CRYSTALLOGRAPHY

PhAI: A deep-learning approach to solve the crystallographic phase problem

Anders S. Larsen, Toms Rekis, Anders Ø. Madsen*

X-ray crystallography provides a distinctive view on the three-dimensional structure of crystals. To reconstruct the electron density map, the complex structure factors $F = |F|\exp(i\phi)$ of a sufficiently large number of diffracted reflections must be known. In a conventional experiment, only the amplitudes $|F|$ are obtained, and the phases ϕ are lost. This is the crystallographic phase problem. In this work, we show that a neural network, trained on millions of artificial structure data, can solve the phase problem at a resolution of only 2 angstroms, using only 10 to 20% of the data needed for direct methods. The network works in common space groups and for modest unit-cell dimensions and suggests that neural networks could be used to solve the phase problem in the general case for weakly scattering crystals.

To this day, the detail and precision of the structures determined by crystallography, which range from simple metals to large membrane proteins, are unsurpassed by any other method. However, the major challenge—the so-called phase problem (*1*)—remains retrieving the phase information from the experimentally determined amplitudes (Fig. 1). In the early days of crystallography, a tremendous effort was put to recover the phase information, which was often done in a trial-and-error fashion complemented with tedious calculations. Later on, the Patterson method (*2*) and Beevers-Lipson strips (*3*) (an aid for speeding-up Fourier synthesis done by hand) allowed for work on more complex structures. A major breakthrough followed in the 1950s and '60s, when so-called direct methods for solving the phase problem were developed by Karle and Hauptmann (*4*, *5*), featuring important contributions from Sayre (*6*, *7*) and others. The former two were awarded the Nobel Prize in 1985 for their outstanding achievement in the development of direct

Department of Pharmacy, University of Copenhagen, Copenhagen, Denmark.
*Corresponding author. Email: a.madsen@sund.ku.dk

methods. Despite some methodological changes, it is textbook knowledge that direct methods require diffraction data of an atomic resolution, that is, a minimum lattice plane distance (d_{min}) equal to 1.2 Å or better (*8*).

The requirement of atomic resolution is, however, an empirical observation (*9*). Some theoretical justification has been proposed, namely that the typical interatomic distances in organic molecules, including proteins, are about 1.2 Å, thus creating distance beats in the pair-distribution function at this distance (*8*). In recent years, the traditional direct methods have been supplemented by dual-space approaches such as charge flipping (*10*, *11–15*), modern direct methods (*16*, *17*), and VLD (vive la difference) (*18–20*). These ab initio approaches can sometimes solve the phase problem at lower resolution (1.4 to 1.6 Å), though the success varies from case to case and is related to data quality and the presence of heavy atoms in the structures (*15*, *21*). For protein crystallography, the Arcimboldo approach (*22*), in which substructure fragments are used as search elements, have been successful at 2-Å resolution. Presently available ab initio methods appear to have reached their limits. A general solution to the phase problem remains unknown.

Mathematically, any combination of structure factor amplitudes with phases can be inverse-Fourier-transformed. However, physical and chemical requirements, such as having an atom-like electron density distribution, impose rules for the possible combinations of phases that are in congruence with a set of amplitudes. Advancements in deep learning have allowed the exploration of such relationships, perhaps in more depth than is possible with the present ab initio approaches. Neural networks readily learn from large datasets and can infer rules from the data that elude humans. This suggests that the training of neural networks could be a winning strategy for solving different problems in crystallography.

The use of neural networks originated in computer vision and natural language processing. It has since found applications in the natural sciences as well, for example, for protein structure prediction (*23*). Diverse artificial intelligence tools have been developed in relation to one-dimensional (1D) phase retrieval in spectroscopy (*24*, *25*) and 2D phase retrieval in real- and complex-valued image analysis (*26*). As far as crystallography is concerned, there have been some studies that involved the deduction of Bravais lattice type, crystal system, or space group symmetry based on powder or single crystal x-ray or electron diffraction patterns (*27–32*).

In this work, we took a data-driven approach, using millions of artificial crystal structures and their corresponding diffraction data, with the aim of solving the phase problem in crystallography. We showed that this ab initio structure solution approach, which is based on deep learning, could be performed at a resolution of only $d_{min} = 2.0$ Å. This resolution corresponds to only 10 to 20% of the data available at atomic resolution, which is often needed for the traditional ab initio approaches. This result clearly demonstrates that atomic resolution is not necessary for ab initio methods and opens new avenues for deep learning–based structure determination. In addition, our approach showed a notable accuracy of the retrieved electron density maps. Here, we demonstrate this approach for the most commonly occurring space groups ($P2_1/c$, $C2/c$, $Pbca$, $Pnma$,

Fig. 1. A flowchart of a standard crystal structure determination. In a conventional diffraction experiment, amplitudes $|F|$ of the measured reflections can be derived. Next, the phase problem needs to be solved to obtain corresponding phases ϕ. An inverse Fourier

transform of the structure factors $F = |F|\exp(i\phi)$ leads to the electron density function within the unit cell from which the initial atomic positions and types are inferred. This initial structure model is then refined against the experimental data. The modern ab initio phasing methods do not retrieve the phase information from the structure factor amplitudes in one step. The electron density function calculated from some initial phases is modified in an iterative fashion until a satisfactory starting model is obtained.
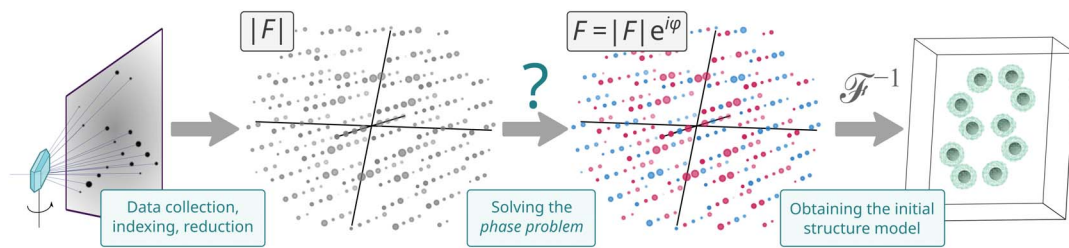
Pbcn, and *C*2/*m*) and, for the sake of computational costs, we set modest limits on the crystallographic unit-cell dimensions. The trained neural network showed an outstanding performance and proves that the phase problem can be solved extremely successfully using deep learning.

Designing and training the neural network

The neural network, which we call PhAI, was constructed to accept structure factor amplitudes $|F|$ and output the corresponding phase values ϕ. The architecture of PhAI is presented in Fig. 2. The number of structure factors that represent a crystal structure depends on the unit-cell size. Based on the computational resources, we set limits on the size of the input data. The input structure factor amplitudes were chosen for reflections having Miller indices (h, k, l) that obey $\sqrt{h^2 + k^2 + l^2} \leq 10$. In practice, this choice limits us to structures of unit-cell dimensions of about 10 Å at atomic resolution. In addition, we chose the most common centrosymmetric space group $P2_1/c$. The centrosymmetry limits the possible phase values to either zero or π rad.

We arranged the input amplitudes in a 3D tensor of a fixed shape of 21 by 11 by 11. This aspect ratio allowed us to work with the unique reflections of the respective Laue symmetry, that is, $2/m$ with $-10 < h < 10$, $0 < k < 10$, and $0 < l < 10$. The neural network has two inputs—one for the amplitudes and another

for the phases. These phases are initially set to zero or random values. A separate input tensor for phases allowed us to implement refinement of the predictions. We refer to this procedure as phase recycling (Fig. 2). The amplitudes and phases are treated with an input block that contains a 3D convolutional layer, after which they are added together. The next stage of the network contains a number of 3D convolutional blocks. The output feature maps are flattened and projected into multilayer perceptron (MLP) tokens and fed into a number of MLP blocks. The output tokens from the MLP blocks are reduced with global average pooling, and a final linear classification head predicts the phase of each reflection. The predicted phases can then be recycled into the network. See the methods section in the supplementary materials for a detailed description of the network.

We trained the neural network using artificial crystal structures that mostly contained organic molecules. We chose to use artificial structures and retain the experimental structures available in crystallographic databases as a benchmark set. The use of artificial structures allowed us to generate a large and diverse training set in the chosen space group. Around 49,000,000 structures were created, with a breakdown of 94.29% organic, 5.66% metalorganic, and 0.05% inorganic crystal structures. Diverse organic molecules and occasionally metal atoms were placed into unit

cells with randomly sampled parameters. Interatomic distances were checked to avoid atom overlap and unrealistically close contacts. The organic molecules were sampled from a database of SMILES strings and were force field optimized before insertion. The supplementary materials contain a detailed description of the structure generation.

During training, isotropic atomic displacement parameters U_{iso}, data resolution d_{min}, and completeness were sampled on the fly. The corresponding structure factor amplitudes and phases were computed using tabulated spherical atomic form factors. During training, the phases were recycled between one and three times. The initial phases were all set to a value of zero, and in subsequent cycles, the network's phase predictions were reused.

The neural network was trained to minimize the loss function that takes into account the allowed origin translations in $P2_1/c$. A detailed description of the loss function and training procedure is given in the methods section of the supplementary materials.

Solving real structures with the neural network

The trained neural network runs on a standard computer with modest computational demands. It accepts a list of *hkl* indices and the corresponding structure factor amplitudes as input. No other input information is necessary, not even the unit-cell parameters
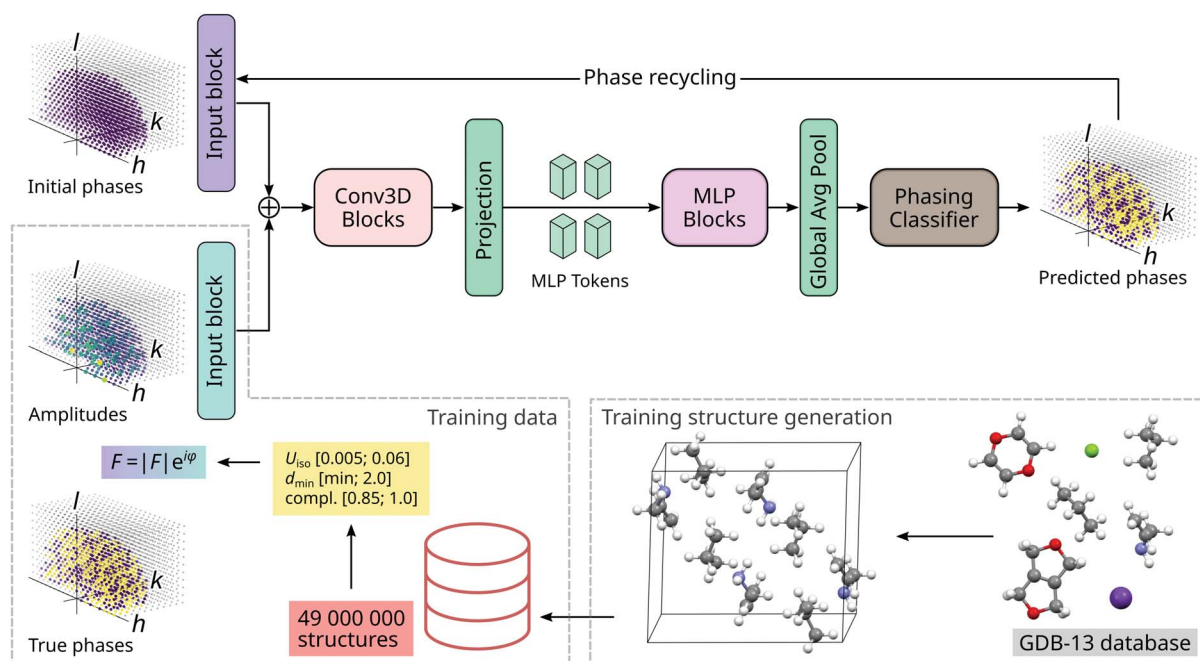

Fig. 2. The PhAI neural network approach to the phase problem. The input to the neural network consists of the amplitudes and phases, which are treated with a convolutional input block, added and inputted to a series of convolutional blocks (Conv3D) followed by a series of MLP blocks. The predicted phases from the linear classifier (phasing classifier) are recycled through the network N_c times. The training data are generated by inserting metal atoms and organic molecules from the GDB-13 database into unit cells. The generated structures are organized into the training data from which the true phases and structure factor amplitudes can be calculated when sampling the temperature factors, resolution, and completeness.
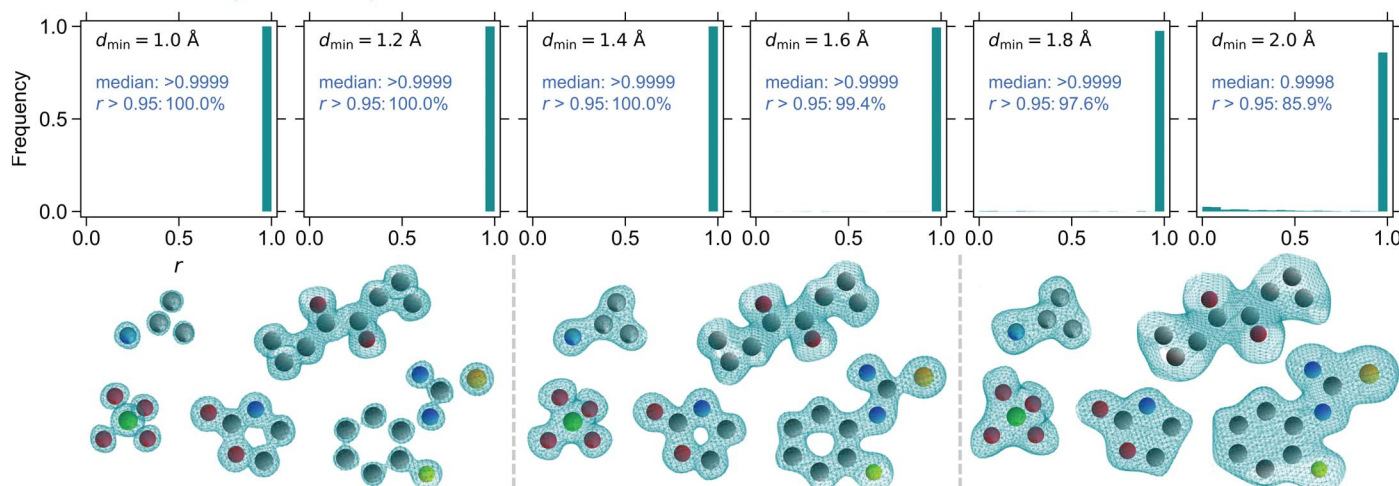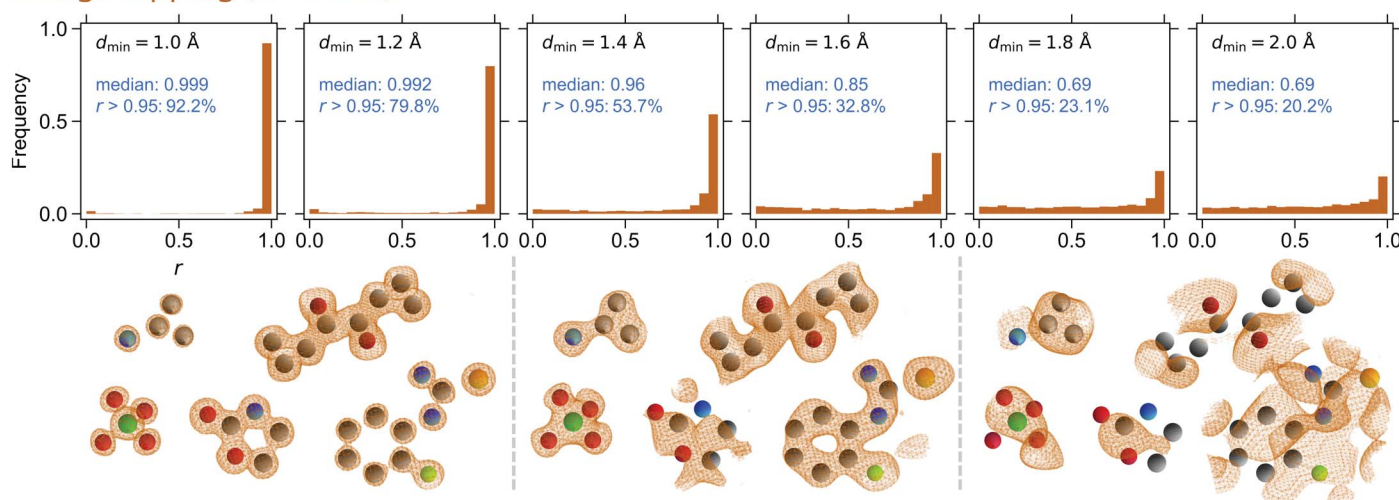
neural network (N = 2387)

charge flipping (N = 2387)

Fig. 3. Histograms of the correlation coefficients _r_ between the phased and true electron density maps. The structures (_N_ = 2387) were retrieved from the crystallographic databases, and the diffraction data were calculated at different resolutions. The phased electron density maps of arbitrarily selected examples are shown. The example electron density maps are provided at three resolution ranges separated by vertical dashed gray lines; from left to right, the ranges are 1.0 to 1.2 Å, 1.4 to 1.6 Å, and 1.8 to 2.0 Å. The phasing was performed by the PhAI neural network and charge flipping for comparison.

of the structure. This is a fundamental difference from all other modern ab initio methods where electron density function modification plays a crucial role to a successful structure solution. The network predicts and outputs phase values instantly.

Initially, we tested the performance of the neural network using calculated diffraction data of real crystal structures. All structures of organic, metal-organic, and coordination compounds corresponding to the space group $P2_1/c$ and respecting the unit-cell-size limitations (≤10 Å) were retrieved from crystallographic databases [the Cambridge Structural Database (CSD) (33), Crystallography Open Database (COD) (34, 35), and Inorganic Crystal Structure Database (ICSD) (36)]. A total of 2387 test cases were obtained (see supplementary

materials section 4.2 for more details). Several values of data resolution, ranging from 1.0 to 2.0 Å, were considered for all gathered structures. For comparison, the charge flipping method was also used to retrieve the phase information [specifically, the program SUPERFLIP (37) was used].

To assess the success of the phase retrieval, the correlation coefficient r between the phased and true electron density map was used. This correlation coefficient can be calculated without generating the actual density maps (38):

$$r = \max_{\mathbf{T} \in T} \left\{ \frac{\sum |F(\mathbf{H})|^2 \cos\left[\phi_t(\mathbf{H}, \mathbf{T}) - \phi_p(\mathbf{H})\right]}{\sum |F(\mathbf{H})|^2} \right\}$$

where $\phi_t(\mathbf{H}, \mathbf{T})$ is the true phase value of reflection $\mathbf{H}$ taking into account a possible

alternative origin translated by $\mathbf{T}$ from the allowed origins T, and $\phi_p(\mathbf{H})$ is the predicted phase value of reflection $\mathbf{H}$.

The trained neural network showed an outstanding performance; it could solve all tested structures ($N = 2387$) if the corresponding diffraction data resolution was good and excelled at structure solution from low-resolution data (see Fig. 3). Although the neural network was hardly trained on inorganic structures, it could perfectly solve such structures, as shown in fig. S3. The charge-flipping method was excellent for dealing with good-resolution data, but its ability to yield a reasonably correct solution gradually decreased as the data resolution got poorer; however, it still solved about 32% of the structures at 1.6-Å resolution. The number of structures determined by charge flipping
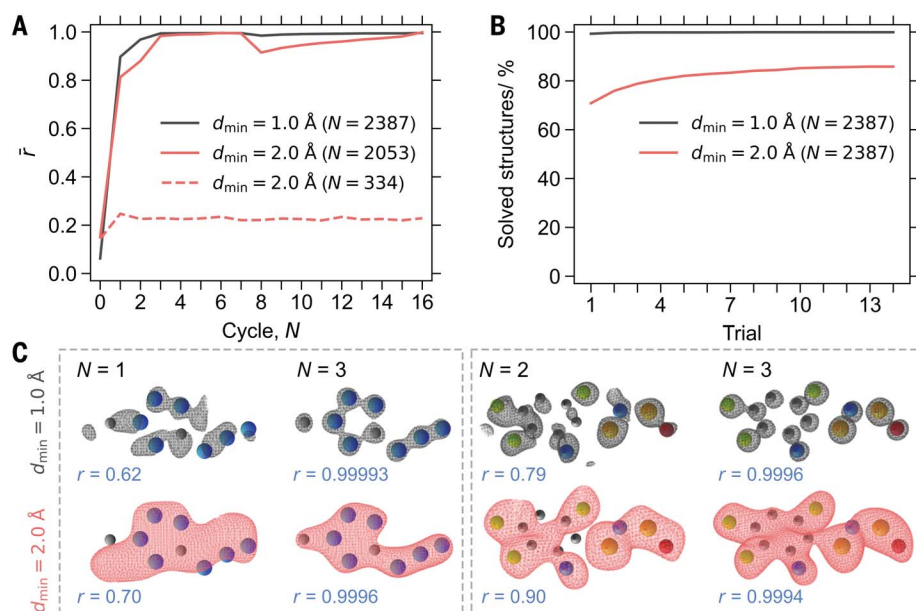
Fig. 4. Phase recycling and solution trials. (**A**) The average correlation coefficient between the true and phased density maps as a function of phase recycling. Starting with random phases (zero cycles), excellent solutions are obtained by recursively improving the predictions within three or four cycles. (**B**) Cumulative fraction of the solved structures ($r > 0.95$) as a function of phasing trials. Each trial was started with random phases and recycled 16 times. (**C**) Phased maps of two example structures using 1.0- and 2.0-Å-resolution data. The maps were calculated from phases obtained after one, two, or three recursive phasing cycles.

could probably be improved by further trials and changing input parameters such as the flipping threshold. In the PhAI approach, such meta-optimization is performed during training and does not need to be performed by the user. These results showed that the common conception of the necessity to have atomic-resolution data to do ab initio phasing in crystallography could be broken. PhAI only required 10 to 20% of the data available at atomic resolution.

The accuracy of the solutions provided by the neural network was exceptionally high (note the median values of the density map correlation coefficients r in Fig. 3). The reflections for which the phases were not predicted correctly were mostly weak reflections that did not contribute meaningfully to the quality of the solution (see also fig. S1). The extreme accuracy of the solutions, and thus of the electron density maps, suggested that robust structure solution procedures could be developed using neural networks. The high accuracy of the density maps implied that their interpretation in terms of atomic positions and types would rely less on the user input. This fact could greatly improve the overall quality of the structures deposited in databases and published in the literature.

Robustness of the neural network

The design of PhAI allows for iterative phase refinement. Along with the input structure factor amplitudes, a list of phases set to zero or random values are provided. The phases can be improved by executing the neural network several times with the input phases set to the predicted values from the previous cycle. We observed that either excellent solutions were found within a few cycles or the network failed to find a solution at all. The latter, however, was only observed for a relatively small fraction of structures at low resolution (Fig. 3).

If the solution could be found with PhAI, it was typically already good ($r > 0.8$) in the first prediction cycle (with zero or random phase values as input). Occasionally, the initial predictions were less accurate, as illustrated by two examples in Fig. 4C. Regardless, excellent solutions were obtained within three or four cycles. This is demonstrated in Fig. 4A, where the trajectories of the average correlation coefficients $\bar{r}$ are shown (see also fig. S2, where trajectories of selected individual cases are shown). For the failed cases of the analyzed 2.0-Å-resolution data, the average trajectory is also shown in Fig. 4A. Individual trajectories for the failed cases are diverse, and some typical examples are shown in fig. S2.

The results show that for the successful cases, the convergence was reached immediately, and similar prediction trajectories were observed for thousands of input datasets, independent of the data resolution. This detail is important to note because solving an unknown structure is more complicated if the behavior of the solution convergence is different from case to case.

Most of the tested input datasets led to an excellent solution with a single trial of random initial phases and three or four cycles, as discussed above. However, there were cases for which no satisfactory solution was obtained even after thousands of cycles. We did not find it particularly effective to increase the number of cycles. Instead, a new trial with different random initial phases was started. At each resolution, the structures that could not be solved within 16 cycles were passed to the next trial. At a resolution of 1.0 Å, >99% of the tested 2387 structures could be solved in the first trial. The remaining structures could be successively solved within the next 14 trials. For 2.0-Å-resolution data of the same 2387 structures, only around 70% could be solved in the first trial. Additional trials increased the percentage to around 86% (see Fig. 4B, where the cumulative percentage of the solved structures after each trial is depicted). The plot suggests that a limit was reached for 2.0-Å-resolution data. Indeed, we could conclude that most of the remaining structures at 2.0-Å-resolution are presently not solvable with PhAI. We performed trials using 90 and 100% of correct phase values in the initial input tensor for the unsolved 334 structures. Only a couple dozen cases were stable during successive execution of the network. We expect that those structures could eventually be solved with PhAI using additional trials with random starting phase values; however, this would increase the cumulative percentage of correct solutions ($r \geq 0.95$) only by 1%, from 86 to 87% for the tested 2387 structures at 2.0-Å-resolution.

These results demonstrate the high robustness of the trained neural network. If the solution could be found by PhAI, (i) it would be found in a small number of trials and (ii) it would be accurate within three or four phasing cycles. The average correlation coefficient of all solved structures (taking into account all resolutions) with PhAI was $\bar{r} = 0.99965 \pm 0.00007$ (99.9% confidence level).

Phase extension

We discovered that PhAI could do phase extension, which means retrieving the phase information for nonmeasured reflections (*39*). The design of the neural network allows it to handle reflections obeying $\sqrt{h^2 + k^2 + l^2} \leq 10$. Taking into account only the unique reflections according to the Laue class $2/m$ and omitting the systematically absent reflections in $P2_1/c$, there were 1042 reflections that corresponded to these criteria. The number of reflections associated with each individual structure varied depending on the unit-cell parameters and data resolution (see Fig. 5A, where an example is given at 1.0- and 2.0-Å resolutions). PhAI output phase values for all 1042 reflections, regardless of the actual number of reflections for which amplitude values were input.

Fig. 5. Phase extension. (**A**) Representation of the available diffraction data of an example structure at resolutions of 1.0 (gray) and 2.0 (red) Å. PhAI allows for the handling of a total of 1042 reflections (gray and teal, or red and teal). Regardless of the number of provided structure factor amplitudes, PhAI always predicted phases for all depicted 1042 reflections. The r values were also excellent when all 1042 reflections were considered. This means that the phases are predicted mostly correctly for the reflections also, for which the structure factor amplitudes are not present in the input data. (**B**) Histogram of the percentage of the available data at resolutions of 1.0 (gray) and 2.0 (red) Å for the tested structures solved by PhAI. The $\bar{r}$ values based on all 1042 reflections indicate that all

phases are mostly predicted correctly. At 2.0-Å resolution, there were, on average, 10% (≈100) reflections with structure factor amplitude information present. PhAI was also able to predict correct phases for the remaining 90% (≈940) reflections, for which the amplitudes are not provided at input.
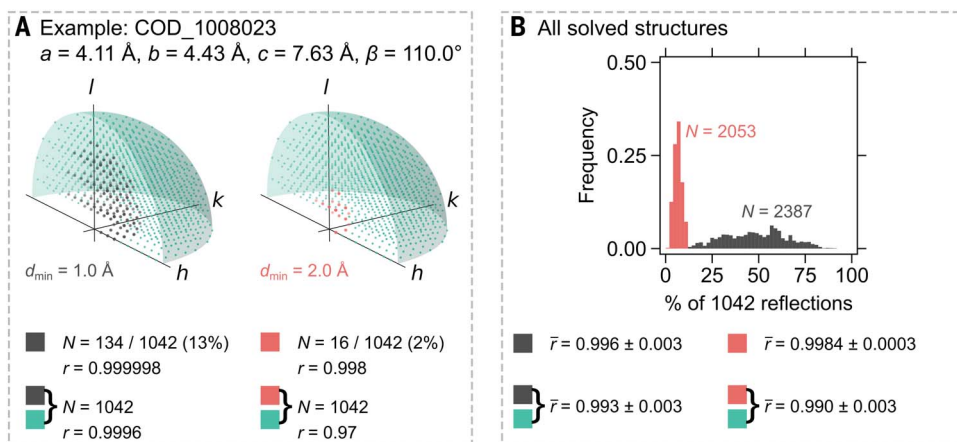
We discovered that the predicted phase values were mostly correct for all 1042 reflections. Note the r values for the specific example structure given in Fig. 5A. At 2.0-Å resolution, the structure factor amplitudes were provided for only 16 reflections (2% of the total that were possible). The phases for those reflections were predicted correctly ($r = 0.998$). Inspecting the predicted phase values for all 1042 reflections revealed that combining those with the amplitude information would result in $r = 0.97$.

In Fig. 5B, results for all solved structures at 1.0- and 2.0-Å resolution are given. The average correlation coefficient values are good. The histogram shows that the diffraction datasets at 2.0-Å resolution contained no more than 12% of the total number of admissible reflections in PhAI. Yet the phase values for the remainder of the 1042 reflections were predicted mostly correctly, as shown by $\bar{r}$. This indicates that PhAI has learned to reconstruct phase values that originate from atomistic electron density maps. We speculate that a similar extension on structure factor amplitudes would be possible to enhance structure solution from low-resolution data.

Application to experimental x-ray diffraction data

So far, the performance of the trained neural network had been demonstrated using calculated structure factor amplitudes as the input data. Spherical atomic form factors and isotropic displacement parameters were used. The real electron densities are much more complex, and it is accordingly represented in the experimentally derived structure factor amplitudes. In addition, there are systematic and random errors, as well as uncorrected effects, such as extinction and absorption. We therefore tested our developed phasing method using experimental data. We could retrieve 495 experi-

mental datasets (see supplementary materials section 5.3). All datasets were truncated so that $d \geq 1.0$.

The datasets were initially tested to see how well the experimental data corresponded to the associated structures. The correlation coefficient was calculated between the observed $|F_{obs}|$ and calculated $|F_{calc}|$ structure factor amplitudes. In addition, data completeness was calculated. The neural network was trained by randomly omitting up to 15% of the reflections, so it was anticipated that even for perfect data (with a high r between $|F_{obs}|$ and $|F_{calc}|$), it would be problematic to find a solution if the completeness was considerably lower than 85%. An overview of the data quality and completeness is summarized in fig. S16.

Low r values between $|F_{obs}|$ and $|F_{calc}|$ for some diffraction datasets indicate that those data were most likely not the actual experimental data but were associated with the respective structures through a mistake. We did not omit any of the data and tested everything, but the data were segregated into two classes, namely, (i) fine and (ii) poor or invalid experimental data. A diffraction dataset was considered fine if the total completeness was ≥85% and the r between $|F_{obs}|$ and $|F_{calc}|$ was >0.9. The rest of the datasets were classified as poor or invalid. There were 302 fine and 193 poor or invalid experimental datasets.

Like before, phase retrieval of the experimental data was done with both the trained neural network and the charge flipping algorithm. r values showing the correlation between the phased and true electron density maps are summarized in the histograms in Fig. 6A.

The neural network's performance on the fine experimental data was excellent. Only two datasets (40, 41) displayed low r values. In both cases, the experimental data were missing the strongest reflection. The neural network was trained by scaling the structure factor

amplitudes to relative values. The absence of the strongest reflection in the data posed a problem, and this is a shortcoming of the present neural network.

Phasing the poor or invalid experimental datasets by the neural network and charge flipping revealed similar results (see Fig. 6). As noted previously, some of the experimental datasets were indeed invalid, and neither charge flipping nor PhAI work for these. The charge-flipping method performed better for low-completeness data. The present version of PhAI has not been trained for such cases.

One of the virtues of PhAI is the ability to phase low-resolution data. In Fig. 6B, we show two examples of phasing powder x-ray diffraction data, namely, naphthalene and barbituric acid (42). We derived structure factor amplitudes from the experimental powder diffraction patterns using Pawley fitting in the software package DASH (43). The amplitudes could be obtained to ≈1.8-Å resolution with 84 and 95% data completeness for naphthalene and barbituric acid, respectively. As shown in Fig. 6B, the structures could be easily discerned from the obtained density maps. The charge-flipping method failed in these cases because of the low resolution. These findings indicate that neural network–based methods could greatly advance structure solutions from powder diffraction patterns. The present methods for solving structures from such data are time consuming and frequently lead to ambiguity and errors (44–46).

Solving structures with larger unit cells and other space groups

The neural network was trained only with structures in $P2_1/c$ and with a maximum unit-cell length less than 10 Å. Structures in supergroups of $P2_1/c$ should, in principle, be solvable with PhAI. Among the most frequently occurring space groups of this kind are $C2/c$, $Pbca$, $Pnma$,

Fig. 6. Application to experimental data. (**A**) Histograms of the correlation coefficients between the retrieved and true electron density maps of structures with fine experimental data (*N* = 302) (top) and poor or invalid experimental data (*N* = 193) (bottom). The phasing was done by PhAI and the charge-flipping approach for comparison. (**B**) Two examples of structure solution with PhAI from powder x-ray diffraction data. The patterns have been indexed and integrated, yielding *N* = 61 and 93 reflections for the naphthalene and barbituric acid examples, respectively. This corresponds to a data resolution of around 1.8 Å in both cases, with a lowered data completeness *C* = 84% in the case of naphthalene. a.u., arbitrary units.
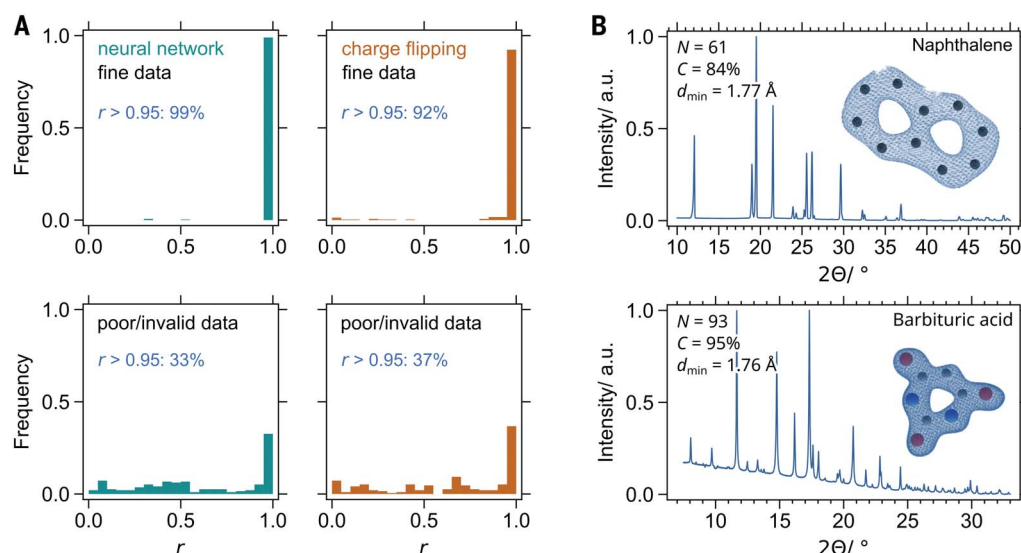
Pbcn, and *C2/m*. A total of 264 structures respecting the unit-cell-size limitation were retrieved from the CSD. Nearly all structures could be solved from atomic-resolution data with PhAI. Around 75% of these structures could be solved from 2.0-Å-resolution data. More details are shown in fig. S3.

We also tested the possibility of solving structures with larger unit cells. One thousand structures in $P2_1/c$ with a maximum unit cell length between 10 and 20 Å were selected from the CSD and tested. In each case, the resolution was limited by the fixed input of PhAI ($h, k, l \leq$ 10). The neural network could perfectly solve around 80% of the structures (see fig. S3).

Conclusions

In this work, we used deep learning for ab initio phasing to determine crystal structures. This technique allowed us to shift method development away from incremental improvements of the established dual-space algorithms that appear to have reached their limits. We lay out how the crystallographic phase problem could be approached using deep learning and propose a framework that can be extended to encompass data from all possible experimental scenarios, such as high-pressure or electron diffraction experiments for which interpretation of the data is complicated by low resolution. We envision that this framework could also be used for data with low completeness or dynamical scattering effects, once these effects are incorporated into the training data, though this remains to be investigated. We present results that clearly demonstrate that ab initio phasing is not limited to atomic-resolution data and have established an approach that provides outstandingly precise results, producing highly accurate electron density maps.

A major requirement and challenge for this deep-learning approach is scaling the neural network, that is, diffraction data for larger unit cells will require substantially larger input and output data and computational cost during the training. Centrosymmetric space groups, which were chosen for this work, have restricted phase values, whereas in space groups without inversion symmetry, the phases can largely have any value. However, there are no fundamental barriers that hinder this development. Further work is required to extend our method to the general case.

REFERENCES AND NOTES

1. G. Taylor, *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1881–1890 (2003).
2. J. Rius, *IUCrJ* **1**, 291–304 (2014).
3. C. A. Beevers, H. Lipson, *Nature* **137**, 825–826 (1936).
4. H. Hauptman, J. Karle, *Solution of the Phase Problem. I. The Centrosymmetric Crystal* (American Crystallographic Association, 1953).
5. H. Hauptman, J. Karle, *Acta Crystallogr.* **7**, 369–374 (1954).
6. D. Sayre, *Acta Crystallogr.* **5**, 60–65 (1952).
7. C. Giacovazzo, *Direct Phasing in Crystallography: Fundamentals and Applications* (International Union of Crystallography, 1998).
8. R. J. Morris, G. Bricogne, *Acta Crystallogr. D Biol. Crystallogr.* **59**, 615–617 (2003).
9. G. M. Sheldrick, *Acta Crystallogr. A* **46**, 467–473 (1990).
10. G. Oszlányi, A. Süto, *Acta Crystallogr. A* **60**, 134–141 (2004).
11. L. Palatinus, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **69**, 1–16 (2013).
12. G. Oszlányi, A. Süto, *Acta Crystallogr. A* **61**, 147–152 (2005).
13. G. Oszlányi, A. Süto, *Acta Crystallogr. A* **63**, 156–163 (2007).
14. G. Oszlányi, A. Süto, *Acta Crystallogr. A* **64**, 123–134 (2008).
15. G. Oszlányi, A. Süto, *Acta Crystallogr. A* **67**, 284–291 (2011).
16. M. C. Burla, C. Giacovazzo, G. Polidori, *J. Appl. Crystallogr.* **46**, 1592–1602 (2013).
17. M. C. Burla *et al.*, *J. Appl. Crystallogr.* **48**, 306–309 (2015).
18. M. C. Burla, C. Giacovazzo, G. Polidori, *J. Appl. Crystallogr.* **43**, 825–836 (2010).
19. M. C. Burla, B. Carrozzini, G. L. Cascarano, C. Giacovazzo, G. Polidori, *J. Appl. Crystallogr.* **44**, 1143–1151 (2011).
20. M. C. Burla, B. Carrozzini, G. L. Cascarano, C. Giacovazzo, G. Polidori, *J. Appl. Crystallogr.* **45**, 1287–1294 (2012).
21. A. van der Lee, *J. Appl. Crystallogr.* **46**, 1306–1315 (2013).
22. D. Rodríguez *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* **68**, 336–343 (2012).
23. J. Jumper *et al.*, *Nature* **596**, 583–589 (2021).
24. R. Junjuri, A. Saghi, L. Lensu, E. M. Vartiainen, *RSC Adv.* **12**, 28755–28766 (2022).
25. R. Houhou *et al.*, *Opt. Express* **28**, 21002–21024 (2020).
26. E. Cha, C. Lee, M. Jang, J. C. Ye, *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 9931–9943 (2022).
27. P. M. Vecsei, K. Choo, J. Chang, T. Neupert, *Phys. Rev. B* **99**, 245120 (2019).
28. Y. Suzuki *et al.*, *Sci. Rep.* **10**, 21790 (2020).
29. L. C. O. Tiong, J. Kim, S. S. Han, D. Kim, *NPJ Comput. Mater.* **6**, 196 (2020).
30. K. Kaufmann *et al.*, *Science* **367**, 564–568 (2020).
31. A. Ziletti, D. Kumar, M. Scheffler, L. M. Ghiringhelli, *Nat. Commun.* **9**, 2775 (2018).
32. W. B. Park *et al.*, *IUCrJ* **4**, 486–494 (2017).
33. C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179 (2016).
34. S. Gražulis *et al.*, *J. Appl. Crystallogr.* **42**, 726–729 (2009).
35. R. T. Downs, M. Hall-Wallace, *Am. Mineral.* **88**, 247–250 (2003).
36. D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, S. Rehme, *J. Appl. Crystallogr.* **52**, 918–925 (2019).
37. L. Palatinus, G. Chapuis, *J. Appl. Crystallogr.* **40**, 786–790 (2007).
38. V. Y. Lunin, M. M. Woolfson, *Acta Crystallogr. D Biol. Crystallogr.* **49**, 530–533 (1993).
39. R. Caliandro *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* **61**, 556–565 (2005).
40. C. Kalaiarasi, M. S. Pavan, P. Kumaradhas, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **72**, 775–786 (2016).
41. J. L. Pereira, R. R. Teixeira, S. Guilardi, D. A. Paixão, *Acta Crystallogr. Sect. E Struct. Rep. Online* **68**, o2995 (2012).
42. M. U. Schmidt *et al.*, *Angew. Chem. Int. Ed.* **50**, 7924–7926 (2011).
43. W. I. F. David *et al.*, *J. Appl. Crystallogr.* **39**, 910–915 (2006).
44. J. van de Streek, M. A. Neumann, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* **70**, 1020–1032 (2014).
45. S. Konar *et al.*, *Cryst. Growth Des.* **22**, 3862–3869 (2022).
46. C. Schlesinger *et al.*, *IUCrJ* **9**, 406–424 (2022).
47. A. S. Larsen, T. Rekis, A. Ø. Madsen, PhAI neural network for crystal structure determination, Version 2. Electronic Research Data Archive University of Copenhagen (2024); http://doi.org/10.17894/ucph.681ff154-a66a-41f4-aee6-9168777de7d9.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn2777
Materials and Methods

Supplementary Text
Figs. S1 to S17
Table S1
References (48–59)