Structure determination using poorly diffracting membrane protein crystals: the H^+ and Na^+,K^+-ATPase case history

Bjørn P. Pedersen, J. Preben Morth & Poul Nissen*

Centre for Membrane Pumps in Cells and Disease, Danish National Research Foundation, Aarhus University, Dept. Molecular Biology. Gustav Wieds Vej 10C, DK – 8000 Aarhus, Denmark

*to whom correspondence should be addressed, pn@mb.au.dk

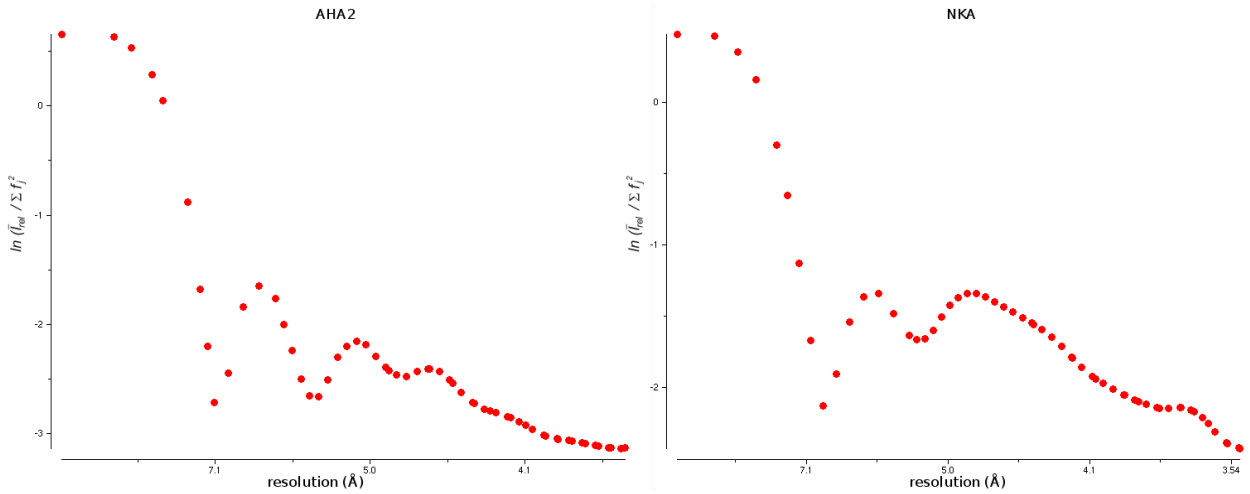
Supplementary Material

Supplementary Methods

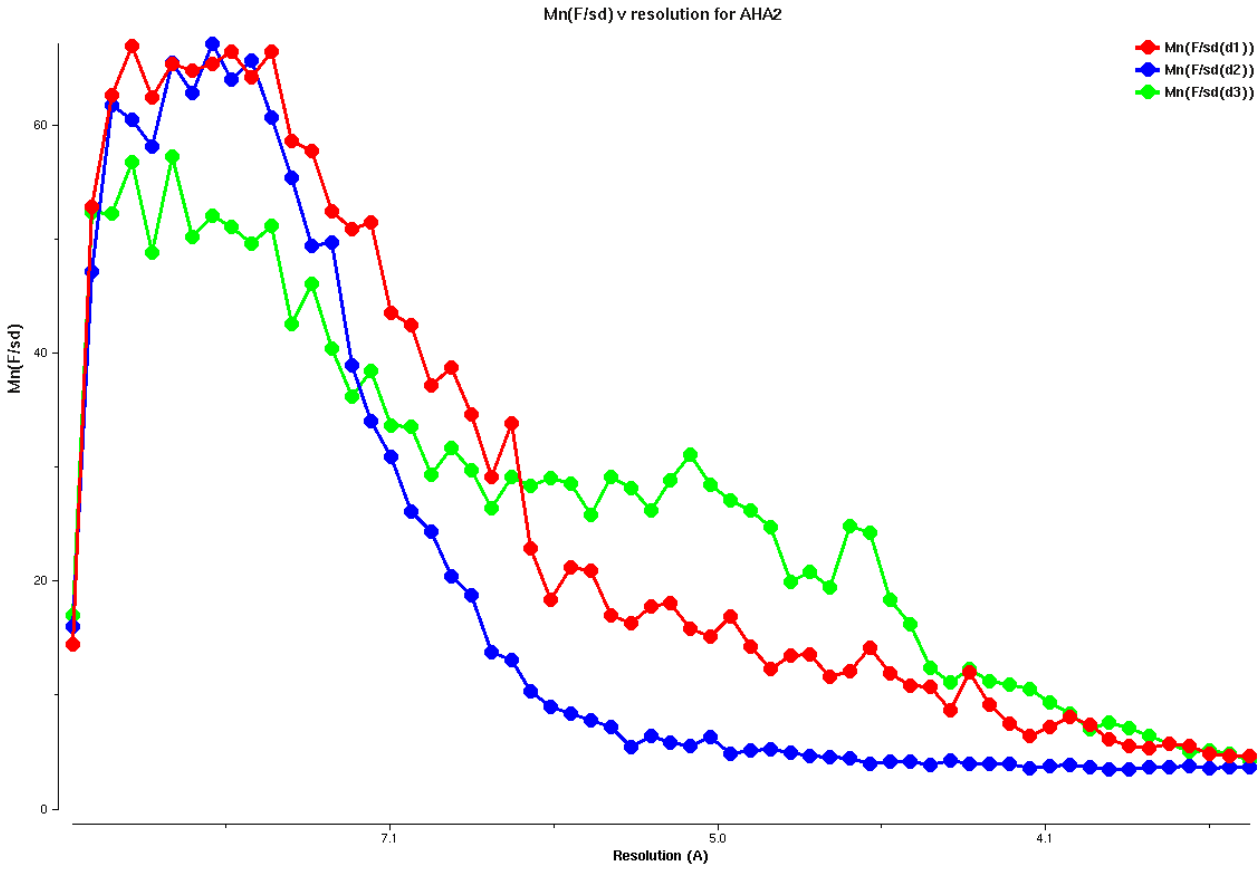
AHA2 is a single-chain P-ATPase enzyme of 92 kDa and was obtained by heterologous expression in *S. cerevisiae*. The yeast membrane fraction was solubilised by DDM and the enzyme purified by Ni^{2+} chelation chromatography using 0.15% DDM. 0.09 mM $C_{12}E_8$ was introduced as a second detergent by dialysis prior to crystallisation at pH 6.5 in hanging or sitting drops using approximately 30% PEG400 as the precipitant and 5-Cyclohexyl-1-pentyl-β-D-maltoside (CYMAL-5) as a third additive detergent at 2.4 mM. The protein sample crystallised thus contained no lipids, and protein monomers in crystals were shielded by detergent micelles. A dehydration protocol based on vapor diffusion procedures and with minimal disturbance of the fragile crystals improved weak diffraction properties to a (still severely anisotropic) resolution of 3.6/5 Å for native crystals, and 6-8 Å resolution for derivative crystals ($HoCl_3$, K_2PtCl_6 and $(Ta_6Br_{12})Br_2$) (Pedersen et al. 2007).

The structure solution of AHA2 followed the approach schematized in Supplementary Figure 4. All HA-phasing was done using *SHARP* (delaFortelle, E. & Bricogne, G. 1997). In the case of phase-combination before density modification (colunm 1 and 2 in Figure 2), this was done by combining the Holmium+Platinum MIRAS phases from *SHARP* with the Tantalum SIRAS phases from *SHARP* using *SIGMAA* with the 'COMBINE MIR2' option. In the optimal case the phases for the MIRAS and the SIRAS run were not combined but input separately into *DMMULTI* with Native 1 and Native 2 being input as unphased crystal forms (see Supplementary Figure 4). Thus the final phase combination leading to the electron density shown in column 3 in Figure 2 was done by real-space averaging over these four crystal forms followed by phase combination within *DMMULTI*.

NKA is a membrane protein complex consisting of a 112 kDa P-ATPase alpha-subunit, a glycosylated beta-subunit of approximately 40 kDa, and a regulatory gamma-subunit of 7 kDa. The complex was solubilised from purified kidney membranes using 35 mM octaethyleneglycol mono-n-dodecylether ($C_{12}E_8$). The supernatant from a subsequent centrifugation was used directly for crystallisation at pH 7.0 with n-Dodecyl-β-D-maltoside (β-DDM) at approx. 0.2% as a second detergent by vapour diffusion in hanging drops using 14% PEG3350 as the precipitant. The protein sample thus contained native lipids and ~1% (200xCMC) detergent, and crystals formed as stacked bilayers. The crystals were thin plates (approx. 700 x 200 x 15 µm) and very fragile, yet allowing for data collection at 3.5 Å resolution for native crystals and at 7-9 Å resolution for derivative crystals (platinum(II) tertpyridine chloride and $(Ta_6Br_{12})Br_2$) (Morth et al. 2007).

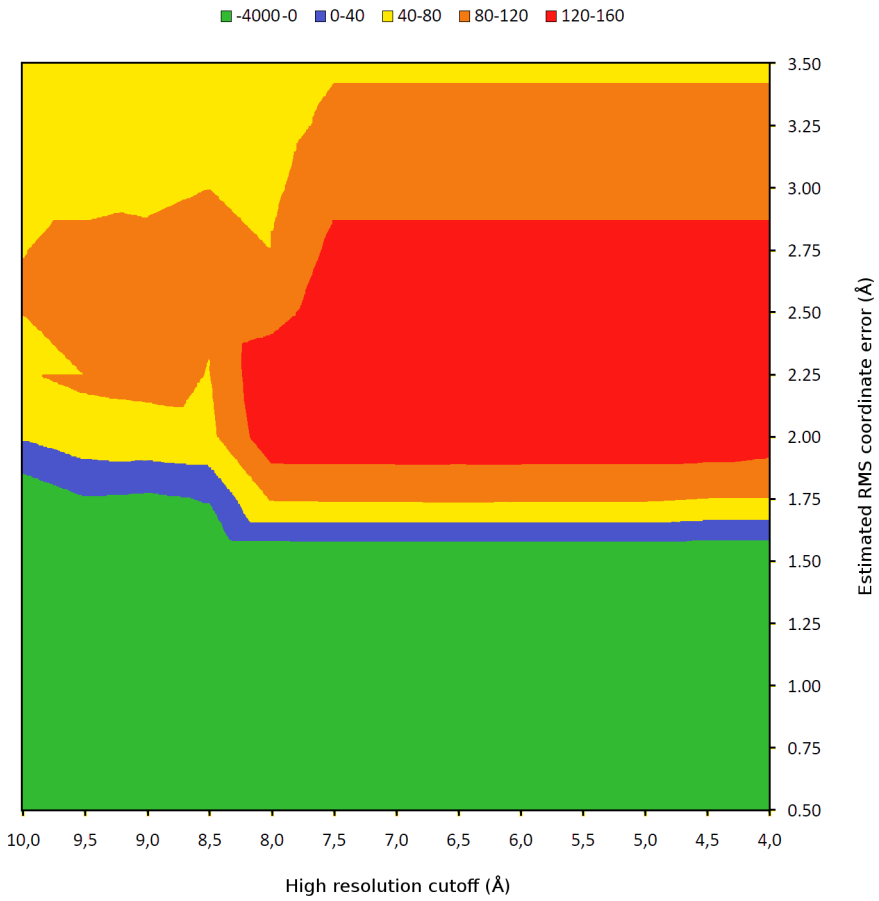


Supplementary Figure 1. Wilson plot of AHA2 and NKA data used.



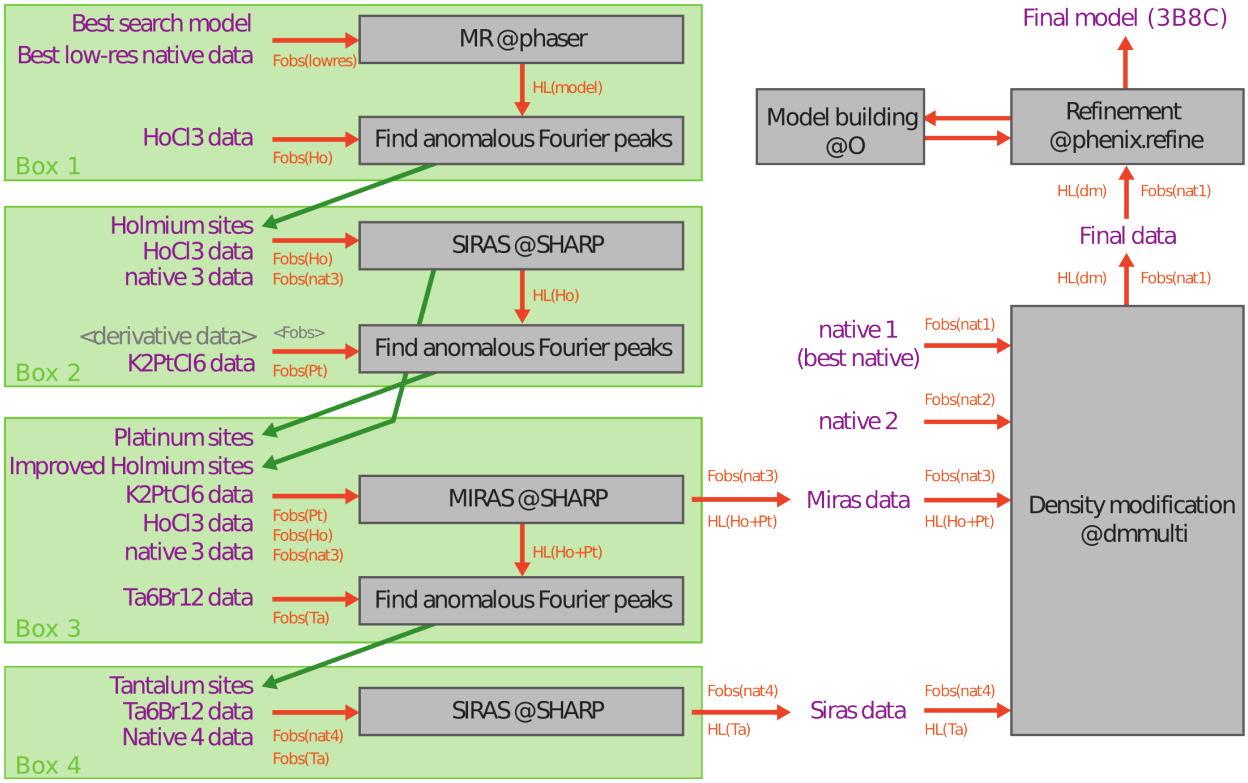
Supplementary Figure 2. Anisotropy plot of AHA2.

The signal to noise ration along the three principal axes clearly demonstrate the anistropic properties of the AHA2 data.



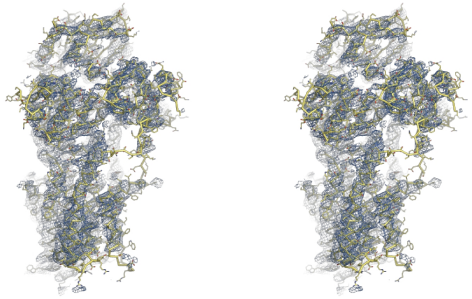
Supplementary Figure 3. Log likelihood gain score of molecular replacement as a function of estimated r.m.s. coordinate error and high resolution cutoff of AHA2.

The final log likelihood gain score of the top-scoring solution from each run is show. Note that underestimating the r.m.s. will lead to negative LLG-scores even for correct solutions. Conversely, a negative LLG-score is a good indicator that the r.m.s. should be increased in future searches.

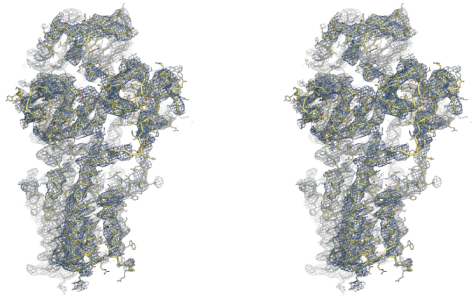


Supplementary Figure 4. Flowchart for the AHA2 structure solution.

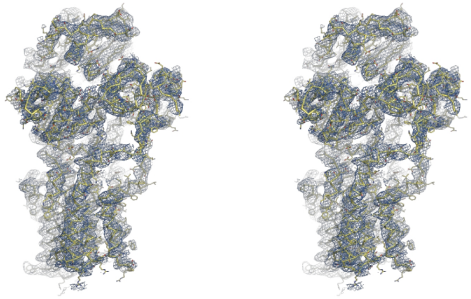
Flowchart of the steps taken to solve the AHA2 structure. The orange arrows denote the flow of information from the different datasets (Fobs: observed experimental amplitudes, HL: Hendrickson-Lattman coefficients). The green boxes denote isolated procedures where no information except xyz-coordinates of HA-positions are interchanged. Box 1: Molecular replacement. Box 2: Initial experimental SIRAS phasing using initial Holmium sites to identify and refine Holmium sites and Platinum sites in derivative data. Box 3: Final experimental MIRAS phasing using refined Ho- and Pt-sites and identification of Tantalum sites. Box 4: Final experimental SIRAS phasing using refined Tantalum sites.



CC(main) = 0.58 CC(side) = 0.25

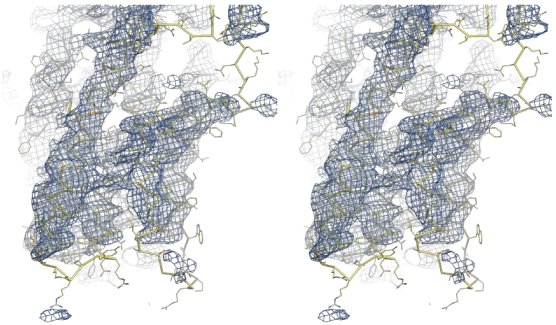


CC(main) = 0.71 CC(side) = 0.38

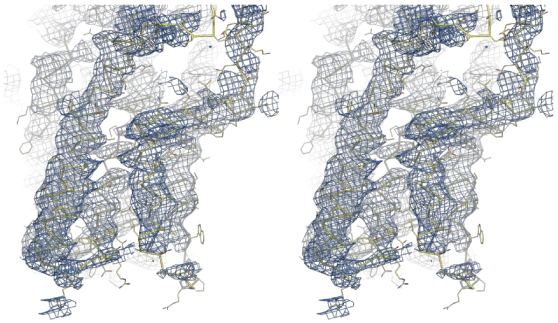


CC(main) = 0.74 CC(side) = 0.43

AHA2
overall



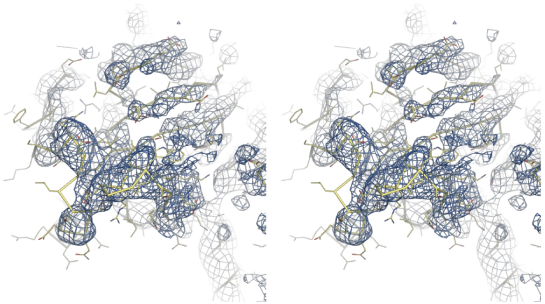
CC(main) = 0.62 CC(side) = 0.31



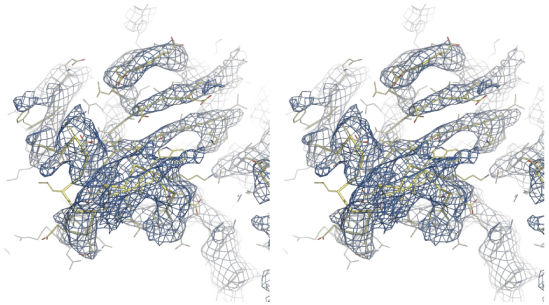
CC(main) = 0.73 CC(side) = 0.42

CC(main) = 0.76 CC(side) = 0.46

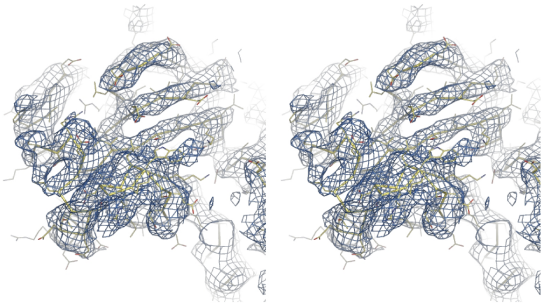
AHA2
TM domain



CC(main) = 0.44 CC(side) = 0.21



CC(main) = 0.62 CC(side) = 0.37



CC(main) = 0.66 CC(side) = 0.44

AHA2
A domain

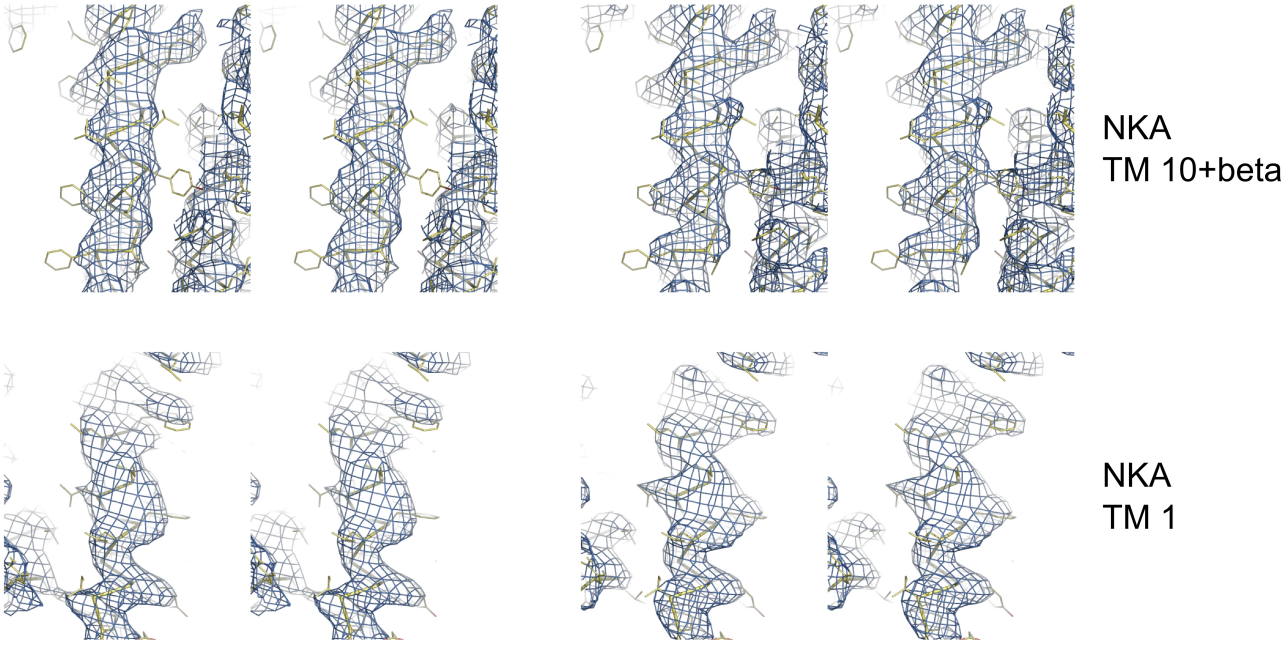
DM map
- Input: Ho+Pt+Ta combined phases from SIGMAA
- 2-fold NCS

DMMULTI map
- Input: Ho+Pt+Ta combined phases from SIGMAA
- 2-fold NCS
- 4-fold multicrystal avr.

DMMULTI map
- Input: Ho+Pt MIRAS experimental phases and Ta SIRAS experimental phases from SHARP
- 2-fold NCS
- 4-fold multicrystal avr.
- phases combined in dmmulti after real-space averaging

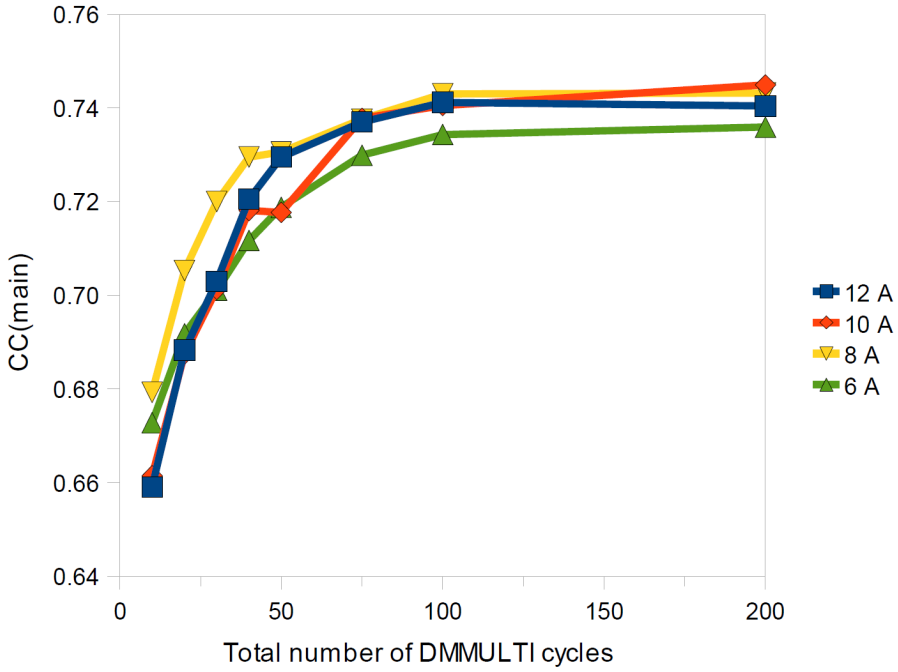
Supplementary Figure 5. Improvements in Electron Density of AHA2 in stereo-view.

This figure is a stereo-view of Figure 2. It show a sequential look at the improvement in experimental electron density maps (1σ) of the H^+-ATPase monomer, transmembrane domain (TM) and actuator domain (A) as phases are improved by multi-crystal averaging (from column 1 to 2) and by phase-combination after real space density averaging (from column 2 to 3). The correlation coefficient is listed for main chain atom density and for side chain atom density as compared to the published model (PDB entry 3b8c).



Supplementary Figure 6. Improvements in Electron Density of NKA.

A sequential look at the improvement in experimental electron density maps (displayed in stereo-view at 1σ) of the transmembrane helices of NKA as phases are improved by multi-crystal averaging. The correlation coefficient is listed for main chain atom density and for side chain atom density as compared to the published model (PDB entry 3b8e).



Supplementary Figure 7.

Correlation coefficient of the main chain density of AHA2 as a function of the number of phase extension cycles in *DMMULTI*. The different curves show the development when starting from different resolution cutoffs from 12 Å to 6 Å (see also Supplementary Table 4).

R_{cross} (%)	Native 2	Native 3	Native 4	$HoCl_3$	K_2PtCl_6	$(Ta_6Br_{12})Br_2$ (4.2 Å)
Native 1 (3.6 Å)	29.7	14.6	33.5	34.5	24.5	33.5
Native 2 (3.6 Å)		28.2	58.4	50.3	24.5	33.1
Native 3 (3.7 Å)			38.1	<u>35.5</u>	<u>19.5</u>	33.1
Native 4 (3.8 Å)				38.8	35.2	<u>24.9</u>
$HoCl_3$ (6.0 Å)					41.7	42.3
K_2PtCl_6 (4.3 Å)						31.0

Supplementary Table 1. Scaling R-factors between Different Datasets of AHA2

R_{cross} is the merging R-factor on intensities calculated between different native and derivative datasets of AHA2. Native 1 was the highest quality dataset and native 2 the highest quality non-isomorphous dataset as assessed by normal data statistics analysis. Native 3 was used for MIRAS with derivative 1 ($HoCl_3$) and derivative 2 (K_2PtCl_6), and native 4 was used for SIRAS with derivative 3 (Ta_6Br_{12})Br_2) (R-factors underlined). Derivative data was scaled to more than 40 different native datasets (data not shown), with R values in the range as seen here to identify the best match for phasing. The $HoCl_3$ dataset was highly non-isomorphous with all native datasets tested with R_{cross} values in the mid-thirty range. Therefore Native 3 was selected to allow the use of the $HoCl_3$ dataset in a MIRAS combination with the K_2PtCl_6 dataset, as this was empirically shown to result in the best experimental phases. All scaling was done within the range of the data where $I/\sigma I > 1.8$ to ensure consistency, and the maximal resolution of each dataset is listed in parentheses in the table. The four native datasets were subsequently used for 4-fold multi-crystal averaging.

Data	a (Å)	b (Å)	c (Å)	Resolution (Å)	R_{Cross} (%)	Derivative match
Native 1	85.29	144.42	312.11	3.6	-	-
Native 2	85.70	143.67	311.78	3.6	29.7	-
Native 3	85.30	143.30	312.30	3.7	14.6	$HoCl_3$ / K_2PtCl_6
Native 4	85.50	143.90	313.70	3.8	33.5	$(Ta_6Br_{12})Br_2$

Supplementary Table 2. Unit cell parameters of the datasets used for AHA2 data processing.

Unit cell parameters of the four native datasets (all in P $2_12_12_1$) used in *DMMULTI*, and the intensity-based R_{cross} value to native 1 which was subsequently used in refinement. Despite small variations in unit-cell parameters, only native 3 scaled in an acceptable way to native 1, and it was advantageous to average the datasets in real space using multi-crystal averaging (as show in figure 2) to combine the phases calculated from tantalum, holmium and platinum with the higher resolution amplitude data in native 1.

Resolution (Å)	R_{cross}
22.0	0.24
12.7	0.17
9.8	0.21
8.3	0.22
7.3	0.24
6.6	0.26
6.1	0.26
5.7	0.28
5.3	0.28
5.0	0.29
4.6	0.30
4.2	0.31
4.1	0.32
3.8	0.36
3.7	0.41
3.5	0.55

Supplementary Table 3. Scaling R-factors between the two different datasets of NKA.

Intensity -based R_{cross} in resolution shells calculated between the two native datasets used for multi-crystal averaging of NKA. While NKA did not shown the same level of non-isomorphism as AHA2, the use of multi-crystal averaging still resulted in a drastic electron density improvement (cf. Supplementary Figure 6).

DMMULTI run	CC(main)
phases extended from 12 Å, FOM(Ho+Pt)=0.61 FOM(Ta)=0.83	0.74
phases extended from 10 Å, FOM(Ho+Pt)=0.48 FOM(Ta)=0.79	0.74
phases extended from 8 Å, FOM(Ho+Pt)=0.38 FOM(Ta)=0.65	0.74
phases extended from 6 Å, FOM(Ho+Pt)=0.26 FOM(Ta)=0.27	0.73
phases extended from 5 Å, FOM(Ho+Pt)=0.15 FOM(Ta)=0.09	0.70
phases extended from 4 Å, FOM(Ho+Pt)=0.09 FOM(Ta)=0.04	0.67

Supplementary Table 4. Correlation coefficient of the main chain density of AHA2 when starting from different resolution cutoffs.

The best fit are seen if the phase extension are applied from a conservative starting point where FOM>0.25. All runs are made with 400 cycles to ensure a plateau was reached (cf. Supplementary Figure 7). This correlates to the point were the strong helix scattering diminishes in strength (cf. Supplementary figure 1) and where data quality as a consequence is reduced.

Supplementary scripts

The following are examples of the scripts used to run phaser, generate and convert masks, calculate translational matrices and run *DMMULTI* as described in the main text. All scripts are written using csh, and should be saved and run typing 'csh <filename>'

Example of *Phaser* (v2.1.4) script.

This type of script is easily modified to test various combinations of parameters, which can be useful when handling numerous *Phaser* runs as might be necessary to find a correct solution. In this particular case one ensemble is searched for two times.

```
#! /bin/csh -f
##################################################
#
# save as 'phaser_run01.csh'
# run by typing 'csh phaser_run01.csh'
# copy and rename to run02 etc for new runs
#
set run = "01"
set rmsd = "2.5"
set highres = "7.0"
set pack = "30"
set number = "2"
set model = "./searchmodel.pdb"
set data = "./best_low_Res_data.mtz"
set seq = "./target.seq"
#
##################################################
# ensure that old runs are not overwritten
if ( -e phaser_run${run}.log && $1 != "force" ) then
  echo ""
  echo "ERROR"
  echo ""
  echo "phaser job $run is already running..?"
  echo "type 'phaser_run${run}.csh force' to force this run"
  exit
endif
nohup nice +19 phenix.phaser << EOF > phaser_run${run}.log &
MODE MR_AUTO
HKLIN $data
LABIN  F=FP SIGF=SIGFP
TITLE test ${data} using ${model}
COMPOSITION PROTEIN SEQ $seq NUMBER $number
RESOLUTION 100.0 $highres
ENSEMBLE $model &
    PDBFILE ${model} &
    RMS $rmsd
SEARCH ENSEMBLE $model NUMBER $number
PACK $pack
FINAL ROT SELECT PERCENT 75.0
FINAL TRA SELECT PERCENT 75.0
SAVE ROT CLUSTER ON DUMP 20
SAVE TRA CLUSTER ON DUMP 20
PERMUTATIONS OFF
ROOT ./phaser_run${run}
EOF
```

Generation of averaging masks.

Masks are easily generated in *O* (v12.0.0) (Jones T.A., Zou J.Y., Cowan S.W. & Kjeldgaard M. (1991). *Acta Crystallogr A*. **47**, 110-119.) using 'ncs-mask_sphere', 'ncs_mask_layer' and 'ncs_mask_write' commands. An example is given below to be used directly in *O* (note that this is not a csh script, but should be run directly in *O*).

```
! Commands to be run in O
! Load pdb-file with best model (model1)
pdb_read ./model.pdb model1 ;;;
! Load map (map1) from dataset you wnat to use in DMMULTI (to get grid)
fm_file ./data.map map1 p212121

! Create new object (obj1) covering only chain A residues 1-800
mol model1 obj obj1
zone a1 a800
end

! Create mask (mask1) around obj1 using grid from map
ncs_mask_sphere map1 obj1 3.0 mask1

! Optional: view the mask to ensure it looks ok
fm_set mask1 80 solid 1 1 magenta
fm_mode mask1 rmsd

! Increase mask and then descrease mask to smoothen and remove 'holes'
ncs_mask_layer mask1 +
ncs_mask_layer mask1 -

! Write out the mask
ncs-mask-write mask1 mask1_from_o.mask
```

Conversion of averaging-masks.

O-format masks can be converted to ccp4-format using the program *MAMA2CCP4* (v6.1).

```
#! /bin/csh -f
##########################################
#
# mama2ccp4-script
#
# Generate CCP4 mask from O-output mask
#
##########################################
mama2ccp4 MASKIN ./mask1_from_o.mask MASKOUT ./chain_a.ccp4_mask
```

Generation of solvent mask.

The ccp4-format averaging mask can be extended to be used as the solvent mask using *MAPMASK* (v6.1) with the following script.

```
#! /bin/csh -f
#######################################
#
# mapmask-script
#
# Generate solvent mask from averaging mask.
# Remember that the input mask should
# cover the entire asymmetric unit!
#
#######################################
mapmask MSKIN ./chain_ab.ccp4_mask MSKOUT ./solvent.mask << EOF
SYMMETRY  P212121
EXTEND OVERLAP
XYZLIM CELL
EOF
```

Calculation of initial translational matrices for use in *DMMULTI*.

These are easily generated by running phaser to place the rudimentary model in the different datasets the best possible. Use a variation of the above mentioned *Phaser* script to accomplish this. Thereafter the translational matrix between any two given chains can be calculated using *LSQKAB* (v6.1) and ready-to-copy-paste output for *DMMULTI* can be generated using the following script:

```
#! /bin/csh -f
#############################################################
#
# lsqkab-script
#
# used to extract rotation and translation matrix
# from lsqkab and set it up for use with dm or dmmulti
#
# in lsqkab, FIT refers to moved_mol.pdb
# and MATCH refers to fixed_mol.pdb
#
# In the example below, the output is the matrix to move
# residue 1-800 of chain A of moved_mol.pdb to chain B of fixed_mol.pdb
#
#############################################################

lsqkab \
 XYZINM ./moved_mol.pdb \
 XYZINF ./fixed_mol.pdb \
 <<EOF | grep -A 4 "      ROTATION MATRIX:"| awk \
     'NR == 2 {printf ("ROTA MATR  %10s %10s %10s -\n", $1, $2, $3)} \
  NR == 3 {printf ("           %10s %10s %10s -\n", $1, $2, $3)} \
  NR == 4 {printf ("           %10s %10s %10s  \n", $1, $2, $3)} \
  NR == 5 {printf ("TRANS      %10s %10s %10s  \n", $5, $6, $7)}'

FIT RESIDUE 1 TO 800 CHAIN A
MATCH 1 TO 800 CHAIN B
END
EOF
```

Example of *DMMULTI* (v6.1) script. This type of script is easily modified to test various combinations of parameters, which can be useful when facing the numerous *DMMULTI* run that might necessary to find a correct solution. In this particular case four datasets are input, two with phase-information and two without (as shown in Supplementary Figure 4).

```
#! /bin/csh -f
##############################
#
# dmmulti run number
set run = "01"
#
# save as 'dmmulti_run01.csh'
# run by typing 'csh dmmulti_run01.csh'
# copy and rename to run02 etc for new runs
#
# example from AHA2
# crystal 1 is pt-ho MIRAS data
# crystal 2 is TaBr SIRAS data
# crystal 3 is best native without phase information
# crystal 4 is "best" nonisomorphous native without phase information
#
# we use the same solvent mask for all crystals
# we use PHI/FOM directly from SHARP
#
# the matrices for XTAL 1,2 are refined from a previous DMMULTI run
# the matrices for XTAL 3,4 comes from MR
# we start slowly (400 cycles) from 12 Å resolution
# we look for two copies of MSKIN1 in the asymmetric cell in each XTAL
#
##############################

# ensure that old runs are not overwritten
if ( -e dmmulti_run${run}.log && $1 != "force" ) then
  echo ""
  echo "ERROR"
  echo ""
  echo "dmmulti job $run is already running..?"
  echo "type 'dmmulti_run${run}.csh force' to force this run"
  echo ""
  exit
endif


nohup dmmulti \
 HKLIN1  ./pt-ho-sharpphases.mtz \
 HKLIN2  ./ta-sharpphases.mtz \
 HKLIN3 ./best_native.mtz \
 HKLIN4 ./other_native.mtz \
 HKLOUT1 ./dmmulti_run${run}_crystal1.mtz \
 HKLOUT2 ./dmmulti_run${run}_crystal2.mtz \
 HKLOUT3 ./dmmulti_run${run}_crystal3.mtz \
 HKLOUT4 ./dmmulti_run${run}_crystal4.mtz \
 SOLIN1 ./solvent.mask \
 SOLIN2 ./solvent.mask \
 SOLIN3 ./solvent.mask \
 SOLIN4 ./solvent.mask \
 MSKIN1 ./chain_a.mask << EOF >dmmulti_run${run}.log&

NCYCLE 400
```

```
# Crystal 1
XTAL 1
LABIN FP=FP1 SIGFP=SIGFP1 HLA=HLA HLB=HLB HLC=HLC HLD=HLD
LABOUT PHIDM=PHIDM1 FOMDM=FOMDM1
MODE SOLV HIST AVER
SOLC 0.75 MASK 0.74 0.25
RESOLUTION 40.0 3.9 ! normally take from mtz-file
SCHEME RES FROM 12.0

AVER DOMAIN 1
ROTATION MATRIX: 1 0 0 0 1 0 0 0 1
TRANSLATION 0 0 0

AVER DOMAIN 1 REFINE
ROTA MATR  -0.34253  0.93950 -0.00305 -
            0.93911  0.34248  0.02808 -
            0.02743  0.00675 -0.99960
TRANS   27.40607 -20.32169 109.45611

# Crystal 2
XTAL 2
LABIN FP=FP2 SIGFP=SIGFP2 PHIO=PHIB FOMO=FOM
LABOUT PHIDM=PHIDM2 FOMDM=FOMDM2
MODE SOLV HIST AVER
SOLC 0.75 MASK 0.74 0.25
RESOLUTION 40.0 3.8 ! normally take from mtz-file
SCHEME RES FROM 12.0

AVER DOMAIN 1 REFINE
ROTA MATR   0.99994 -0.00002  0.01081 -
            0.00004  1.00000 -0.00169 -
           -0.01081  0.00169  0.99994
TRANS          0.53     0.42     0.24

AVER DOMAIN 1 REFINE
ROTA MATR  -0.34737  0.93767  0.00990 -
            0.93718  0.34679  0.03790 -
            0.03210  0.02245 -0.99923
TRANS         27.85   -20.29   109.93

# Crystal 3
XTAL 3
LABIN FP=FP3 SIGFP=SIGFP3
LABOUT PHIDM=PHIDM3 FOMDM=FOMDM3
MODE SOLV HIST AVER
SOLC 0.75 MASK 0.74 0.25
RESOLUTION 30.0 3.6 ! normally take from mtz-file
SCHEME RES FROM 12.0

AVER DOMAIN   1 REFINE
ROTA MATR   0.99999  0.00205  0.00167 -
           -0.00205  0.99999  0.00330 -
           -0.00166 -0.00330  0.99999
TRANS          0.03    -0.07    -0.05

AVER DOMAIN   1 REFINE
ROTA MATR  -0.35049  0.93656 -0.00384 -
            0.93613  0.35045  0.02888 -
            0.02839  0.00653 -0.99958
TRANS         27.31   -20.23   109.36
```

```
# Crystal 4
XTAL 4
LABIN FP=FP4 SIGFP=SIGFP4
LABOUT PHIDM=PHIDM4 FOMDM=FOMDM4
MODE SOLV HIST AVER
SOLC 0.75 MASK 0.74 0.25
RESOLUTION 30.0 3.6 ! normally take from mtz-file
SCHEME RES FROM 12.0

AVER DOMAIN   1 REFINE
ROTA MATR   0.99993  0.00519  0.01049 -
           -0.00518  0.99999 -0.00108 -
           -0.01050  0.00102  0.99994
TRANS         -0.29     0.15     0.06

AVER DOMAIN   1 REFINE
ROTA MATR  -0.34670  0.93795 -0.00646 -
            0.93744  0.34673  0.03150 -
            0.03179  0.00487 -0.99948
TRANS         27.82   -20.52   109.38

EOF
```