DEN-assisted refinement at low resolution (lower than ~ 3.5 A resolution)

In this section we will discuss the use of deformable elastic network (DEN) restraints to assist refinement at low resolution. First, please go through the previous section on low resolution refinement.

Reference

G.F. Schroeder, M. Levitt, and A.T. Brunger, Super-resolution biomolecular crystallography with low-resolution data, Nature 464, 1218-1222 (2010).

Generation of the reference model

The reference model can be a homology model or a composite model consisting of several homology models. The homology models for individual domains or molecules do not have to be contiguous, and they can be placed in arbitrary orientations and positions as long as the inter-domain distances are excluded from the DEN generation (see below). To generate such reference models, the program Modeller can be used. It is recommended to follow the "basic" example in the Modeller tutorial. This tutorial describes the search for a suitable template among all known PDB structures using sequence alignment techniques, followed by a structure-specific alignment for the selected template, and then the homology model building itself using this alignment with the template structure.

There are alternative methods for structure-based alignments, especially for difficult cases. Of particular interest is the PROMALS3D method, which has a web submission site that is very easy to use. The resulting alignment can then be used for homology building using Modeller.

If the structure consists of a multi-domain complex, you can create a composite reference model containing all (sub)domains that can be obtained by homology modeling or that are available at higher resolution. You can simply append the individual models to generate the composite reference model. If you use a upper sequence separation cutoff (the default value is 10), you can place the individual (sub-)domains in arbitrary positions and orientations since inter-domain distances will be generally ignored with this sequence separation cutoff (see also the Online Methods section in the above reference). If more specific selections of DEN distances (or exclusion of all inter-domain distances) are desired use the individual DEN groups as exemplified in the example below.

Note that the reference model must use the same chainIDs (or SEGIds), RESIDs, and residue names as the model that you're refining (the starting model for refinement). However, it is OK if the reference model is incomplete (i.e., there are fewer atoms in the reference model compared to the starting model). Also, it is OK if the reference model has more atoms than the starting model (for example if the starting model uses poly-ala for some of the regions of a protein). Such missing or additional atoms will produce informative message in the CNS output file. Carefully check these messages and the *.den files(s) that are created by the particular refinement job to make sure that this is the behavior that is desired. In any case, the refinement job should run without problems even if these messages are present.

DEN refinement example

Diffraction data: full-length p97 in complex with ADP at 4.25 A resolution.

Initial structure: consists of the first refined structure of full-length p97 (PDB ID 1YQI, see figures below). As we found out later this structure had register shift errors in the D2 ADP binding domain. These shifts exists in all three protomers.

Reference structure: was generated with Modeller using the high resolution crystal structure of the N-D1 domains of p97 (PDB ID 1E32), and a homology model for the D2 domain based on the structure of the D1 domain. Note that the high resolution structure of the D2, PDB ID 3CF0, was not used in this case in order to test the power of the method if only limited high-resolution information is available; the all-atom RMSD between our homology model of the D2 domain and the 3CF0 structure is 3.1 A. Note that the orientation and position of the reference structure is arbitrary since the distance restraints are invariant with respect to rotations and translations.

Target structure: is the re-refined structure of full-length p97 in complex with ADP using the high-resolution crystal structures of the N-D1 and D2 domains as starting point (PDB ID 3CF3).

References:

- J.M.Davies, A.T.Brunger, W.I.Weis, Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. Structure 16, 715-726, 2008.
- B.DeLaBarre and A.T. Brunger, Nucleotide dependent motion and mechanism of action of p97/VCP, J. Mol. Biol. 347, 437-452 (2005).
- B. DeLaBarre and A.T. Brunger, Complete Structure of p97/Valosin-Containing Protein Reveals Communication Between Nucleotide Domains, Nature Structural Biology 10, 856-863 (2003).

The particular example we discuss in this section is (the refine_den.inp file is taken from the refine.inp input file, i.e., the only differences are the particular input parameter settings):

        cns_solve < refine_den.inp > refine_den.out

Compared to the example discussed in the previous section, the following changes are made

1. A different initial model is read that has register shift errors in the D2 ADP binding domain.
2. DEN restraints were obtained from reference structure. In general, it is important that this reference model refers to the sequence numbers in the coordinates that are read by the CNS refinement job. The reference model does not have to be complete - there can be atoms that are undefined compared to the coordinate set that is being refined. Furthermore, the positions and orientations of distinct domains or fragments can be arbitrary as long as at least one of the following conditions is true:
    - they have different chain identifiers,
    - they are selected as distinct groups in the DEN group statements (see below),
    - they are separated by more than 10 residues and the DEN sequence separation limit is set to 10 (see below).

   Alternatively, the particular inter-domain or inter-molecule interactions can be excluded by using the DEN atom selection statements.

Explanation of some of the DEN and refinement parameters:

1. A total of 10 macrocycles are performed in this example. During the first cycle, the DEN equilibrium distances are set to the current refined model. This is accomplished by setting the initial DEN minimum to the starting coordinates. Furthermore, the parameter relax_ncycle is set to 3 which allows the model to undergo initial refinement without influence by the reference model. This setting enables one to use a relatively crude starting model that can be different from the reference model. During the subsequent

cycles, the DEN reference restraints are slowly introduced, therefore preventing numerical instabilities if they were introduced right at the beginning.

2. Each macrocycle consists of slow-cooling torsion angle dynamics refinement and restrained grouped B-factor refinement. The starting temperature of the slow-cooling simulated annealing refinement is set to 3000 K. This starting temperature should be lowered (to 300-1000 K) if the model has already undergone refinement and is presumably close to the true structure.

3. During the last 2 of the 10 macrocycles, the DEN restraints are turned off. This allows one to asses if the refined structure is in a minimum that persists independent of the DEN restraints.

4. Note that in this example the van der Waals radii are reduced by 75% during all macrocycles except the last two cycles (parameter full_vdw_cycles).This reduction in van der Waals parameters sometimes improves convergence of the refinement.

5. The DEN restraints are automatically generated from a specified subset of interatomic distances in the reference model using a distance (d) cutoff range of $3 < d < 15$, an upper sequence separation limit of 10 (within the same ChainID/SEGId), a lower sequence separation limit of 0 (within the same ChainID/SEGId), and selection of two distinct subgroups in the reference model, residues 1:458 and 488:699, corresponding to the N-D1 and D2 domains of p97, respectively). With a "ndistance ratio" of 1, on average, one distance per atom is randomly selected out of this subset of interatomic distances, resulting in a total of ("ndistance ratio" * <number of atoms>) DEN distances. To include all distances within a given chain (as specified by a unique ChainID/SEGId) as possible candidates for the random selection, set the upper sequence separation limit to very large number (e.g., 999999). To include all distances between all atoms (regardless of ChainID/SEGId) set both the lower and upper sequence separation limit to -1.

6. The DEN gamma value (den_gamma) is set to 0.5 and DEN weight wden (den_scale) is set to 100. See below for optimization of these values.

7. The exponent of the DEN restraint term is set to 2. Possible choices are 2 and 4. Better numerical stability is achieved with an exponent of 2 albeit with slightly less rapid convergence, possibly requiring more refinement cycles for convergence.

In the particular example (refine_den.inp) the DEN restraints were generated "on-the-fly". Different random number seeds will result in different random selections of DEN inter-atomic distances. It is important to check that this automatic generation produced the desired distances by manually inspecting the file refine_den.den that is produced by the refinement job.

Instead of the automatic generation of DEN distances within the refinement job, a file containing DEN restraints can be read. Such a DEN restraints file can be generated with the CNS task file den_make.inp (located in the General section of the CNS Input Files). This task file allows one to keep the same DEN restraint file for all refinements that use DEN.

A note for structures that are already refined

Note, that the periodic DEN updates that allow the DEN network to be deformed are only performed in conjunction with slow-cooling torsion angle dynamics in the refine.inp file. If this type of annealing schedule is turned off, then the DEN restraints will be fully active during positional refinement and will not undergo any deformations during refinement. If extensive annealing is not desired, lower the starting temperature of the slow-cooling schedule from 3000 K to 300 K. In addition, the number of macrocycles should be reduced, along with setting the number of initial refinement cycles and number of final refinement cycles (both options in the DEN section) to smaller number or zero.

Weights

Atom-specific weighting schemes of DEN restraints are not yet implemented in the CNS task files although it

is possible for an advanced used to modify the appropriate NOE statements in the refine.inp and den_make.inp task files to introduce different NOE classes that would enable different weights for each class.

NCS restraints

As in the previous section, this example uses NCS restraints (see the p97_adp_ncs.def file) and non-standard (ADP) ligands (see the p97_adp_cns.top and p97_adp_cns.par files).

Additional restraints

An additional restraints file (e.g., dna-rna_restraints.def) can be provided as one of the parameters in the refine.inp task file. However, any distance information that is included in such an additional restraints file is overwritten by the DEN restraints. Thus, any such distances should be provided as part of a customized den_make.den file.

Refinement repeats

It is generally advisable to perform many repeats of the annealing refinement with different seeds that determine the initial random velocities for the molecular dynamics simulations and the seed for the DEN distance selections, as well as different den_gamma (gamma) and den_scale (wden) values. This can be easily done by setting up a C Shell script file that modifies the CNS refine_den.inp task file. A particular example is:

    ./refine.csh 55 0.6 100

This csh script changes the seed in the refine_den.inp file to 55 (first parameter), the den_gamma value to 0.6 (second parameter), the den_scale value to 100 (third parameter), the output_root to "refine_den_55_0.6_100", saves the modified file to refine_den_55_0.6_100.inp, and then runs the job. The script then deletes the large ".map" files, but keeps the ".hkl" map coefficient files, so electron density files can be generated (or directly displayed using the ".hkl" files using COOT). The script also compresses the output file to reduce disk space usage.

Multiple such csh scripts can be run to set up multiple jobs. This is particularly useful in a distributed computing environment (see below). Generally, the coordinates with the lowest Rfree should be used for further analysis. As control, a few jobs with den_scale set to zero should be run - these runs correspond to refinement without DEN restraints.

Determination of the best parameter pair combination by sorting with respect to Rfree

The refined coordinates are then sorted with respect to Rfree. We use a Perl script to do this:

    ./sort.pl refine_den_*.pdb > sort.list

This Perl script creates a new directory "sorted" that contains the final coordinates and the corresponding Fourier coefficient map files (*.hkl files) sorted by Rfree, i.e., sorted/sort_1.pdb is the best coordinate set, the one with the lowest Rfree, and sorted/sorted_1.hkl is the corresponding Fourier coefficient file. The script also produces a convenient summary:

sort.list

B-factor sharpening of electron density maps

B-factor sharpening of the electron density maps produced by `refine_den.inp` can be accomplished in several ways. First, the electron density map files (`refine_den*.map`) can be read by Coot and then B-factor sharpened using Calculate -> Map Sharpening. This method also works when reading the CNS structure factor file that contains the coefficients for the 2Fo-Fc and Fo-Fc maps (`refine_den.hkl`) into Coot using File -> Auto OpenMTZ. Second, the CNS structure factor file `refine_den.hkl` can be read by the CNS `fourier_map.inp` script and B-factor sharpening applied upon computing electron density maps (the coefficients for the Fo-Fc map are in the array called F1, the coefficients for the 2Fo-Fc map are in the array F2).

Averaging electron density maps obtained from multiple independent refinements

The electron density maps resulting form the best structures (i.e., lowest Rfree structures) can be averaged by using the `model_map.inp` script file (see section [Refinement at high to medium resolution](#)). This type of averaging can be very powerful to improve model phases (L. M. Rice, P. D. Adams, Y. Shamoo, and A. T. Brunger, Phase improvement by multi-start simulated annealing refinement and structure factor averaging, J. Appl. Crystallography 31, 798-805, 1998). The `model_map.inp` script file also has an option for B-factor sharpening.

Setup for distributed (grid) computing

If you have [Xgrid](#) installed on your MAC OS X server, here is a script file for the [Gridstuffer](#) program to run a large number of DEN refinement jobs with different seeds, den_gamma and den_scale values:

[refine_xgrid_stuffer.txt](#)

which uses the unix shell file [refine.csh](#) which modifies the refine.inp file as described above and runs the CNS job. A similar approach can be used in any distributed environment, not just Xgrid.

Results of the DEN refinements

As mentioned above, the starting structure for this tutorial is the earlier model of the ADP-bound form of p97, PDB ID 1YQI. The target structure is the re-refined structure of p97 (PDB ID 3CF3). The following figure compares both structures around the D2 nucleotide binding site for one of the protomers:

Starting structure (1YQI, magenta) vs target structure (3CF3, blue)
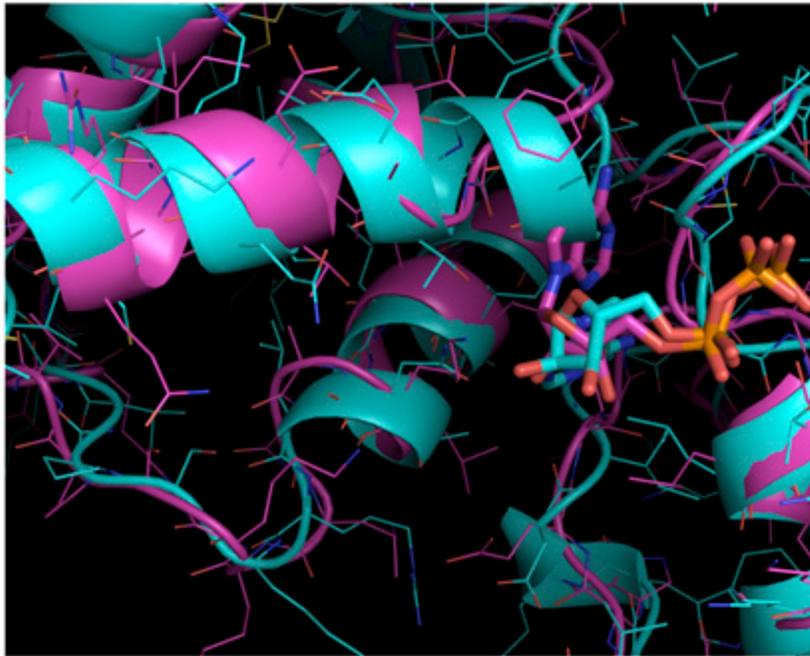

Note the large differences (register shifts) for some of the helices and the nucleotide conformation.

We performed 590 DEN-refinements using the protocol described above (and implemented in the file refine_den.inp): for each parameter pair (den_gamma x den_scale = [0,0.2,0.4,0.6,0.8,1.0] x [0, 3, 10, 30, 100, 300]) 20 refinements were performed with different random number seeds for the initial velocity assignments for simulated annealing. The structure with the minimum Rfree for each parameter pair was used for further analysis (we used the refine_xgrid_stuffer.txt file to run the jobs in our Xgrid environment). The following figure shows a contour plot of these minimum Rfree values. The contour plot shows that the best Rfree values are obtained for wden/gamma combinations in the upper left corner, with an absolute minimum at den_gamma (gamma)=0, den_scale (wden)=30.
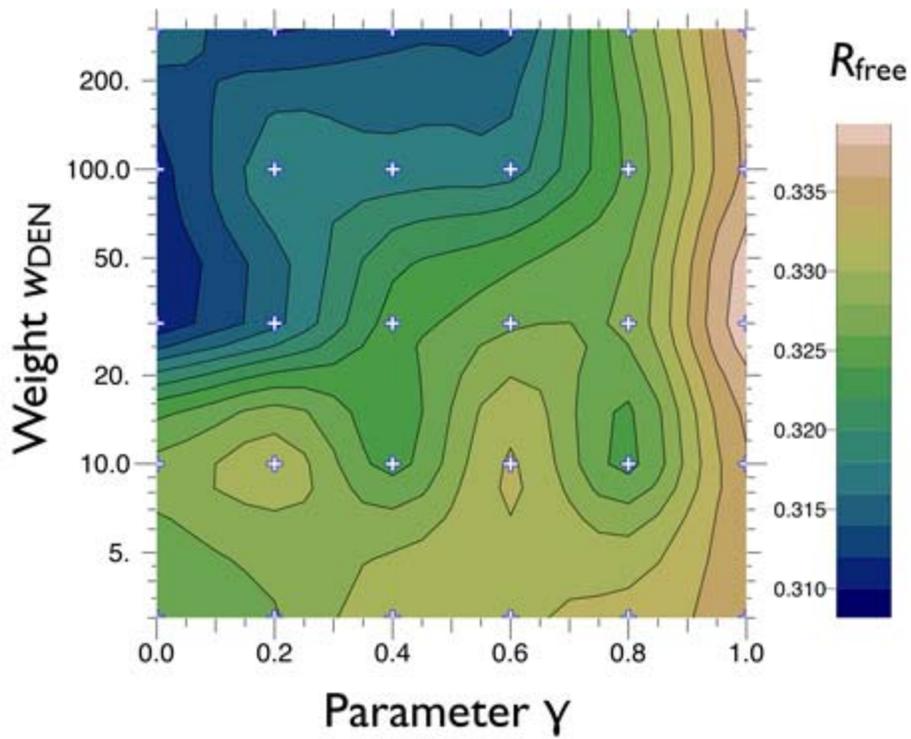
The following figure shows the all-atom RMSD between the DEN-refined structures and the target structure (PDB ID 3CF3):
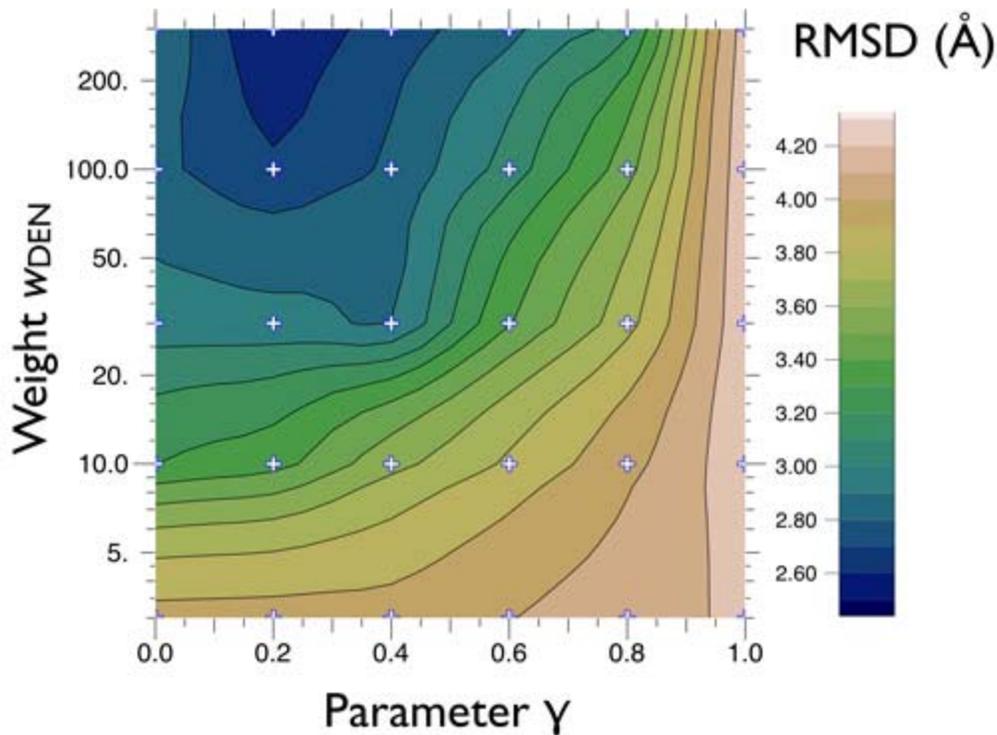
Note the similarity of the RMSD and Rfree contour plots. Structures with low Rfree produce a low RMSD to the target structure 3CF3.

DEN refinement accomplishes very large shifts resulting in an improvement of the RMSD to the target structure. The Rwork and Rfree values for the best DEN-refined structure (refine_den_19_0.0_30.0.pdb) are 0.274 and 0.308, respectively; the all-atom RMSD to the target structure (3CF3) is 2.91 A. The remaining differences are in certain loops of the D2 domain and a long protruding helix. In contrast, the all-atom RMSD for the best structure without DEN (refine_den_11_1.0_0.pdb) is significantly larger (4.32 A); the Rwork and Rfree values are 0.296 and 0.331, respectively.

The following figure compares the structures refined with and without DEN:

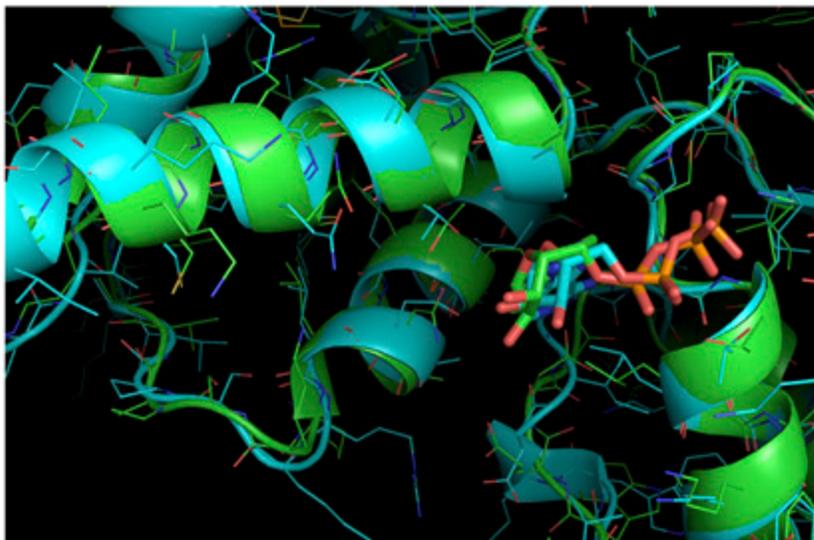# DEN-refinement (green) vs refinement without DEN (yellow)
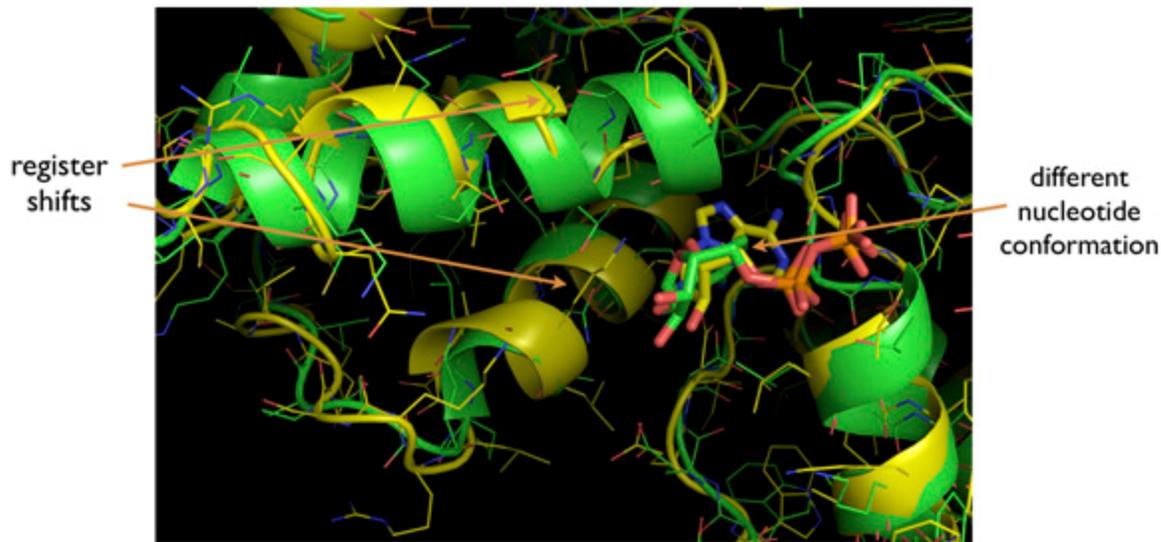
p97 D2 domain nucleotide binding site

Note the large differences for the helix conformations (register shifts) and the large difference in nucleotide conformation.

The next figure compares the DEN refined structure with the target structure.

DEN-refinement (green) vs target structure (blue), PDB ID 3CF3

p97 D2 domain nucleotide binding site

The helices around the nucleotide binding site are in the correct place and the nucleotide binding site is more typical for nucleotide binding proteins. This example demonstrates the very large radius of convergence of DEN-refinement.

It is not always necessary to perform as many trials as in this particular example. In fact, most of the DEN refinements with in the upper quadrant of the gamma, wden parameter space accomplish a significant improvement in R values and RMSD to the target structure. So, it would have been possible to obtain a reasonable DEN-refined structure with just one trial per parameter pair.

Script to run this tutorial

Back to tutorials   Previous section