrala R, Levitt M. Decoys 'R' Us: a database of incorrect conformations to improve protein structure prediction. *Protein Sci* 2000;9:1399–401. [PubMed: 10933507]
- Samudrala R, Xia Y, Huang E, Levitt M. Ab initio protein structure prediction using a combined hierarchical approach. *Proteins* 1999;(Suppl):194–8. [PubMed: 10526368]
- Saul FA, Poljak RJ. Crystal structure of human immunoglobulin fragment Fab New refined at 2.0 Å resolution. *Proteins* 1992;14:363–71. [PubMed: 1438175]
- Sengupta J, Nilsson J, Gursky R, Spahn CM, Nissen P, Frank J. Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. *Nat Struct Mol Biol* 2004;11:957–62. [PubMed: 15334071]
- Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 1997;268:209–25. [PubMed: 9149153]
- Smith TJ, Chase ES, Schmidt TJ, Olson NH, Baker TS. Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. *Nature* 1996;383:350–4. [PubMed: 8848050]
- Smith TJ, Olson NH, Cheng RH, Liu H, Chase ES, Lee WM, Leippe DM, Mosser AG, Rueckert RR, Baker TS. Structure of human rhinovirus complexed with Fab fragments from a neutralizing antibody. *J Virol* 1993;67:1148–58. [PubMed: 7679742]
- Topf M, Sali A. Combining electron microscopy and comparative protein structure modeling. *Curr Opin Struct Biol* 2005;15:578–85. [PubMed: 16118050]

- Topf M, Baker ML, John B, Chiu W, Sali A. Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. *J Struct Biol* 2005;149:191–203. [PubMed: 15681235]
- Topf M, Baker ML, Marti-Renom MA, Chiu W, Sali A. Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. *J Mol Biol* 2006;357:1655–68. [PubMed: 16490207]
- Volkman N. A novel three-dimensional variant of the watershed transform for segmentation of electron density maps. *J Struct Biol* 2002;138:123–32. [PubMed: 12160708]
- Volkman N, Hanein D. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol* 1999;125:176–84. [PubMed: 10222273]
- Volkman N, Hanein D. Docking of atomic models into reconstructions from electron microscopy. *Meth Enzym* 2003;374:204–225. [PubMed: 14696375]
- Volkman N, DeRosier D, Matsudaira P, Hanein D. An atomic model of actin filaments cross-linked by fimbrin and its implications for bundle assembly and function. *J Cell Biol* 2001;153:947–56. [PubMed: 11381081]
- Volkman N, Hanein D, Ouyang G, Trybus KM, DeRosier DJ, Lowey S. Evidence for cleft closure in actomyosin upon ADP release. *Nat Struct Biol* 2000;7:1147–55. [PubMed: 11101898]
- Zemla A, Venclovas C, Moulton J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. *Proteins* 1999;(Suppl):22–9. [PubMed: 10526349]
- Zhao B, Carson M, Ealick SE, Bugg CE. Structure of scorpion toxin variant-3 at 1.2 Å resolution. *J Mol Biol* 1992;227:239–52. [PubMed: 1522588]

Figure 1. Representative plots of mean deviation between decoys and target structure (in Å, y-axes) versus density scores (x-axes)

A-C: Effect of additive noise. Shown are the results for scorpion toxin variant-3 (1sn3, Zhao et al., 1992) in the four-state lattice model decoy set. Density maps were generated from the target structure at 8 Å resolution with no noise added (A) and Gaussian random noise (B) or Laplacian impulse noise (C) added at signal-to-noise ratio of 0.5.

D-F: Effect of resolution. Shown are the results for engrailed homeodomain (1hhd-C, Kissinger et al., 1990) in the fragment insertion simulated annealing decoy set. Density maps used were generated at 8 (D), 15 (E) and 20 (F) Å. The grey dotted lines indicate the approximate front of the dead zone (see L).

G-H: Results for the noise-free, 8 Å resolution density maps of members of the casp3 decoy set. The corresponding decoys were generated in a blind test setting. Shown are (G) an Eps15 homology domain fragment (1eh2, de Beer et al., 1998) and (H) the DNA-binding motif in MarA (1bl0, Rhee et al., 1998).

I: Results for the noise-free, 8 Å resolution density map of human immunoglobulin fragment fab (8fab, Saul and Poljak, 1992) from the immunoglobulin homology decoy set.

J: Results for the noise-free, 8 Å resolution density map of the c-terminal domain of ribosomal protein L7/L12 from *Escherichia coli* (1ctf, Leijonmarck and Liljas, 1987). This molecule was used in three different decoy sets which were the four-state lattice model set (blue), the ab-initio set (red) and the local minima set (yellow).

K: Results for the noise-free, 8 Å resolution density map of phage 434 cro protein (2cro, Mondragon et al., 1989). This molecule was used in three different decoy sets which were the four-state lattice model set (blue), the fragment insertion simulated annealing set (red) and the local minima set (yellow).

L: Schematic drawing of the general landscape of the density-score model-accuracy relationship. For all decoy sets that contain structures with better than ~3 Å RMSD (red dotted line) a funnel of high correlation exists between high density scores and model accuracy (pink). The density score is a very efficient model selector within the funnel. A “dead zone” (light blue) of structures with a large variability of RMSDs but similar density-scores exists to the left of the pure funnel zone. In this zone there is little correlation between model accuracy and the density score which is not useful for model selection in this area.

M-N: Results for synthetic (M) and experimental (N) density at 28 Å resolution from a Fab fragment that neutralizes human rhinovirus (1for, Liu et al., 1994; Smith et al., 1996). Gaussian noise was added to the synthetic data (signal-to-noise ratio of 0.8).

O: Correlation of RMSD as a function of density-score rank between the experimental (x-axis) and synthetic data (y-axis) of the Fab fragment 1for (M-N). The homology models were sorted according to density scores for both densities and then the RMSDs for each rank were plotted.

Table 1

Evaluation of decoy sets by density scoring

Reso ^a	P ¹	P ²	P ⁵	P ¹⁰	P ^{1/10}	P ^{10/10}	CC ^b	RMSD ^c
Four-state lattice model set								
8	0.167	0.333	0.857	0.833	1.000	1.000	0.824	1.316
10	0.143	0.429	0.714	0.857	1.000	1.000	0.798	1.350
15	0.429	0.571	0.571	0.857	0.857	1.000	0.721	1.361
20	0.143	0.143	0.143	0.571	0.857	1.000	0.639	1.769
<i>Gaussian random noise</i>								
8	0.286	0.571	0.571	0.714	0.714	1.000	0.815	1.613
10	0.286	0.429	0.857	0.857	0.571	1.000	0.810	1.515
15	0.286	0.286	0.429	0.714	0.714	1.000	0.705	1.657
20	0.143	0.143	0.286	0.571	0.714	0.857	0.626	1.793
<i>Laplacian impulse noise</i>								
8	0.143	0.286	0.571	0.714	0.857	1.000	0.812	1.612
10	0.143	0.143	0.429	0.429	0.714	1.000	0.807	1.813
15	0.286	0.286	0.571	0.714	0.714	1.000	0.684	1.561
20	0.143	0.286	0.429	0.571	0.429	1.000	0.602	1.641
Local minima set								
8	0.400	0.500	0.600	0.700	0.600	1.000	0.297	5.425
10	0.300	0.400	0.400	0.400	0.700	1.000	0.295	6.368
15	0.200	0.200	0.300	0.300	0.700	1.000	0.286	7.018
20	0.200	0.200	0.200	0.200	0.600	1.000	0.267	7.077
<i>Gaussian random noise</i>								
8	0.200	0.300	0.400	0.400	0.500	1.000	0.284	6.666
10	0.200	0.200	0.200	0.200	0.500	1.000	0.283	7.524
15	0.200	0.200	0.200	0.300	0.700	0.900	0.261	7.403
20	0.200	0.200	0.200	0.300	0.600	0.900	0.250	6.327
<i>Laplacian impulse noise</i>								
8	0.300	0.400	0.400	0.400	0.700	1.000	0.295	6.283
10	0.300	0.400	0.500	0.600	0.600	1.000	0.304	6.061
15	0.200	0.200	0.400	0.600	0.600	1.000	0.300	6.183
20	0.400	0.400	0.500	0.500	0.400	1.000	0.229	6.315
Fragment insertion set								
8	0.250	0.250	0.250	0.250	0.500	1.000	0.347	5.009
10	0.000	0.250	0.250	0.250	0.250	0.750	0.337	5.134
15	0.000	0.000	0.250	0.250	0.250	0.500	0.293	5.181
20	0.000	0.000	0.000	0.000	0.000	0.250	0.244	5.806
<i>Gaussian random noise</i>								
8	0.250	0.250	0.250	0.250	0.250	0.750	0.340	5.104
10	0.250	0.250	0.250	0.250	0.250	0.750	0.349	5.142
15	0.000	0.000	0.250	0.250	0.250	0.250	0.305	5.426
20	0.000	0.000	0.000	0.000	0.000	0.250	0.274	5.806
<i>Laplacian impulse noise</i>								
8	0.250	0.250	0.250	0.250	0.250	1.000	0.341	5.221
10	0.000	0.250	0.250	0.250	0.250	0.750	0.367	5.134
15	0.000	0.250	0.250	0.250	0.250	0.500	0.321	5.378
20	0.000	0.000	0.000	0.000	0.000	0.250	0.228	7.668
Casp3 set								
8	0.000	0.250	0.250	0.250	0.250	0.500	0.410	7.627
10	0.000	0.250	0.250	0.250	0.250	0.500	0.410	7.627
15	0.000	0.000	0.000	0.000	0.250	0.500	0.412	9.421
20	0.000	0.000	0.000	0.000	0.250	0.250	0.387	9.767
<i>Gaussian random noise</i>								
8	0.000	0.250	0.250	0.250	0.250	0.500	0.410	7.956
10	0.000	0.250	0.250	0.250	0.250	0.500	0.414	7.933

Reso ^a	P ¹	P ²	P ⁵	P ¹⁰	P ^{1/10}	P ^{10/10}	CC ^b	RMSD ^c
15	0.000	0.250	0.250	0.250	0.250	0.500	0.419	8.046
20	0.000	0.000	0.000	0.000	0.250	0.250	0.392	10.37
<i>Laplacian impulse noise</i>								
8	0.000	0.250	0.250	0.250	0.250	0.500	0.404	7.627
10	0.000	0.250	0.250	0.250	0.250	0.500	0.410	7.627
15	0.000	0.250	0.250	0.250	0.250	0.500	0.412	9.361
20	0.000	0.000	0.000	0.000	0.250	0.250	0.392	10.27
Ab-initio set								
8	0.000	0.000	0.000	0.000	0.000	0.000	0.032	10.18
10	0.000	0.000	0.000	0.000	0.000	0.000	0.032	10.18
15	0.000	0.000	0.000	0.000	0.000	0.000	0.038	9.586
20	0.000	0.000	0.000	0.000	0.000	0.000	0.037	9.639
<i>Gaussian random noise</i>								
8	0.000	0.000	0.000	0.000	0.000	0.000	0.017	10.72
10	0.000	0.000	0.000	0.000	0.000	0.000	0.018	9.535
15	0.000	0.000	0.000	0.000	0.000	0.125	0.024	8.697
20	0.000	0.000	0.000	0.000	0.000	0.000	0.030	9.496
<i>Laplacian impulse noise</i>								
8	0.000	0.000	0.000	0.000	0.000	0.000	0.017	10.54
10	0.000	0.000	0.000	0.000	0.000	0.125	0.021	9.876
15	0.000	0.000	0.000	0.000	0.000	0.000	0.039	9.701
20	0.000	0.000	0.000	0.000	0.000	0.125	0.040	9.324
Homology set								
8	0.558	0.574	0.787	0.869	0.836	1.000	0.733	1.560
10	0.426	0.508	0.705	0.770	0.836	0.984	0.761	1.609
15	0.426	0.525	0.672	0.820	0.803	1.000	0.697	1.486
20	0.410	0.557	0.656	0.885	0.770	0.984	0.655	1.518
<i>Gaussian random noise</i>								
8	0.246	0.443	0.525	0.689	0.770	0.984	0.748	1.755
10	0.361	0.459	0.590	0.803	0.803	1.000	0.745	1.570
15	0.393	0.508	0.623	0.836	0.770	0.984	0.684	1.587
20	0.361	0.525	0.656	0.820	0.754	0.951	0.656	1.644
<i>Laplacian impulse noise</i>								
8	0.295	0.459	0.607	0.787	0.705	0.951	0.739	1.716
10	0.328	0.508	0.639	0.787	0.770	0.967	0.744	1.630
15	0.361	0.492	0.639	0.869	0.787	1.000	0.688	1.588
20	0.230	0.393	0.525	0.738	0.656	0.984	0.644	1.677

^aReso refers to resolution, p1,2,5,10 to the probability of identifying the lowest, second, fifth, tenth lowest RMSD structure as the best scoring solution. p1/10 is the probability to find the lowest RMSD structure in the 10 best scoring solutions, p10/10 is the probability to find one of the 10 lowest RMSDs among the 10 highest scoring solutions.

^bbCC is the correlation coefficient between the density score and the mean deviation between the target structure and the respective decoy.

^cRMSD is the root-mean-square deviation between the highest scoring solution and the target structure.