Title
Q|R: quantum-based refinement

Journal
Acta Crystallographica Section D, Structural Biology, 73(1)

ISSN
2059-7983

Authors
Zheng, Min
Reimers, Jeffrey R
Waller, Mark P
et al.

Publication Date
2017

DOI
10.1107/s2059798316019847

Peer reviewed

Q|R: Quantum-based Refinement

Min Zheng[1], Jeffrey R. Reimers[2], Mark P. Waller[2]*, Pavel V. Afonine[3]*

[1] *Theoretische Organische Chemie, Organisch-Chemisches Institut and Center for Multiscale Theory and Computation, Westfälische Wilhelms-Universität Münster, Corrensstraße 40, 48149 Münster, Germany*
[2] *International Center for Quantum and Molecular Sciences, Shanghai University, Shanghai, China*
[3]*Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

* waller@shu.edu.cn, PAfonine@lbl.gov

Synopsis

Quantum-based refinement software is being developed to refine bio-macromolecules against crystallographic or cryo-electron microscopy data.

Abstract

Quantum-based refinement utilizes chemical restraints derived from quantum chemical methods, instead of the standard parameterized library-based restraints used in refinement packages. The motivation is two fold: firstly, the restraints have the potential to be more accurate and secondly, the restraints can be more easily applied to new molecules such as drugs or novel co-factors. Here we introduce a new project called Q|R aimed at developing a quantum-based refinement of bio-macromolecules. The central focus of this long-term project is to develop software that is built on top of open source components. A development version of Q|R was used to compare quantum-based refinements to standard refinement using a small model system.

Keywords: Quantum refinement, structural biology, X-ray diffraction, neutron diffraction, cryo-EM, CCTBX.

1. Introduction

Crystallography accounts for about 90% of all structures in Protein Data Bank (PDB; Bernstein *et al*., 1977; Berman *et al*., 2000), and is therefore the leading tool for obtaining three-dimensional structures of bio-macromolecules. Cryo-electron microscopy (cryo-EM) is rapidly becoming its major competitor (Bai *et al*., 2015; Cheng, 2015). These two methods are rather different from a technological and conceptual perspective (Frank, 2006; Rupp, 2010), however they both yield a map that is used to build an initial atomic model into. Model refinement against experimental data is the next common step in the process for both of these two structure solution techniques. For cryo-EM, the experimental data is used to construct a map, and that map normally does not change during the refinement procedure. For crystallography, the experimental data are the measured intensities of structure factors, and, since phases are lost in diffraction experiment, the map is typically calculated using model phases. This implies that the map is constantly changing, since it depends on the model that changes during refinement. It turns out that despite these technical and methodological nuances, the computational refinement tools are very similar if not identical for both techniques. Therefore we now refer to crystallographic or cryo-EM experimental data as *experimental data* or simply *data*.

A general refinement protocol is shown schematically in Figure 1. Given an atomic model and experimental data, the refinement engine calculates a refinement target and its derivatives with respect to atomic parameters, which are then sent to an optimizer (typically, a minimizer). The minimizer updates the model parameters, and then sends them back to the refinement engine that then calculates a new target value and set of derivatives and returns them back to the minimizer. This process is carried out iteratively until convergence is achieved. Finer details and specific implementation depend on particular software and experimental data (X-ray, neutron or cryo-EM, for example). Model refinement against experimental data is an optimization process of changing parameters that describe the model to satisfy a goal (or target) function. A target function relates model parameters to experimental data and, if needed, *a priori* knowledge (for reviews, see Tronrud, 2004; Watkin, 2008; Afonine *et al*., 2015). In the case of bio-macromolecules the data are almost always of insufficient quality to be used alone in refinement, and thus using *a priori* knowledge is almost always needed, with the exception being ultra-high resolution data that constitute less than 0.5% of all entries in the PDB.

A priori knowledge is typically introduced as constraints or as a weighted term to the refinement target function

$$T = T_{data} + w * T_{restraints} \quad (1)$$

and is hereafter called restraints. Here T_{data} is typically referred to as experimental or the data term, the term that scores model to data fit, $T_{restraints}$ represents restraints and w is the relative weight.

Most popular refinement packages such as Refmac (Murshudov *et al*., 2011), Shelxl (Sheldrick, 2008), CNS (Brunger *et al.*, 1998), BUSTER-TNT (Bricogne *et al*., 2016) or *phenix.refine* (Afonine *et al*., 2012) use a sum of potentials (e.g. harmonic) to restrain specific features of atomic model, such as bond lengths or angles, or planes of planar groups. Typically it is a sum of six terms

$$T_{restraints} = T_{bond} + T_{angle} + T_{planarity} + T_{chirality} + T_{torsion} + T_{nonbonded\_repulsion} \quad (2)$$

where each term is responsible for a particular feature: covalent bonds and angles, planes, chiral volumes, torsion angles and preventing nonsensical steric clashes.

This kind of restraint is sufficient most of the time at data resolutions of 2.5-3Å or better. However, for lower resolutions that account for about 20% of crystallographic data in PDB, or for resolutions typically found in cryo-EM, these restraints are insufficient. Indeed, at a typical macromolecular resolution (around 2Å) there is insufficient information to determine the atomic level of detail, but it does contain information about secondary and higher order structural organization. Restraints such as those in (2) are needed to compensate for this lack of information. Lower resolution data may not only lack atomic details but also higher-order details. The impact of poorly performing restraints (2) during a low-resolution refinement is at least two-fold. Firstly, the geometry of a refined model may not be sound; for example, alpha-helices and beta-strands may be distorted while still fitting the map and satisfying the restraints (2). Secondly, the data over-fitting may be significant because the amount of data (experimental plus restraints) may be severely outweighed by the amount of model parameters.

To address these problems, additional restraints have been used to augment (2) (see, for example, Oldfield, 2000; Echols *et al*., 2010; Headd *et al*., 2012; Sobolev *et al*., 2015)

$$T_{restraints\_plus} = T_{restraints} + T_{SS} + T_{Ramachandran} + T_{rotamer} + T_{reference} \quad (3).$$

Here T_{SS} is secondary structure restraints, which are essentially restraints on hydrogen bond distances and angles, $T_{Ramachandran}$ restrain torsion angles of protein main chain against Ramachandran plot (Ramachandran *et al.* 1963), $T_{rotamer}$ restrain amino-acid side-chains to valid rotameric states and $T_{reference}$ restraints a model being refined against low-resolution data to a known better quality model solved against higher resolution data. Other, similar restraints may be envisaged.

Restraints for a standard refinement target are functions of (2) and possibly (3), while these additional restraints are clearly an improvement, they are not without problems. For example, they require manual annotation (a user needs to tell the program what the secondary structure is) and they are still simple potentials. These potentials are fitted to reproduce some average value taken from a compiled library, and do not take into account finer details such as local environment and nearby charges.

Quantum chemical methods have the potential to play a transformational role in refinement by delivering restraints in much less *ad hoc* way, and this can potentially lead to more chemically meaningful structures (Carlsen 2015). In quantum-based refinement the restraints are derived from a quantum chemical calculation. Performing an accurate and efficient quantum chemical calculation for macromolecules remains a challenge in computational chemistry (Borbulevych *et al*., 2014; Borbulevych *et al*., 2016; Goerigk *et al.*, 2014). However, several attempts at using quantum chemical calculations as a source of restraints for crystallographic refinement have been reported before and can be categorized as follows:

Hybrid QM/MM. A refinement procedure can be focused on an "*active*" region of a protein. The advantage is obviously that one does not waste computational resources trying to better describe the (potentially uninteresting) environment region. The QM/MM-based refinement method advocated by Ryde and co-workers (Ryde, 2003; Ryde & Nilsson 2003; Ryde & Nilsson 2003; Ryde, 2003; Nilsson *et al*., 2004) was pioneering in this area. The ComQum (Ryde, 1996) software package was developed for this task. The challenge of hybrid QM/MM based methods is that one needs to carefully select the active QM region, ensuring that a sufficiently large enough region is taken. This can be time-consuming and labor intensive to carry out convergence studies, and furthermore, finding a balanced force-fields and *ab initio* combination remains an open area for QM/MM modelling.

Semi-empirical. Seminal work by Merz and co-workers have managed to address a lot of issues in quantum refinement (Yu *et al*., 2005; Yu, Li *et al*., 2006; Yu, Hayik *et al*., 2006; Li *et al*., 2012; Fu *et al*., 2013; Fu *et al*., 2013; Borbulevych *et al*., 2014). Using semi-empirical calculations more-or-less alleviates the issue of computational scaling. The DivCon (Dixon & Merz, 1996) software was used for this purpose and has been interfaced with *Phenix* (Adams *et al*., 2010). Employing semi-empirical methods is attractive due to their inherent computationally more efficient scaling. However, the accuracy and robustness (e.g. metalloenzymes) issues may prove to be too much of a drawback in the long run (Korth & Thiel 2011).

Linear-Scaling Density Functional Theory. The work by Reimers and co-workers (Canfield *et al*., 2006) employed a divide and conquer based QM/MM optimization approach to study a 150,000-atom photosystem-I trimer. The whole protein is divided into individually optimized regions, with each region (and its immediate environment) treated by a density functional theory (DFT) and the remaining protein by molecular mechanics. This study used the forces coming from the DFT calculation to optimize the structure. This calculation found the structural feature that held the trimer together. Serious errors in the coordinates the chlorophyll "special pair" were identified. The orientations of 35 residue sidechains were optimized to make improved hydrogen bonding networks.

Quantum methods such as semi-empirical, *Hartee-Fock, Quantum Monte Carlo or DFT* can be used to calculate restraints for co-factors, co-crystals, drugs bound to active sites *etc*. The primary concern with quantum based methods is the tremendous amount of computing resources required, however recent progress in developing very efficient code, accelerated by general purpose processing units (GPUs), now offers an exciting glimpse into a promosing future for quantum-based refinement.

2. Methods

We have developed a refinement package that combines crystallographic and cryo-EM data with restraints computed using standard quantum-chemical methods. Our quantum refinement code is called Q|R. The source code was written as a lightweight standalone Python (for example, v2.7) program and the source code is made freely available at https://github.com/qrefine/qr-core. Q|R uses the CCTBX open source library (Grosse-Kunstleve & Adams, 2002; Grosse-Kunstleve *et al*., 2002) to construct a standard refinement protocol, very much like *phenix.refine* and most other *Phenix* tools. CCTBX is used to compute the data term in (Eqn. 1) and its derivatives, do scaling and account for bulk-solvent, and drive refinement using a standard L-BFGS (Liu & Nocedal, 1989; Byrd *et al.*, 1995) minimizer.

CCTBX is also used to compute $T_{restraints}$ (Eqn. 1) and its derivatives in standard refinement. The electronic energy from a quantum chemical method is used to calculate the restraints ($T_{restraints}$) in quantum refinement. In order to achieve this, Q|R interfaces with ASE v3.8.1 (Bahn & Jacobsen, 2002) to enable easy access to many quantum chemical calculators. These calculators are thin wrappers around major quantum chemical codes. We generated a custom ASE calculator for Terachem v1.5 (Ufimtsev & Martinez, 2009), and modified some existing ASE calculators, and they are all available on https://github.com/qrefine/qr-plugin-ase. In this study we have chosen to investigate three different quantum methods: semi-empirical (PM7; Stewart, 2013) in Mopac v2016 (Stewart, 2016), *ab initio* (HF/6-31G-D3; Grimme *et al.*, 2010) using Terachem v1.5, and a Density Functional (RI-BP86/SV(P); Von Arnim & Ahlrichs, 1998; Becke, 1988; Schäfer *et al.*, 1992) from Turbomole v7.0.1 (Ahlrichs, 2015). The choice of individual quantum methods was arbitrary at this point, because the goal of this present study was to validate quantum-based refinement, not to carry out a systematic survey of candidate methods. The three different quantum chemical approaches chosen here are vastly different methodologies.

The relative weight, denoted as w in Eqn. 1, is initially taken as the ratio of the gradient norm of the restraint and data terms. This weight is scaled up or down using a heuristic approach based on crystallographic statistics such as R_{work}, R_{free} and R_{free}-R_{work} and geometric descriptors (Afonine *et al.* 2011).

To validate the approach and its implementation, the following test was carried out. A short 13 amino acid well-ordered and resolved helix was taken as a reference from the X-ray structure of aldose reductase (PDB code 1US0) refined at 0.66Å resolution. A helix was extracted from the structure, and then all of the side chains were removed to form a poly

glycine reference model, see Figure 2. Since this is a very high quality structure derived from high resolution data, the geometry of this helix is likely to be very close to representing the reality. This helix was then placed into a 16*18*30Å P1 unit cell box, which should be sufficiently large enough to have only minimal intermolecular interactions. A low resolution and highly incomplete set of structure factors describing all reflections in the 4-6 Å range was calculated from this model. After adding 5% of random noise to the amplitudes of these structure factors, we refer to this set as experimental data F_{obs}.

As starting coordinates for quantum refinement, five sets of perturbed structures of increasing magnitude (range 0.3 – 1.5 Å) were constructed by running molecular dynamics simulations, starting from the original model. The simulation used a simplified potential (Eqn. 2) from *phenix.dynamics*. This potential does not include an explicit hydrogen bonding term and is therefore cannot maintain the hydrogen bonding interactions during the simulation, see Figure 3. This diverse set of structures obtained from the different perturbation strengths should provide insight into the behavior of the quantum-based refinement. The structures were considered to be within typical convergence radius of refinement. To test the robustness of our implementation, each degree of perturbation was repeated 10 times using different snapshots sampled from the molecular dynamics simulations. The original model (prior to perturbations) is taken as the reference structure in all-subsequent analyses. All data presented in this work, including scripts to reproduce reported statistics, figures and plots, are available at https://github.com/qrefine/qr-tests-1us0.

3. Results and Discussion

In order to validate our approach and exercise the implementation (such as eliminate bugs, optimize runtime performance and find about convergence radius) we choose to work with a semi-artificial system described in the methods section. The advantage of working with such a system is two-fold. First, it is small and therefore allows sampling diverse refinement scenarios and parameters in a manageable amount of time (minutes to hours on a typical laptop and not days or weeks of computer time). This is extremely important during the development stage of a project as this allows a quick turnaround, which in turn promotes a continuous development process. Second, since we construct this system (as opposed to using real experimental data) we have full control over all its properties, and, most importantly, we know what the expected answer is. This development model has been used for more than a decade during development of CCTBX and many *Phenix* tools, including writing from scratch its refinement engine *phenix.refine* and has proven to be very efficient. Here we adopt this paradigm for development of our Q|R code.

These perturbed models described in the methods section above were subjected to Q|R refinement or standard CCTBX based refinement using the calculated scattering data. Since the starting model is known and data are calculated from it, it is trivial to score the refinement outcomes against a known answer and compare the scores between different refinement approaches, namely standard and quantum.

It is clear from Figures 3 that perturbed models become increasingly further removed from the reference model as the perturbation strength increases. Figure 3 shows a lengthening of the chain owing to a greater loss of critical hydrogen bonds with increasing degree of perturbation. At 0.3 Å RMS deviation, most hydrogen bonds are retained whereas by 1.5 Å most are lost. Hence, the challenge for refinement becomes greater as the perturbation strength increases. This gives a well-controlled set of models that can challenge quantum-based and standard refinement methods. In the set of smallest perturbations (0.3Å RMSD) the majority of hydrogen bonds are retained, while in the set of most heavily perturbed structures (1.5Å RMSD) almost all hydrogen bonds are destroyed. Refinement is expected to return the structure back to the original reference model, but this task becomes more challenging as the perturbation gets stronger.

We refined each and every one of the 50 perturbed structures with the quantum-based methods semi-empirical (PM7), *ab initio* (HF/6-31G-D3*),* density functional theory (RI-BP86/SV(P))*,* and standard (CCTBX), and the results are displayed in Figures 4 and 5.

Figure 4 shows crystallographic R-factors: R_{work} and R_{free}, and the gap R_{free}-R_{work} respectively. Since we introduced 0.05 error into calculated Fobs, the expected R_{work} for converged refinement is 0.05, which corresponds to the refined structure perfectly matching the structure of the known answer. It is desirable that R_{free} stays close to R_{work}, indicating less overfitting. From Figure 4 we observed CCTBX based refinements using standard restraints both R_{work} and R_{free}-R_{work} gap are marginally higher across all perturbation sizes when compared to the quantum-based refinements. The lower R_{work} and R_{free}-R_{work} gap the better the fit, therefore quantum-based refinements are outperforming standard refinement for this model system.

In addition to R factors, the number of hydrogen bonds recovered from the perturbed starting points is also used to check the quality of the refined helix structure. The range of valid hydrogen bond lengths was considered to be between 1.7 and 2.2 Å; bonds outside this range were considered *distorted*. Hydrogen bond distances in the helix extracted from 1US0 model range from 1.8 to 2.1 Å, see Figure 2. We can clearly see in Figure 5 that the refinements using the quantum-based restraints are recovering more of the hydrogen bonds than the refinements that employed standard restraints. Expectedly, the geometry of the helix cannot recover the perturbed hydrogen bonds during refinement with standard restraints, because the restraints do not contain relevant information. This can be understood as the refinement does not contain any explicit hydrogen bonding term, or even electrostatic interactions, which are dominant in hydrogen bonding.

4. Conclusions

The Q|R project is focused on developing software and methods for refining bio-macromolecules using chemical restraints derived from quantum mechanics. Q|R is currently under active development by researchers at Shanghai University together with *Phenix* developers. We have detailed our initial development implementation built on open source components, which we consider as a solid starting point. In addition, we have shown a validation example where the quantum-based refinement was able to recover more of the disrupted hydrogen-bonded network in a model system, providing a glimpse of what quantum refinement can provide in the future.

Previous attempts to develop software for quantum-based refinement have been made. A *Phenix* plugin for their linear-scaling semi-empirical DivCon code was developed by QuantumBio (www.quantumbioinc.com). Prior to this, the ComQum code was developed to locally improve a crystal structure using hybrid QM/MM methods. The development of the Q|R code is different from these two codes for three main reasons. Firstly, we have a multi-disciplinary team of developers from bio-crystallography and quantum chemistry working together. Secondly, we see Q|R as being a stable bridge between the well-established large quantum chemical code bases and the open source bio-crystallographic refinement tools available, e.g. in the CCTBX library. Therefore, we are strictly adhering to best practices in software development for long-term sustainability. Thirdly, we are focused on developing a high quality code base using an open source model, and are welcoming new contributors.

It is well known that QM calculations require significant computational resources, and therefore issues related to scalability will need to be addressed in future work. Further challenges also await us, such as crystallographic symmetry and static disorder to name but a few. Solving these challenges will require significant teamwork, sustained over a long period of time, to overcome these scientific and technical challenges. Quantum refinement has the potential of becoming a standard technique for assisting structural biologists in obtaining high-quality structures.

Acknowledgments

MZ and MPW would like to acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG) for funding from the SFB858 project. MPW would like to acknowledge support from the Shanghai Eastern Scholar Program. PVA thanks the NIH (grant GM063210) and the *Phenix* Industrial Consortium. PyMol (DeLano, 2002) was used for molecular graphics. ICQMS is gratefully acknowledged for support.

References

P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart. Acta Cryst. D66, 213-221 (2010).

Afonine, P.V., Echols N., Grosse-Kunstleve, R., Moriarty N.W., Adams, P.D. (2011) Computational Crystallography Newsletter 2, 99–103.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D. Acta Cryst. D68, 352-367 (2012).

Afonine, P.V., Urzhumtsev, A., Adams, P.D. Macromolecular crystallographic structure refinement. in Arbor, Celebrating 100 years of modern crystallography 2015, 191:a219.

Ahlrichs, Reinhar, Turbomole V7.0, 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com

Bai, X., McMullan, G. & Scheres, S. (2015). Cell. 40(1), 49–57.

Bahn, S. R. & Jacobsen, K. W. (2002). Computing in Science & Engineering,4(3), 56-66. More info available from: https://wiki.fysik.dtu.dk/ase/

Becke, A. D. (1988). Physical review A, 38(6), 3098.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). Nucleic Acids Res. 28, 235–242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). J. Mol. Biol. 112, 535–542.

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2016). BUSTER. Cambridge, United Kingdom: Global Phasing Ltd.

Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, F., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst*. **54**, 905–921.

Borbulevych, O., et al. (2016). Acta Crystallographica Section D-Structural Biology 72: 586-598.

Borbulevych, O. Y., et al. (2014). Acta Crystallographica Section D-Biological Crystallography 70: 1233-1247.

Byrd, Richard H., et al. (1995). SIAM Journal on Scientific Computing 16(5) : 1190-1208.

Canfield, P., et al. (2006). Journal of Chemical Physics 124(2): 15.

Carlsen, M. & Røgen, P. (2015) Proteins, 83, 1616-1624.

Cheng, Y. (2015). Cell. 161(3), 450-457.

DeLano, W.L. (2002). *The PyMOL Molecular Graphics System,* DeLano Scientific, San Carlos, CA, USA. *http://www.pymol.org*

Dixon, S. L. & Merz Jr, K. M. (1996). The Journal of chemical physics, 104(17), 6643-6649.

Echols, N., Headd, J.J., Afonine, P. & Adams, P. (2010; July). Computational Crystallography Newsletter. https://www.phenix-online.org/newsletter/ .

Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State. Oxford University Press.

Fu, Z., et al. (2013). Journal of Chemical Theory and Computation 9(3): 1686-1693.

Grimme, S., Antony, J., Ehrlich, S., & Krieg, H. (2010). The Journal of chemical physics, 132(15), 154104.

Grosse-Kunstleve, R. W. & Adams, P. D. (2002). J. Appl. Cryst. 35, 477-480.

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. (2002). J. Appl. Cryst. 35, 126-136.

Goerigk, L. and J. R. Reimers (2013). Journal of Chemical Theory and Computation 9(7): 3240-3251.

Goerigk, L., Collyer, C. A. & Reimers, J. R. (2014). The Journal of Physical Chemistry B, 118(50), 14612-14626.

Headd, J.J., Echols, N., Afonine, P.V., Grosse-Kunstleve, R.W., Chen, V.B., Moriarty, N.W., Richardson, D.C., Richardson, J.S. & Adams, P.D. Acta Cryst. D68, 381-390 (2012).

Korth, M. & Thiel, W. (2011). Journal of Chemical Theory and Computation, 7(9), 2929-2936.

Liu, D. C. & Nocedal, J. (1989). Math. Program. 45, 503–528.

Li, X., et al. (2012). Journal of Computational Chemistry 33(3): 301-310.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367.

Nilsson, K., et al. (2004). Biophysical Journal 87(5): 3437-3447.

Oldfield, T.J. (2000). Acta Cryst. (2001). D57, 82-94.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V. (1963). J. Mol. Biol. 7, 95-99.

Ryde, U. (1996). Journal of computer-aided molecular design 10(2): 153-164.

Ryde, U. (2003). Current Opinion in Chemical Biology 7(1): 136-142.

Ryde, U. and K. Nilsson (2003). Journal of the American Chemical Society 125(47): 14232-14233.

Ryde, U. and K. Nilsson (2003). Journal of Molecular Structure-Theochem 632: 259-275.

Rupp, B. (2010). Biomolecular crystallography: principles, practice, and application to structural biology. New York: Garland Science.

Schäfer, A., Horn, H., & Ahlrichs, R. (1992). The Journal of Chemical Physics,97(4), 2571-2577.

Sheldrick, G.M. (2008). Acta Cryst. A64, 112-122.

Sobolev, O., Afonine, P.V., Adams, P.D. & Urzhumtsev, A. (2015). J. Appl. Cryst. 48, 1130-1141.

Stewart, J. J. (2013). Journal of molecular modeling, 19(1), 1-32.

Stewart, J. J., MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016).

Tronrud, D.E. (2004). Acta Cryst. (2004). D60, 2156-2168.

Ufimtsev, I. S. & Martinez, T. J. (2009). Journal of Chemical Theory and Computation, 5(10), 2619-2628.

Watkin, D. (2008). J. Appl. Cryst. (2008). 41, 491-522.

Yu, N., et al. (2006). Journal of Chemical Theory and Computation 2(4): 1057-1069.

Yu, N., et al. (2006). Protein Science 15(12): 2773-2784.

Yu, N., et al. (2005). Acta Crystallographica Section D-Biological Crystallography 61: 322-332.

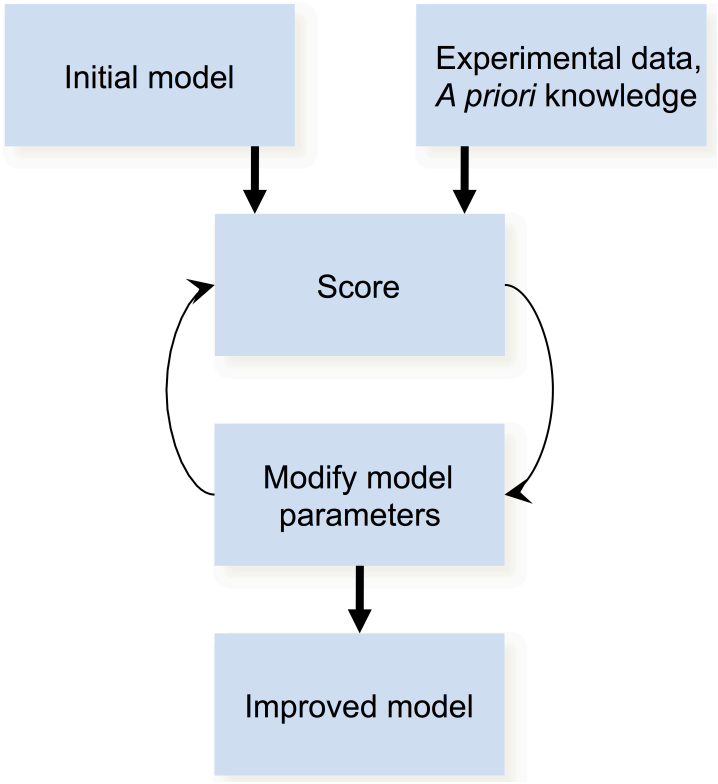

Figure 1. A general model refinement workflow.

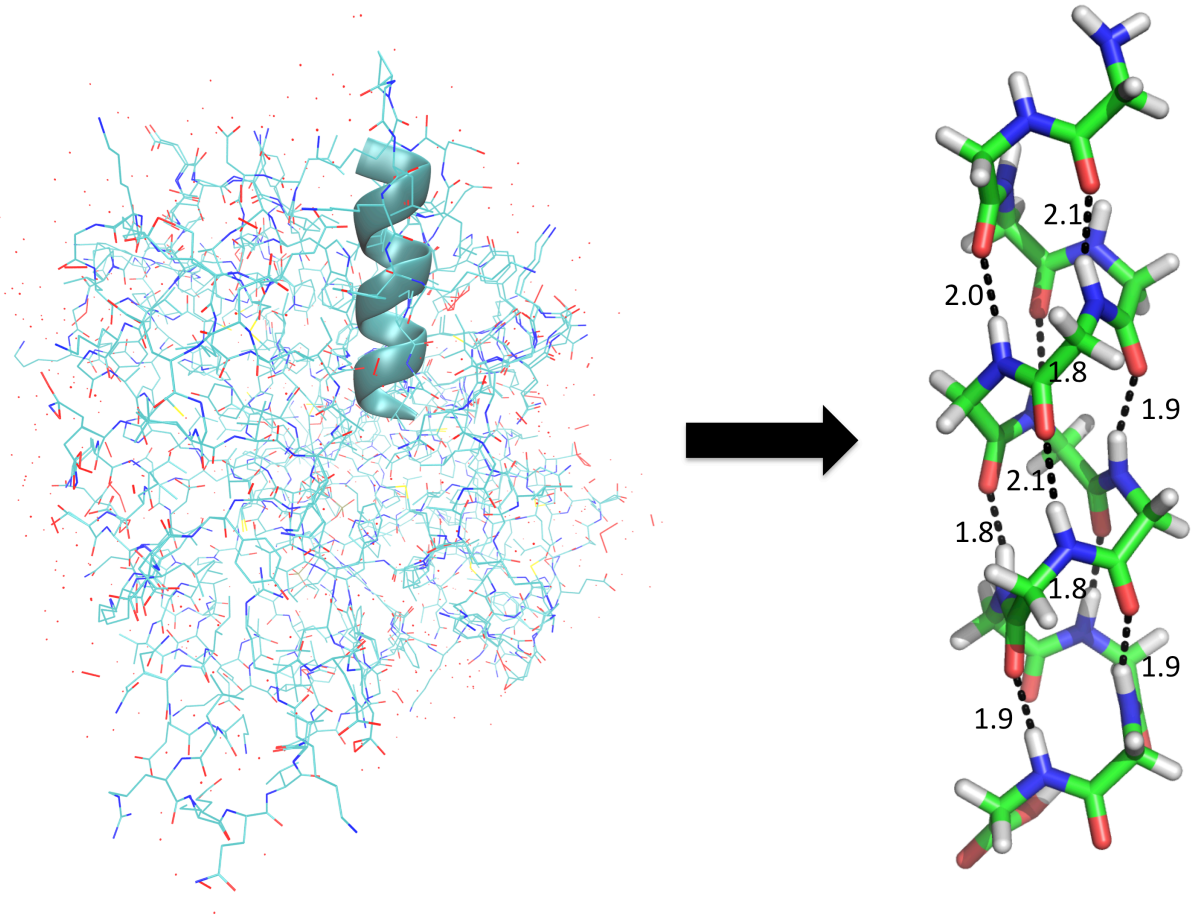

Figure 2. Aldose reductase PDB structure (left) and extracted helix model (right) with hydrogen bond distances shown in Å.

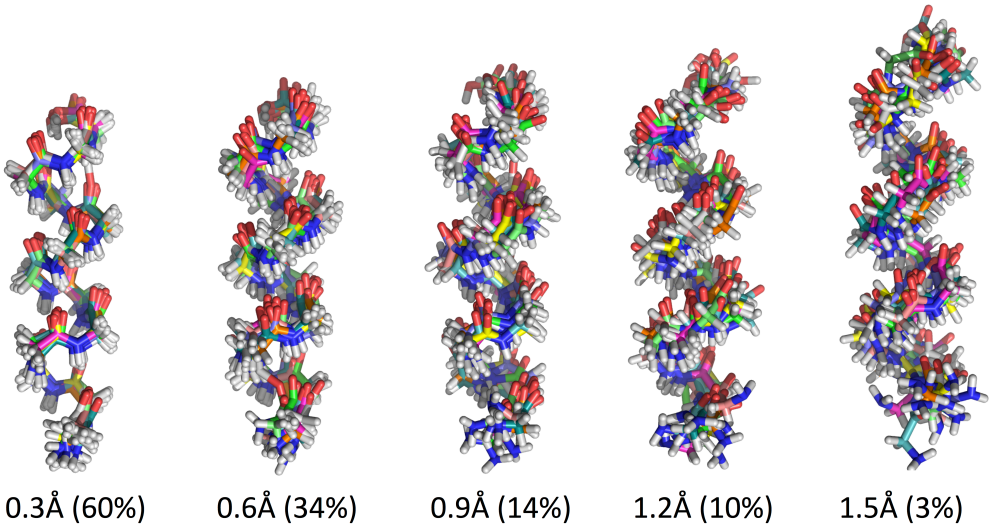

Figure 3. Perturbed models with RMS deviation from the starting model of 0.3, 0.6, 0.9, 1.2 and 1.5Å, overlaying 10 models per perturbation. The average percentage of conserved H-bonds at each perturbation level is shown in brackets.

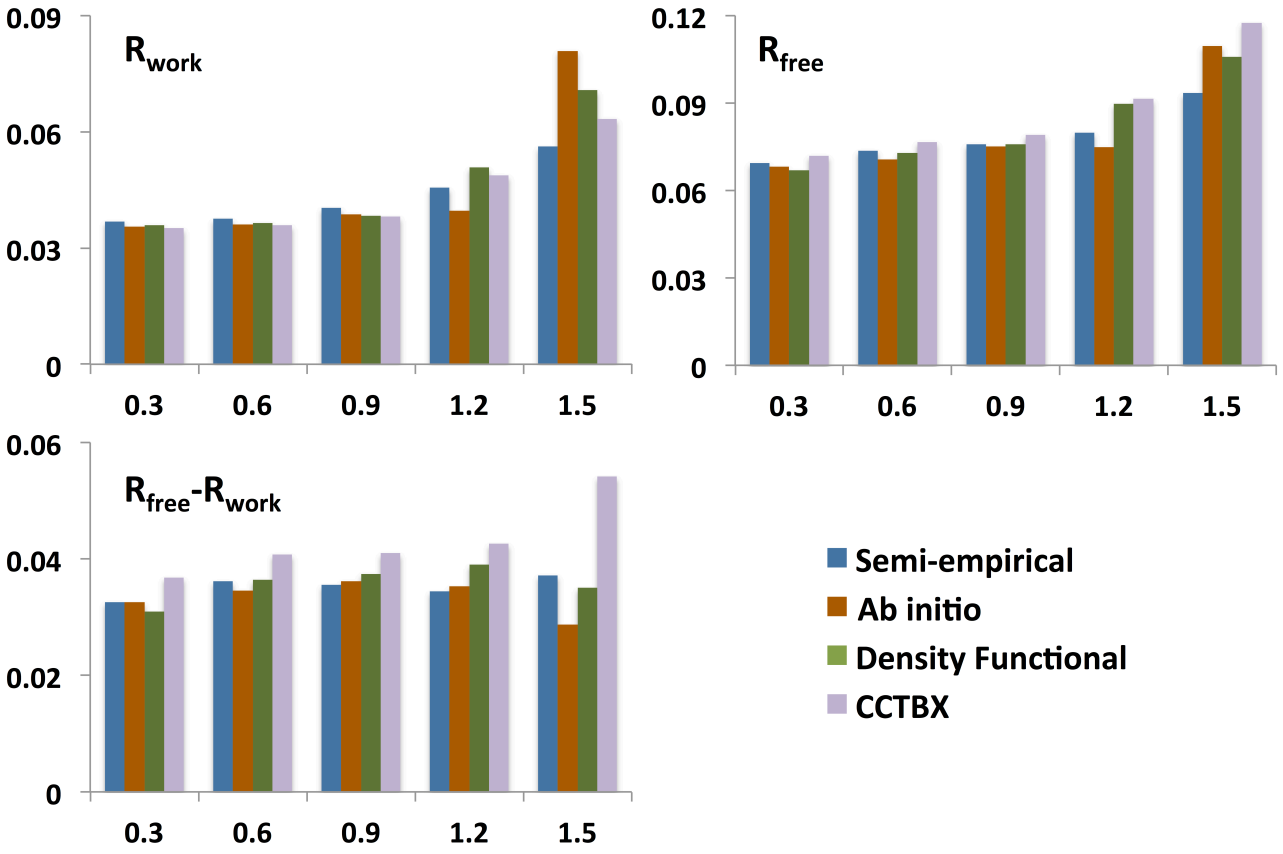

Figure 4. Average R_{free} (top, left), R_{work} (top, right) and R_{free}-R_{work} (bottom) as a function of perturbation strength (Å) for semi-empirical (PM7), *ab initio* (HF/6-31G-D3), density functional (RI-BP86/SV(P)) and standard refinement (CCTBX). Average (10 trials per perturbation) starting R_{work} are 0.15, 0.27, 0.35, 0.44 and 0.55 correspondingly for each perturbation dose from 0.3 to 1.5 Å. Random noise (5%) was added to F_{obs}, therefore R is expected to be around 0.05 which would correspond to the ideal structure.

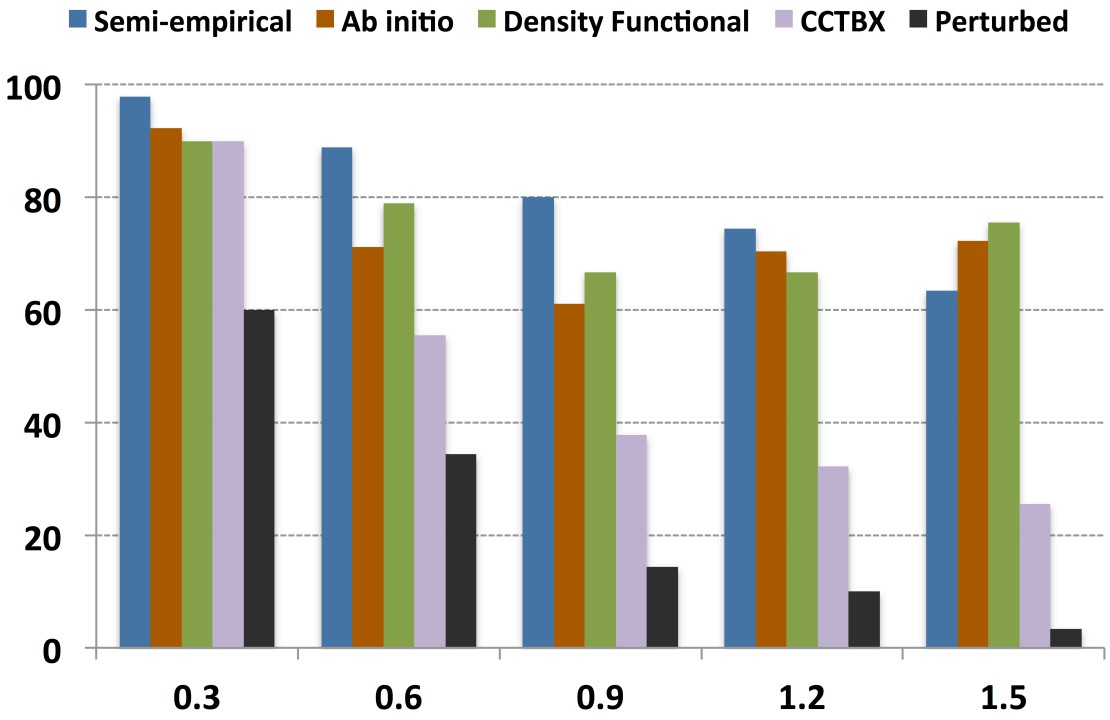

Figure 5. The average percentage of recovered H-bonds as a function of perturbation strength (Å) after refinement using either semi-empirical (PM7), *ab initio* (HF/6-31G-D3), density functional (RI-BP86/SV(P)), or standard refinement (CCTBX), and the percentage of hydrogen bonds that remained in the perturbed models.