A method for intuitively extracting macromolecular dynamics from structural disorder.

Nicholas M. Pearce\textsuperscript{1,2,*} and Piet Gros\textsuperscript{1}

\textsuperscript{1}Structural Biochemistry, Department of Chemistry, Faculty of Science, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, the Netherlands

\textsuperscript{2}Department of Chemistry and Pharmaceutical Sciences, Free University of Amsterdam, Amsterdam, the Netherlands

\*Correspondence and requests for materials should be sent to n.m.pearce@vu.nl

ABSTRACT Macromolecular dynamics manifest as disorder in structure determination, which is subsequently accounted for by displacement parameters (also called temperature factors, or B-factors) or alternate conformations. Though B-factors contain detailed information about structural dynamics, they are the total of multiple sources of disorder, making them difficult to interpret and thus little-used in structural analysis. We report here an approach for decomposing molecular disorder into a parsimonious hierarchical series of contributions, providing an intuitive basis for quantitative structural-dynamics analysis. We demonstrate the decomposition of disorder on example SARS-CoV-2 and STEAP4 structures, from both crystallographic and cryo-electron microscopy data, and reveal how understanding of the macromolecular disorder leads to deeper understanding of molecular motions and flexibility, and suggests hypotheses for molecular mechanisms.

Introduction

Macromolecular crystallography records a temporal and spatial average over billions of copies of a molecule in a crystal. Variations in the relative positions of these molecules lead to a blurring of the observed electron density\textsuperscript{2}. There are many sources of disorder within a crystal (Figure 1): crystal imperfections lead to a global contribution for the whole unit cell\textsuperscript{2}; static or dynamic molecular displacement leads to systematic disorder over whole molecules, or sections of molecules\textsuperscript{3}; and finally, atomic motions describe the individual dynamics of an atom relative to its surroundings. The same is true for cryo-electron microscopy data: while large-scale structural changes lead to images being separated into distinct classes, global errors in image alignment and local structural differences lead to disorder equivalent to that of crystallography. Disorder is thus a fundamental feature of macromolecular structural data. Conversely, disorder is a relatively untapped source of structural information: large-scale disorder obscures local disorder\textsuperscript{4}, thereby limiting any biological interpretations.

Local continuous disorder in an atomic model is described using atomic displacement parameters (ADPs; also called temperature...
factors, or B-factors). At high resolution, anisotropic ADPs (a-ADPs) describe "thermal ellipsoids" of disorder. At more moderate resolutions, we use isotropic ADPs (i-ADPs), often in combination with Translation-Libration-Screw (TLS) models that generate sets of a-ADPs describing collective rigid-body motions for groups of atoms. For lower resolutions, one i-ADP per residue or region are even possible. When B-factors are displayed, they are shown as spheres (for i-ADPs) or ellipsoids (for a-ADPs). The size of an ellipsoid in a particular direction is indicative of the disorder in that direction. More precisely, the surface of the sphere or ellipsoid is a probability contour, which contains the atom within the surface a chosen fraction of the time.

Comparative analysis of disorder between macromolecular structures is dominated by the empirical "B-factor normalisation" approach\[^1\], which allows only for qualitative, relative comparisons of i-ADPs\[^1\]. Additionally, while it is possible to extract physical motions from refined TLS descriptions\[^2\]\[^10\], such analyses may confound multiple sources of disorder (e.g. analysis of domain motions will also contain contributions from molecular disorder), and may not, in any case, result in physical motion\[^2\]. On the other hand, at high resolution, where models are most likely to contain atomic-level dynamical information, there are no general methods for deconstructing a-ADPs into interpretable components corresponding to different length-scales. New methods validate the global distributions and local values of model B-factors\[^11\]\[^2\] but refinement and validation approaches have understandably focussed on optimising the quality of the model, rather than producing an interpretable disorder model that is easily utilisable for structural analysis.

One convenient property of displacement parameters, \(U\), is that they are additive, and thus independent contributions can each be represented in the overall disorder model\[^3\]:

\[
U_{\text{total}} = U_{\text{crystal}} + ... + U_{\text{domain}} + ... + U_{\text{atomic}}. \tag{1}
\]

In this work, we present an extensible model that explicitly re-factors disorder into a hierarchical series of contributions. This physically-motivated formalism separates the different disorder contributions at different scales, as in \(U\), is generally applicable to both a- and i-ADPs, and enables a quantitative analysis of static/dynamic disorder in atomic structures. The disorder identified at different length scales reveals domain motions and loop flexibility which are likely linked to function, and this approach opens the door to new ways of visualising and understanding macromolecular structures.

**Results**

**A hierarchical disorder model**

Our new disorder model – the Extensible-Component Hierarchical TLS (ECHT) B-factor model – comprises a hierarchical series of TLS groups that describe disorder at different length-scales:

\[
U_{\text{total}} = \sum_{l=1}^{n_{\text{tls}}} U_{l}^{\text{tls}} + U_{\text{atomic}}. \tag{2}
\]

where \(n_{\text{tls}}\) is the number of TLS levels in the hierarchical model, \(U_{l}^{\text{tls}}\) are the TLS disorder contributions for level \(l\), and \(U_{\text{atomic}}\) is a set of i-/a-ADPs. The groups on each level are chosen to mirror the hierarchy of physical sources of macromolecular disorder (Figure 2): lower levels contain large-scale groups e.g. for each chain or domain, whilst higher levels contain smaller-scale groups, e.g. for each secondary structure element, residue or side-chain. For comparison, conventional TLS-refined models (one TLS level with an isotropic atomic component) can be retrieved by setting \(n_{\text{tls}} = 1\) and using an isotropic \(U_{\text{atomic}}\).

To parameterise an ECHT model, we utilise an elastic-net\[^4\]\[^13\] based approach which assigns disorder to a smaller scale (e.g. residue) only when it is incompatible with disorder at a

\[^1\] ADPs has previously been used to mean atomic displacement parameters or anisotropic displacement parameters. Here, we use ADPs for atomic displacement parameters, and a-ADPs or i-ADPs to refer to anisotropic ADPs or isotropic ADPs, respectively.
Figure 2: Hierarchical disorder model partitioning. A possible model partitioning is shown for six residues. The bottom (first) level contains a TLS group for all atoms, the second level into two TLS groups, and so on. The final (highest) level contains an ADP for each atom. The total ADP for an atom is given by summing the contributions from each level.

Figure 3: Optimisation and ECHT disorder profile of a structure of the SARS-CoV-2 main protease (7k3t). (a) The average B-factor of each level for each optimisation cycle. Optimisation begins with large weights on the number of model parameters, forcing disorder to be modelled only using large-scale groups, i.e. at the chain level. As optimisation cycles increase, the penalty on the number of model parameters decreases and disorder is increasingly allowed at smaller and smaller scales, improving model fit until the model converges. (b) The sum of the (squared) average B-factors for each independent model component (e.g. TLS group), representing model complexity, for each optimisation cycle. These values are used to penalise model complexity during elastic net optimisation. (c) The final ECHT disorder profile after optimisation, shown as the magnitude of the atomic B-factors averaged over each residue and coloured by level.

larger scale (e.g. molecule). This effectively minimises the number of parameters required to describe the disorder of a protein and thus produces a parsimonious model of disorder for the structure (see Methods). By separating out potentially mutually-confounding disorder components, these ECHT decompositions lead to quantitative structural models for molecular flexibility on different length scales.

Application to Structural Analysis

To demonstrate the method, we analysed the disorder patterns in a structure of considerable societal importance: that of crystal structure of the main SARS-CoV-2 main protease (Mpro: PDBID 7k3t; resolution 1.2Å; refined with individual a-ADPs). The elastic net optimisation of the ECHT model and the B-factor decomposition profile can be seen in Figure 3 and show that significant disorder is present at all levels (Supp. Table 1). Once these mutually-confounding components are separated, we identify flexibility in the regions surrounding the catalytic site where structural changes have been observed previously upon substrate binding (Figure 4). Moreover, the same key components are identified as flexible that are identified in molecular dynamics simulations, the P2 helix, the P5 loop, and the C-terminus. The disorder is also identified at the appropriate level, with the disorder for the P2 helix identified in the secondary structure level, while for the P5 loop the disorder is principally identified in the residue level.

Another Mpro structure, collected at room temperature, contains a dimer in the asymmetric unit, allowing us to compare the monomers’ disorder profiles (PDBID 6xhu; resolution 1.8Å; refined with TLS and individual i-ADPs; Supp. Figures 3). Overall, the ECHT components are highly correlated between the two monomers.
Figure 4: ECHT decomposition of a structure of the SARS-CoV-2 main protease (7k3t). The protein is shown as lines and ellipsoids, coloured by B-factor for each structure independently from blue (zero) to green to red (maximum); the symmetry-related copy creating the obligate homodimer is shown as semi-transparent surface. The binding site histidine (HIS 41) is also shown as a transparent surface. B-factor ellipsoids are contoured at $p=0.95$. (a) Re-refined structure from PDB_REDO14 (1.2Å resolution; $R_{work}/R_{free}$ 0.14/0.16; refined with a-ADPs; B-factors 7.7-77.6Å$^2$). (b-f) Disorder components of each ECHT level (maximum B-factor in brackets): (b) chain (7.4Å$^2$), (c) secondary structure (30.2Å$^2$), (d) residue (35.9Å$^2$), (e) backbone & sidechain (14.4Å$^2$), (f) atomic (19.0Å$^2$). (c,d) Flexible segments and residues line the sides of the catalytic site (see Supp. Figure 1). (c) The P2 helix, on the edge of the catalytic site, shows a large disorder component at the secondary structure level. The C-terminal also shows a large disorder component (Supp. Figure 2). (d) The P5 loop is also observed to be particularly flexible in the residue level. (e) Backbone motions are generally small, with the majority of disorder being isolated to the surface sidechains. (f) Atomic disorder highlights internal motions of residues that cannot be modelled by the rigid body approximation; these principally highlight longer sidechains and backbone carbonyls, as well as presumably absorbing inflated B-factors from modelling errors. Images rendered in pymol.
Table 1: Distribution of B-factors for ECHT decompositions of SARS-CoV-2 surface glycoprotein (6vxx & 6vyb). The columns contain the average B-factors for each ECHT level in both Å² and as a percentage of the total B-factor.

| Level         | 6vxx (closed) | 6vyb (open) |
|---------------|---------------|-------------|
| Molecule      | 12.4 (43.1%)  | 26.4 (55.1%)|
| Domain        | 10.3 (36.0%)  | 12.9 (27.0%)|
| Secondary     | 3.9 (13.4%)   | 5.8 (12.1%) |
| Structure     | 1.7 (5.8%)    | 2.1 (4.3%)  |
| Residue       | 0.5 (1.7%)    | 0.7 (1.5%)  |
| Atomic        |               |             |

(Supp. Figure 7), however some regions display significant differences: the region below the main catalytic site (residues 62-80) is significantly more flexible in the first monomer (chain A; c.f. Supp. Figures 3 and 4), despite the fact that this region forms crystal contacts with symmetry-related molecules in both molecules in the asymmetric unit. When a dimer level is included in the ECHT description, we can extract a collective dimer motion and individual monomer-level rocking motions around the dimer interface (Supp. Figure 8).

A third Mpro crystal structure (PDBID 6wqf; 2.3Å; individual i-ADPs), also collected at room temperature, again displays similar disorder patterns, including the disorder of the region around residues 62-80 (Supp. Figure 9), suggesting that this disorder reflects an intrinsic domain flexibility which is only visible in certain crystal forms, or at higher temperatures. As this structure was refined with isotropic B-factors, only the magnitude of the anisotropic TLS components are fitted against the input structure’s i-ADPs. While residues surrounding the binding site show flexibility comparable to that identified in both 7k3t and 6xlu, 6wqf shows significantly more secondary structure disorder (Supp. Table 1).

We subsequently applied the method to another crucial SARS-CoV-2 protein, the surface glycoprotein spike protein, in both closed and open conformations (PDBIDs 6vxx and 6vyb; 2.8Å and 3.2Å; individual i-ADPs), which were determined using cryo-EM (Figure 5). In each of these, the ECHT description contains a level for the whole homotrimer, as well as for each sub-domain (for domain definitions see Supp. Table 2). Supp. Figure 11, which were identified by manual inspection of the structures. As expected, given the resolution of the structures, disorder is assigned mostly to large-scale disorder in the ECHT profile, with the molecule and domain levels accounting for 79% (6vxx) and 82% (6vyb) of the total disorder (Table 1).

The molecular ECHT levels show a gradient of disorder over the model (likely reflecting image alignment uncertainties) and at the tips of the structures, this component masks smaller-scale disorder components. Accounting for and removing this global disorder component allows for interrogation of the region surrounding the receptor-binding domains (RBDs), which are required for the binding to ACE2 receptors. In the closed conformation (6vxx), the ECHT domain level shows clear hinge-like disorder components that reveal the propensity for RBDs to adopt an alternate conformation. In the open conformation (6vyb), one RBD is observed in an extended conformation, and is highly disordered at both the domain and the secondary-structure level, indicating that the domain is highly flexible. However, the disorder patterns of the remaining two closed RBDs also show significant differences: one RBD (chain C) maintains the distinct hinge-like disorder pattern, while the other (chain A) now displays a smaller, more homogenous, disorder profile at the domain level. Away from the RBDs, however, equivalent atoms in each part of the trimer demonstrate very similar disorder components at each level (Supp. Figure 14).

We next analysed the ECHT profiles of two cryo-EM structures of the Six-Transmembrane Epithelial Antigen of the Prostrate (STEAP4; PDBIDs 6hcy & 6hd1; resolutions 3.1Å & 3.8Å; both refined with one i-ADP per residue). 6hcy was determined from data where Fe³⁺-NTA molecules were present in the substrate binding site, no iron molecule was modelled; only the iron was certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.
Figure 5: ECHT decompositions of SARS-CoV-2 surface glycoprotein (6vxx & 6vyb). Atoms are shown as lines and ellipsoids, coloured by B-factor for each structure independently from blue (zero) to green to red (maximum). B-factor ellipsoids are contoured at $p=0.95$. Side and top views of trimer for (a-d) the closed state (6vxx), and (e-h) the open state (6vyb). a) Deposited structure for 6vxx (2.8Å resolution; individual i-ADPs, B-factors 3.5-102.4Å$^2$). (b-d) Disorder components of each ECHT level (maximum B-factor in brackets): (b) molecule (32.0Å$^2$), (c) domain (45.5Å$^2$), (d) secondary structure (20.7Å$^2$). Not shown (see extended data): residue level (12.2Å$^2$) & atomic level (37.0Å$^2$). (e) Deposited structure for 6vyb (3.2Å resolution; individual i-ADPs, 16.4-181.1Å$^2$). (b-d) Disorder components of each ECHT level (maximum B-factor in brackets): (b) molecule (67.2Å$^2$), (c) domain (85.3Å$^2$), (d) secondary structure (44.3Å$^2$). Not shown (see extended data): residue level (28.4Å$^2$) & atomic level (39.5Å$^2$). (g) The molecular components for both structures show similar profiles, with a gradient of disorder across the molecule. (c) The domain-level disorder reveals hinge-like components for the closed RBDs. (g) The domain-level disorder reveals conservation of the hinge-like disorder for chain C, but a loss of this component for chain A. Inset images show the disorder for the RBDs only. (d,h) The secondary structure level reveals intra-domain flexibility. Images rendered in pymol.
no iron was added in 6hd1. The detail in these disorder models is naturally limited by the resolution and B-factor model, but the ECHT disorder analysis reveals a large number of interesting features. Expected features include a large homogenous global disorder component (Figure 6b), and increased disorder of the globular intracellular domain (Figure 6c). Additionally, at the secondary structure level, increased flexibility is clearly visible for the two outermost transmembrane helices (Figure 6d) and for loops around the substrate binding site (Figure 6f & 6g; Supp. Figure 17), which is likely necessary for substrate recognition and binding, since Fe^{3+}/Cu^{2+} ions must navigate through rings of alternating positive and negative charges around the binding site to be reduced by the integral HEME molecule\textsuperscript{19}. Furthermore, disorder in the secondary structure and residue levels supports a potential mechanism for electron transport, whereby shuttling of the bound FAD molecule facilitates transfer of electrons between the NADPH and HEME molecules\textsuperscript{19} (Figure 7).

Lastly, the intracellular domains form a tightly-ordered trimeric interface, whilst the other half of the domain is more flexible, suggesting a “molten” sub-domain whose flexibility may be related to turnover of NADPH molecules bound to the intracellular domain (Supp. Figures 19-20). All these features are also present in the related structure 6hd1 (c.f. Supp. Figures 15-21).

Comparison of different temperatures

To further investigate the ECHT decompositions, we analysed 30 pairs of structures collected at room- and at cryogenic temperatures\textsuperscript{20}, and looked at the changes in disorder patterns associated with the change in temperature. The ECHT decompositions generally reveal an increase in disorder at all scales between the pairs of structures, in line with our expectation that the “warming up” of the protein affects all scales of motion (Supplementary Figure 22). Several notable exceptions show an increase in all levels except the chain level, which decreases. This is likely because the parsimonious model reduces the common disorder component to the lowest possible level, and as the disorder profile becomes less homogeneous at higher temperatures, the parsimonious model now requires reassignment of disorder to higher levels to account for distinct motions. This reassignment evidently necessitates a reduction in the disorder of the chain level in some cases.
Discussion

By averaging over thousands to billions of molecules, macromolecular diffraction and microscopy data contain detailed information about molecular dynamics, but this information is obscured by experimental artefacts, and the layers-upon-layers of motions that make the total disorder pattern impossible to interpret in terms of individual components. In this work, we have presented a physically-motivated approach for decomposing the disorder in an atomic structure into interpretable components. This generalises existing TLS approaches to multiple length-scales, presents a general framework for disorder decomposition that applies for both isotropic and anisotropic disorder, and enables quantitative structural analysis.

The advent of cryo-EM raises the exciting opportunity to study molecules outside the context of a crystalline environment, but there clearly remains an opportunity to study macromolecular dynamics in crystals. Both large-scale and small-scale disorder contain useful information for structural analysis: large-scale molecular disorder dictates how well molecules are resolved and reveal domain motions (Figure 5); secondary structure-level disorder helps form hypotheses for structural mechanisms (Figure 6); and small-scale disorder details the rigidity/plasticity of binding sites (Figure 4). Since this disorder is now characterised on an absolute scale (describing the fluctuations of an atom in Å), the relevant components can be used in a variety of quantitative structural and bioinformatic applications.

The output ECHT model is naturally dependent on the choice of levels, and how these levels are partitioned (e.g. choice of secondary structure). Therefore, groups must be chosen carefully and critical analysis of the model is essential. Conversely, through the elastic net approach, redundant or overly-partitioned levels are generally penalised, providing a measure of robustness. The quality of the input model and the refinement protocol will also have an effect on interpretation: regions of particular interest should be checked to ensure they are of high quality, and overly strong B-factor restraints must be avoided,
which might cause atoms to have overly-similar disorder profiles, thereby channelling disorder to larger scales than the data might otherwise suggest. Furthermore, when optimising an ECHT model against models refined with i-ADPs, multiple TLS parameter sets may provide a similar fit to the input i-ADPs. Currently, visual analysis of the output model is necessary to ensure it is physically reasonable, and does not contain large changes in the magnitude or direction of disorder between adjacent residues. The addition of appropriate model restraints to enforce these restrictions, as well as the implementation of symmetry constraints, remains an avenue of future work.

Naturally, a disorder component at a particular level does not necessarily imply correlated motion, only that atoms have compatible disorder profiles; correlated motion can only be confirmed by orthogonal methods, such as diffuse scattering (21). The use of TLS-described rigid-body motions is also a simplification of the continuous motions that proteins will undergo in reality, however, we have shown here that layers of rigid-like components can still be used to gain insight into the magnitude and localisation of macromolecular disorder patterns. The elastic-net optimisation approach is also naturally applicable to any other disorder formalism.

Methods

Input Structures

The ECHT disorder model is fitted to a refined atomic structure. Input structures can be parameterised with either i-ADPs or a-ADPs, or a mixture. For TLS-refined structures, the TLS contribution can either be included in the atomic ADPs (resulting in a-ADPs), or excluded (resulting in i-ADPs).

Hierarchical Model Partitioning

Default TLS group partitions are created for each: protein chain; local secondary structure element; residue; and residue backbones/side chains. Secondary structure is automatically identified with the DSSP algorithm (22) as implemented within the cctbx (23). C3 atoms are included with backbone atoms. Glycine, Alanine and Proline residues are not considered in the backbone/sidechain levels. An “atomic” level is created with an a-ADP or i-ADP (depending on the input disorder model) for each atom. Part of an example composition is shown in Supplementary Figure 24. All ECHT levels mentioned in this work use the above model composition unless stated otherwise. Custom levels can also be defined, such as for individual domains within a chain, or across chains. To aid interpretability, levels should generally be ordered such that the largest groups appear at lower levels.

ECHT Model Parameterisation

The overall schema of the parameterisation process is shown in Supp. Figure 25. In each macrocycle, the algorithm alternates between optimising the TLS parameters for each level using a simplex-based approach, and optimising the amplitudes of all model components (TLS groups for TLS levels, and ADP values for the atomic level), using a gradient-based method with elastic net penalties. Between the macrocycles, elastic net penalties are decayed by a chosen factor (0-1). Smaller values lead to a faster runtime but may miss model components (e.g. lead to a non-parsimonious model); large values lead to a longer runtime, but a much more detailed model. In this work, the decay factor is chosen to be 0.8, which was found to be a good balance between speed and model quality. Some minor differences may appear between symmetry-related monomers in a model; these can be a good indicator that the decay factor is too small. Within each macrocycle, each series of TLS and amplitude optimisations is repeated until the amplitudes of the components stop changing between each microcycle.
Target function

All optimisation uses a least-squares target function,
\[ \sum_a \omega_a \cdot (U_{\text{target}} - U_{\text{model}}) \text{,} \]
where \( \omega_a \) is the weight for atom \( a \). For amplitude optimisation, the input \( U \) are used as \( U_{\text{target}} \). For the optimisation of individual TLS group parameters, the target \( U \) values for level \( k \) are given by
\[ U_{\text{target}}^k = U_{\text{input}} - \sum_{l \neq k} U_{\text{model}}^l \text{,} \]
where \( U_{\text{model}}^l \) are the currently parameterised \( U \) values for level \( l \), and the sum is over levels not currently being optimised.

In this work, the weights used are
\[ \omega_a \propto \frac{1}{|U_{\text{input}}^a|} \text{,} \]
which reduces the effect of atoms with large B-factors on the optimisation of large-scale levels – for which they contain little information – during the initial cycles. Input weights are normalised such that
\[ \sum_a (\omega_a) = 1 \text{.} \]

TLS-group normalisation

The \( U \) values arising from a TLS group are rewritten as
\[ U_{\text{tls}}^g = A_{\text{tls}}^g \cdot \tilde{U}_{\text{tls}}^g \text{,} \]
for group \( g \). The scaling between \( A_{\text{tls}}^g \) and \( \tilde{U}_{\text{tls}}^g \) is chosen such that
\[ \frac{1}{n_{\text{atoms}}^g} \sum_{a \in g} 1 \cdot \text{Tr} \left( \left[ U_{\text{tls}}^g \right]_a \right) = 1 \text{,} \]
where the sum \( a \in g \) is over all atoms that belong to group \( g \), and \( n_{\text{atoms}}^g \) is the number of atoms in group \( g \). This normalisation scales the TLS-matrix values so that the amplitudes \( A_{\text{tls}} \) are proportional to the average B-factor of each group.

Note, however, that an amplitude of \( A_{\text{tls}} = 1 \AA^2 \) corresponds to an average B-factor of \( 8\pi^2 \AA^2 \approx 72\AA^2 \) for this group.

TLS-level optimisation

TLS matrices are initialised with an isotropic T-matrix, \( \text{diag}(1,1,1) \), and zero-value L-, and S-matrices. Amplitudes are initially set to zero, and become non-zero during amplitude optimisation (see below). TLS-matrices are optimised using a simplex search method implemented in cctbx\(^{23} \), the starting simplex for optimisation of the matrix values is obtained by independently incrementing each normalised matrix component (6 T-elements, 6 L-elements, 8 S-elements) in the coordinate basis of the L-matrix. During simplex optimisation, invalid combinations of TLS matrix parameters are rejected using the TLS-decomposition protocol described in Urzhumtsev et al. 2011\(^{22} \). After matrix optimisation, the TLS matrices and amplitudes are renormalised according to \( \text{TLS-matrices} \).

Atomic-level optimisation

Atomic-level optimisation is performed using the same least-squares target and simplex search method as the TLS levels with the constraint that the resulting ADP is positive semi-definite to within some tolerance \( \epsilon \).

Inter-level amplitude optimisation

After each component optimisation, the magnitudes of all TLS groups and atomic ADPs are optimised using the lbfgs algorithm implemented in cctbx\(^{23} \), with the constraint that no amplitudes can be negative. The amplitudes for TLS groups are defined in \( \text{TLS-matrices} \). The amplitudes of each atom are defined similarly as \( A_{\text{atom}} = \left| U_{\text{atom}}^a \right| = \frac{1}{2} \text{Tr}(U_{\text{atom}}^a) \); the atomic normalisation then becomes
\[ U_{\text{atom}}^a = A_{\text{atom}}^a \cdot \tilde{U}_{\text{atom}}^a \text{.} \]

To minimise the number of parameters needed to describe the disorder, the following elastic-net penalties\(^{23} \) are added to the target function:
\[ \Omega^\alpha \cdot \left[ \sum_g A_{\text{tls}}^g + \sum_a A_{\text{atom}}^a \right] \text{,} \]
and
\[ \Omega^\beta \cdot \left[ \sum_g (A_{\text{tls}}^g)^2 + \sum_a (A_{\text{atom}}^a)^2 \right] \text{,} \]
where \( \Omega^\alpha \) and \( \Omega^\beta \) are the lasso (sum of amplitudes) and ridge-regression (sum of squared amplitudes) weights, respectively. These can be redefined in terms of a mixing parameter
\[ \gamma = \Omega^\alpha / (\Omega^\alpha + \Omega^\beta) = \Omega^\alpha / \gamma_0 \text{,} \]
where \( \gamma = 1 \) is lasso regression and \( \gamma = 0 \) is ridge regression. Larger values of \( \gamma \) bias towards the parsimonious model, whilst smaller values lead to more non-zero components. In this work a mixing value of \( \gamma = 0.9 \) is used.

At the end of every optimisation macrocycle, the elastic-net weights are reduced by a factor \( \delta \):
\[ \Omega_{n+1} = \Omega_n \cdot \delta \text{,} \]
where \( n \) is the macro-cycle number and \( \delta \) is a weight decay factor between 0 and 1.
Convergence
Optimisation cycles continue until the $U$ values for all atoms are changing by less than $\Delta B_{\text{cutoff}}$ over a fraction of the total cycles. Additionally, a cutoff can be given which terminates the optimisation when a threshold is reached for the RMSD between the input $U$ and $U_{\text{model}}$. Note that when the atomic level is not included in the optimisation, there is no guarantee that a given RMSD cutoff will ever be reached.

Program Output
The implementation outputs the disorder contributions for each level as separate structures. A large number of analytical graphs are automatically generated to allow the visual analysis of the disorder; these are combined into one html results file for ease of use.

Implementation & Data Availability
This method as described above is implemented within the PanDEMIC project (Pan-Dataset Ensemble Modelling of Iso-structural Crystals; named prior to the ongoing coronavirus pandemic) and the script panddas python package [https://panddas.bitbucket.io]. The program will be distributed within CCP4 but can equally be installed within a phenix environment with an up-to-date cctbx installation. Decompositions presented in the results sections will be uploaded to Zenodo upon publication but can now be found at https://surfdrive.surf.nl/files/index.php/s/W4y3Rqgpyp1164E0 Any further data presented in this paper are available from the authors upon reasonable request.

Acknowledgements
We particularly thank Helen Ginn, James Parkhurst and Billy Poon for their help and guidance with c++ and cctbx. We also thank Loes Kroon-Batenburg, Tim de Klijn and Jitse van der Horn for useful general discussions regarding disorder, and the coronavirus structural taskforce (insidecorona.net) for discussions of coronavirus structures. NMP recognises funding from a long-term fellowship from the European Molecular Biology Organisation (EMBO), fellowship number ALTF 609-2017, and from a NWO Veni Fellowship, fellowship number VI.Veni.192.143.

Author Contributions
NMP and PG designed the research and wrote the manuscript. NMP developed and implemented the algorithm, and performed all analyses.

References
1. Rupp, B. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology (Garland Science, 2010).
2. Sherif, S. et al. Description of overall anisotropy in diffraction from macromolecular crystals. Acta Crystallogr. Sect. A 43, 118–121 (1987).
3. Winn, M. D. et al. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. Sect. D Biol. Crystallogr. 57, 122–133 (2001).
4. Carugo, O. Atomic displacement parameters in structural biology. Amino Acids 50, 775–786 (2018).
5. Trueblood, K. N. et al. Atomic Displacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature. Acta Crystallogr. Sect. A Found. Crystallogr. 52, 770–781 (1996).
6. Schomaker, V. et al. On the rigid-body motion of molecules in crystals. Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem. 24, 63–76 (1968).
7. Painter, J. et al. TLSMD web server for the generation of multi-group TLS models. J. Appl. Crystallogr. 39, 109–111 (2006).
8. Parthasarathy, S. et al. Analysis of temperature factor distribution in high-resolution protein structures. Protein Sci. 6, 2561–2567 (2008).
9. Urzhumtsev, A. et al. From deep TLS validation to ensembles of atomic models built from elemental motions. Acta Crystallogr. Sect. D Biol. Crystallogr. 71, 1668–1683 (2015).
10. Afonine, P. V. et al. From deep TLS validation to ensembles of atomic models built from elemental motions. II. Analysis of TLS refinement results by explicit interpretation. Acta Crystallogr. Sect. D Struct. Biol. 74, 1668–1683 (2018).
11. Masmaliyeva, R. C. et al. Analysis and validation of macromolecular B values. Acta Crystallogr. Sect. D Struct. Biol. 75, 505–516 (2019).
12. Masmaliyeva, R. C. et al. Local and global analysis of macromolecular atomic displacement parameters. Acta Crystallogr. Sect. D Struct. Biol. 76, 926–937 (2020).
13. Zou, H. et al. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B (Statistical Methodol. 67, 301–320 (2005).
14. Joosten, R. P. et al. PDB\_REDO: automated re-refinement of X-ray structure models in the PDB. J. Appl. Crystallogr. 42, 376–384 (2009).
15. Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.8 2015.
16. Kneller, D. W. et al. Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. Nat. Commun. 11, 3202 (2020).
17. Kneller, D. W. et al. Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: insights into enzyme mechanism and drug design. IUCrJ 7, 1028–1035 (2020).
18. Walls, A. C. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281–292.e6 (2020).
19. Oosterheert, W. et al. Cryo-EM structures of human STEAP4 reveal mechanism of iron(III) reduction. Nat. Commun. 9, 4337 (2018).
20. Fraser, J. S. et al. Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc. Natl. Acad. Sci. 108, 16247–16252 (2011).
21. De Klijn, T. et al. Rigid-body motion is the main source of diffuse scattering in protein crystallography. IUCrJ 6, 277–289 (2019).
22. Kabsch, W. et al. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637 (1983).
23. Grosse-Kunstleve, R. W. et al. The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. J. Appl. Crystallogr. 35, 126–136 (2002).
24. Winn, M. D. et al. Overview of the CCP 4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 235–242 (2011).
25. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221 (2010).