Molecular nucleation mechanisms and control strategies for crystal polymorph selection

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

LETTER

Molecular nucleation mechanisms and control strategies for crystal polymorph selection

Alexander E. S. Van Driessche[1], Nani Van Gerven[5,6], Paul H. H. Bomans[2,3], Rick R. M. Joosten[2,3], Heiner Friedrich[2,3], David Gil-Carton[4], Nico A. J. M. Sommerdijk[2,3] & Mike Sleutel[5,6]

The formation of condensed (compacted) protein phases is associated with a wide range of human disorders, such as eye cataracts[1], amyotrophic lateral sclerosis[2], sickle cell anaemia[3] and Alzheimer's disease[4]. However, condensed protein phases have their uses: as crystals, they are harnessed by structural biologists to elucidate protein structures[5], or are used as delivery vehicles for pharmaceutical applications[6]. The physiochemical properties of crystals can vary substantially between different forms or structures ('polymorphs') of the same macromolecule, and dictate their usability in a scientific or industrial context. To gain control over an emerging polymorph, one needs a molecular-level understanding of the pathways that lead to the various macroscopic states and of the mechanisms that govern pathway selection. However, it is still not clear how the embryonic seeds of a macromolecular phase are formed, or how these nuclei affect polymorph selection. Here we use time-resolved cryo-transmission electron microscopy to image the nucleation of crystals of the protein glucose isomerase, and to uncover at molecular resolution the nucleation pathways that lead to two crystalline states and one gelled state. We show that polymorph selection takes place at the earliest stages of structure formation and is based on specific building blocks for each space group. Moreover, we demonstrate control over the system by selectively forming desired polymorphs through site-directed mutagenesis, specifically tuning intermolecular bonding or gel seeding. Our results differ from the present picture of protein nucleation[7–12], in that we do not identify a metastable dense liquid as the precursor to the crystalline state. Rather, we observe nucleation events that are driven by oriented attachments between subcritical clusters that already exhibit a degree of crystallinity. These insights suggest ways of controlling macromolecular phase transitions, aiding the development of protein-based drug-delivery systems and macromolecular crystallography.

How do protein crystals nucleate? What is or are the pathway(s) from isolated protein molecules to mesoscopic and finally macroscopic crystals? There have been three independent nanometre-scale observations of protein nucleation at solid–liquid interfaces[13–15], revealing both direct and indirect pathways, but these works used atomic force microscopy—a surface technique that is blind to events taking place within the liquid. In another approach, *in situ* liquid-cell transmission electron microscopy was used to map the nucleation pathways of calcium carbonate[16] and, more recently, of the protein lysozyme[17], but that technique currently lacks the lateral resolution needed to resolve the structure of the nuclei and the particles that precede them.

To obtain an experimental window onto the formation of a crystal nucleus in liquid at molecular resolution, we use cryo-transmission electron microscopy to image vitrified samples that have been plunge frozen at various time intervals. We study the nucleation pathways of glucose isomerase, a protein with applications in the biofuel and food

industries as a crystalline suspension[18]. Depending on the solution conditions, glucose isomerase can crystallize into at least two different space groups[19], or (as we show here) can aggregate into a disordered, gelled state. Using ammonium sulfate as a precipitant, we find that the protein exhibits a polymorph transition from an $I222$ (rhombic) to a $P2_12_12$ (prismatic) space group as a function of the precipitant concentration (Extended Data Fig. 1). Turbidity measurements reveal that the induction time for nucleation decreases exponentially as the ammonium sulfate concentration increases from 1.2 M to 1.65 M (Extended Data Fig. 2). However, no conditions are identified that lead to liquid–liquid phase separation or gelation. Cryo-transmission electron microscopy (cryo-TEM) imaging of the earliest quenched sample (plunge frozen after 20 seconds) in 1.5 M ammonium sulfate (a mixed $I222/P2_12_12$ condition) shows the presence of elongated particle assemblies ('nanorods'; Fig. 1a–e). Given the overall particle dimensions and the electron-microscopy silhouette of the subunits, we identify the building blocks of these nanorods to be single protein molecules (Fig. 1a–c). The nanorods are on average two molecules in width (1.7 ± 1; $n = 60$) and 12 molecules in length (12.4 ± 5; $n = 60$), with an intermolecular distance of 8.2 ± 0.1 nm ($n = 51$) along the long axis (Fig. 1d and Extended Data Fig. 3). Single-file protein chains are also observed (Fig. 1c), as well as trimers, tetramers and larger polymers at later time points (10–20 min; Fig. 1e). Although successive images show a gradual increase in the nanorod concentration as a function of time (Fig. 1a, b; see Extended Data Fig. 4 for the dependence on ammonium sulfate concentration), there is no increase in their length (Extended Data Fig. 3g).

At around 15 to 30 minutes after protein–precipitant mixing, larger structures begin to emerge. We detect (sub)micrometre-sized fibres of 43 ± 7 nm ($n = 88$) in width (Fig. 1f). The molecular columns that run along the fibre axis have a characteristic centre-to-centre distance of 8.0 ± 0.1 nm ($n = 27$), in line with the spacing measured for the nanorods. The associated two-dimensional fast Fourier transform (2D-FFT) image does not show sharp diffraction spots, but rather diffraction arcs in one or two directions (Fig. 1f, inset). Such arcs indicate that there is local ordering, but also substantial deviation from the crystallographic directions. The aspect ratio of these fibres ranges from 10 to 30, a considerable increase with respect to the aspect ratio of the nanorods (around 6), indicating that fibre broadening is slow compared with elongation. We also see bundles of individual fibres that are making loose lateral contacts with each other (Fig. 1g, h). These bundles have varying degrees of misalignment at the interfibre level, leading to different levels of disorder.

Within the same time frame, faceted nanocrystals start to appear, with morphologies and intermolecular distances that fit the $P2_12_12$ and $I222$ space groups (Fig. 2a–e and Extended Data Table 1). The crystallinity of both particle types is reflected in the emergence of sharp diffraction spots in the 2D-FFT. The smallest observed rod-like $P2_12_12$

[1]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, F-38000 Grenoble, France. [2]Laboratory of Materials and Interface Chemistry and Center of Multiscale Electron Microscopy, Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands. [3]Institute for Complex Molecular Systems, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands. [4]Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia, 48160 Derio, Bizkaia, Spain. [5]Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [6]Structural and Molecular Microbiology, Structural Biology Research Center, VIB, Pleinlaan 2, 1050 Brussels, Belgium.
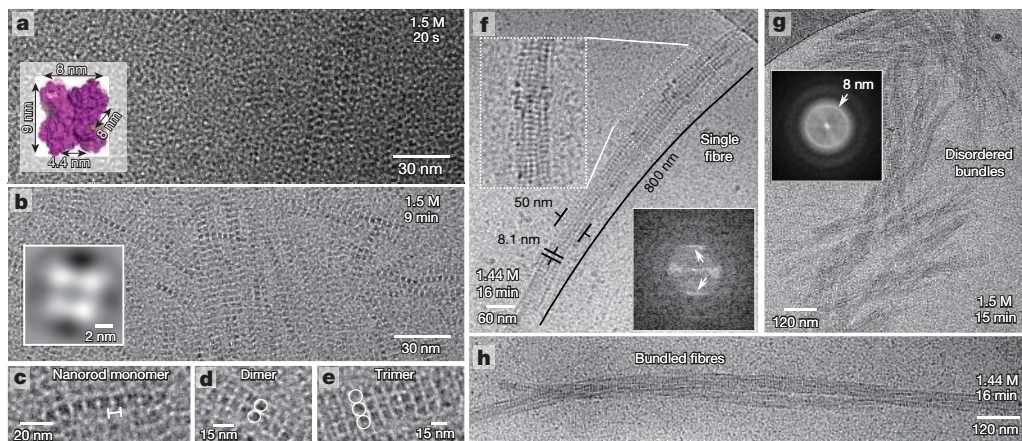
Figure 1 | Pre-nucleation assemblies of glucose isomerase induced by ammonium sulfate. a, b, Cryo-TEM image of a glucose isomerase solution in 1.5 M ammonium sulfate, at 20 seconds (**a**) and 9 minutes (**b**) after mixing. The inset in panel **a** shows a van der Waals representation of a glucose isomerase molecule. The inset in panel **b** shows a class-average electron microscopic image of the nanorods' building blocks. **c–e,** Magnified images of a nanorod monomer (**c**), dimer (**d**) and trimer (**e**) composed of glucose isomerase molecules, with a centre-to-centre distance of 8.2 ± 0.1 nm ($n = 51$, marked in panel **c**). **f–h,** Emergence of larger structures (at 15–30 min). **f,** A micrometre-sized fibre, with the upper inset being a magnified image and the lower inset being the corresponding 2D-FFT image, showing diffraction arcs in two directions. **g, h,** Bundling of fibres into aggregate structures with varying degrees of alignment. The inset in panel **g** is a 2D-FFT image.

crystal measures 380 nm by 120 nm (aspect ratio = 3), and has a width that exceeds those of the fibres (Fig. 2a). The characteristic interplanar spacing parallel to the long crystal axis is 8.1 ± 0.1 nm ($n = 26$), again in line with the value obtained for the nanorods and fibres. The nanorod alignment parallel to the nearest facet of the crystallite in Fig. 2b suggests that oriented attachment is a mode of incorporation into the crystalline phase. For the rhombic *I*222 space group, the smallest crystals that we find have an edge length of ± 100 nm (Fig. 2d, e) and characteristic distances of 5 nm and 7 nm.

With polyethylene glycol (PEG$_{1000}$ or PEG$_{1500}$) as the precipitant, glucose isomerase exhibits a similar polymorph transition from rhombic to prismatic crystals, albeit over a relatively narrow PEG concentration range (Extended Data Figs 1, 5). However, the highest PEG concentration produces a different effect to the highest ammonium sulfate concentration, in that glucose isomerase solidifies rapidly into a kinetically arrested gelled state (Extended Data Fig. 6). Cryo-TEM imaging of a range of PEG conditions reveals striking similarities to the nucleation pathways observed with ammonium sulfate. At 5% (w/v) PEG$_{1000}$ (an *I*222-only condition), we detect only mesoscale, rhombic crystals that exhibit fringe patterns compliant with the expected interplanar spacing of the *I*222 space group (Fig. 3a and Extended Data Table 1). Under conditions that lead to nucleation of both of the space groups and the gel (86 mg ml^{-1}, 4.5% (w/v) PEG$_{1500}$; Extended Data Fig. 5b), fibre-like structures appear 2–3 minutes after protein/precipitate mixing; these structures have a characteristic intermolecular distance of 8.0 ± 0.2 nm ($n = 24$) along the long axis, and measure 41 ± 6 nm ($n = 166$) in width (Fig. 3b–e). Interestingly, we observe no nanorods in any of the time points for this sample series, or for any glucose isomerase/PEG sample (Fig. 3b–e). At later time points, we see the grouping of these fibres into structures of increasing dimensions, exhibiting lateral stacking of individual fibres but still separated by a thin solvent layer (Fig. 3d, e). Identical sample replicates that failed to crystallize, but ended up in the gel state, reveal the presence of fibres that are morphologically similar to those described above
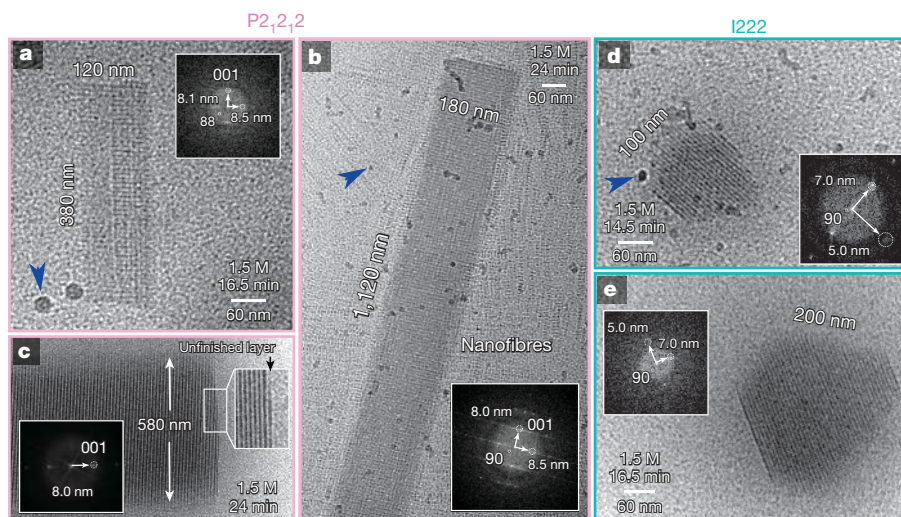
Figure 2 | Emergence of faceted prismatic and rhombic nanocrystals 15–30 minutes after mixing glucose isomerase and ammonium sulfate. a, A $P2_12_12$ nanocrystal. Inset, 2D-FFT image, with diffraction spots (measured 8.1 nm × 8.5 nm, 88°; theoretical (110) plane 7.8 nm × 8.1 nm, 90°). **b,** Larger $P2_12_12$ crystal displaying higher-order diffraction. **c,** Mature $P2_12_12$ crystal with a clear fringe pattern with a spacing of 8 nm). The magnified image of the facet edge resolves an unfinished molecular layer, suggesting that crystal growth at this point proceeds by incorporation of single glucose isomerase molecules. **d, e,** The smallest detected rhombic, crystalline objects display diffraction spots in two directions, that is, 7 nm and 5 nm at 90° (theoretical (101) plane 6.7 nm × 5 nm; (011) plane 7.3 nm × 4.3 nm). Blue arrows show contamination by ethane.
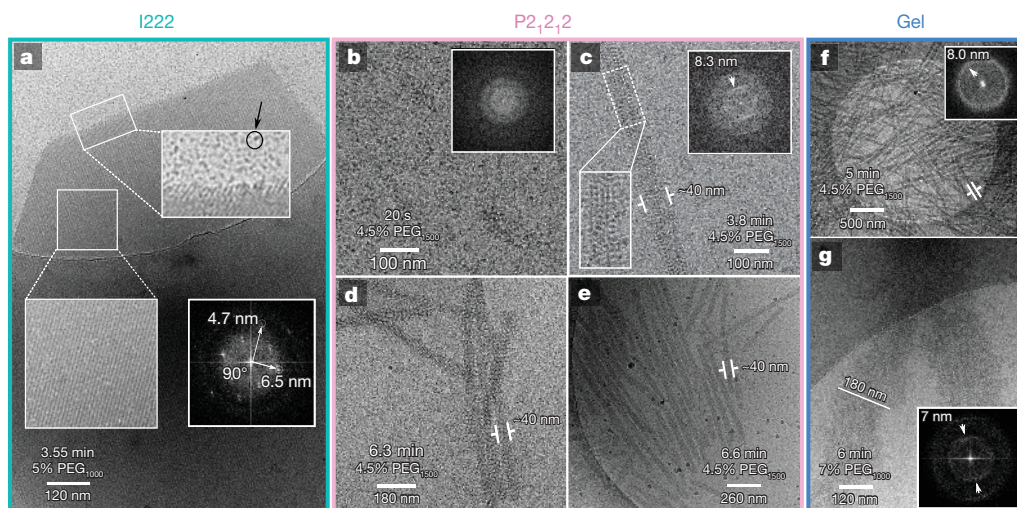
Figure 3 | Cryo-TEM imaging of PEG-induced crystallization of glucose isomerase. a, Micrometre-sized rhombic crystals obtained 4 minutes after mixing glucose isomerase and PEG at a low depletion attraction (in 5% PEG$_{1000}$). Lower left inset, a magnified image of the resolved crystal lattice. Lower right inset, the corresponding 2D-FFT image. Upper inset, magnification of the upper facet edge, resolving free monomers in solution. **b–e**, Time series of a glucose isomerase/PEG sample with 4.5% PEG$_{1500}$ reveals a $P2_12_12$ nucleation pathway reminiscent of that observed with ammonium sulfate (Fig. 1), but with a clear absence of nanorods at all measured time points. **f, g**, Formation of a glucose isomerase/PEG hydrogel at an intermediate depletion attraction (**f**, with 4.5% PEG$_{1500}$) and a high depletion attraction (**g**, with 7% PEG$_{1000}$).

(of width 44 ± 5 nm; $n = 100$), but at drastically higher concentrations (Fig. 3f). With higher depletion–attraction forces (through the use of 7% PEG$_{1000}$) but lower protein concentrations (37.5 mg ml^{-1})—conditions that slow down the rate of aggregation—cryoTEM imaging before kinetic arrest (6 min after protein/PEG mixing) reveals the formation of fractal-like aggregates with a distinct lack of rotational order. This can also be seen from the two concentric arcs in a 2D-FFT image (Fig. 3g). The FFT image is reminiscent of those of the disordered fibre bundles observed in the ammonium sulfate experiments, but shows higher packing density, as indicated by the 7.0 ± 0.2 nm ($n = 16$) spacing (Fig. 3g) and broader fibre cross-section (of width 100 ± 50 nm; $n = 20$).

The striking resemblance at the microscopic level between the $P2_12_12$ crystallization pathways (Fig. 1, with ammonium sulfate; or Fig. 3b–e, with PEG) and the gelation pathway strongly suggests that both phases originate from the same precursor states (that is, fibres). This observation prompted us to perform seeding experiments using glucose isomerase/PEG hydrogels, to see whether we could selectively elicit $P2_12_12$ crystals. For this, we transferred a glucose isomerase/PEG hydrogel fragment to a freshly prepared mother liquor solution that leads exclusively to $I222$ crystals (Fig. 5a). Time-lapse imaging of the solution–gel interface reveals the rapid and exclusive formation of $P2_12_12$ crystals on, or protruding from, the gel phase, demonstrating that gels can be used as polymorph-specific seeding agents. If the hydrogel is instead transferred to a similar $I222$-exclusive condition but a lower PEG concentration (3% (w/v) PEG$_{1500}$), then the gel phase gradually dissolves as $P2_12_12$ crystals emerge over time (Fig. 5f, g).

Both the early-stage nanorods (formed in high ammonium sulfate concentrations) and the later-stage fibres (formed in high ammonium sulfate or PEG) exhibit high aspect ratios, suggesting that there are substantial differences in lattice-contact strengths ($|C_i|$) for these conditions. To understand the origins of these differences, we analyse the mode of intermolecular bonding within the nanorod structure and compare it with known crystal lattice contacts. Using the glucose isomerase atomic structures for both space groups, we generate nine plausible nanorod models (Extended Data Fig. 7). On the basis of clear discrepancies between the intermolecular distances in these models, and a comparison of the cryoTEM silhouette and the van der Waals model, the only plausible orientation of the nanorod from the $P2_12_12$ space group is in the (001) direction (Fig. 3 and Extended Data Fig. 2).

There are two lattice contact types in the $P2_12_12$ space group—C_1 along the (001) direction, and C_2 along the (110) direction, involving the formation of six and seven hydrogen bonds, respectively (Fig. 4a, b and Extended Data Table 2). The nanorod anisotropy suggests that
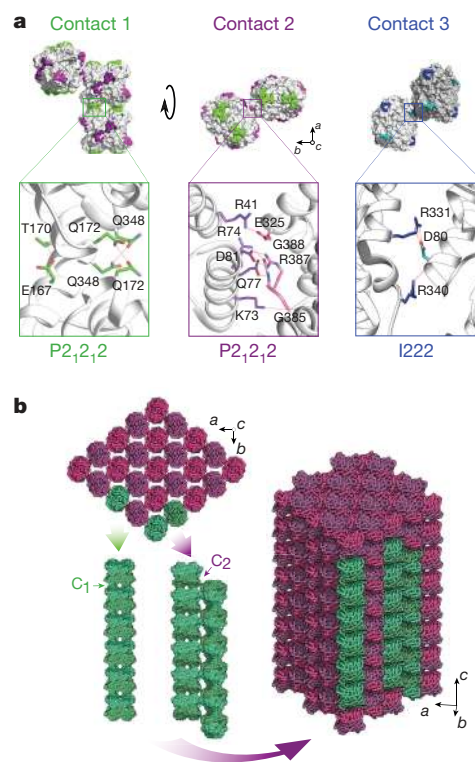
Figure 4 | Lattice contact analysis for both space groups of glucose isomerase. a, Contact types for the $P2_12_12$ and $I222$ space groups, with magnifications of the corresponding surface patches and hydrogen bonds, and showing the amino acids at the contact sites (T, threonine; Q, glutamine; E, glutamate; R, arginine; D, aspartate; K, lysine; G, glycine). Hydrogen bonds are shown with solid lines. **b**, Depiction of the molecular columns that run through a $P2_12_12$ glucose isomerase crystal, with the c-axis corresponding to the long axis of the nanorods observed in cryo-TEM. C_1 and C_2 are contacts 1 and 2.
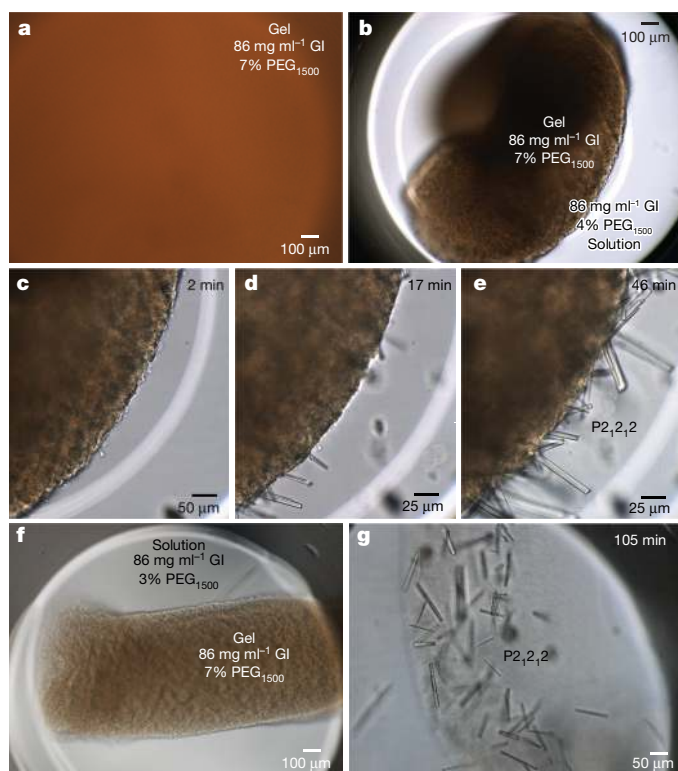
Figure 5 | Gel-mediated nucleation of $P2_12_12$ crystals. a, Glucose isomerase/PEG hydrogel formed with 86 mg ml^{-1} glucose isomerase (GI) and 7% (w/v) PEG$_{1500}$. **b**, An aliquot of the gel shown in panel **a** was transferred to a freshly prepared solution of 86 mg ml^{-1} glucose isomerase and 4% (w/v) PEG$_{1500}$. **c–e**, Consecutive magnified images of the gel–solution interface, at respectively 2 minutes, 17 minutes and 46 minutes after transfer of the gel aliquot. **f**, An aliquot of the gel shown in panel **a** was transferred to a freshly prepared solution of 86 mg ml^{-1} glucose isomerase and 3% (w/v) PEG$_{1500}$. **g**, Formation of $P2_12_12$ crystals is linked to gradual dissolution of the gel phase.

$|C_1|$ is much greater than $|C_2|$. This bond hierarchy can be rationalized by considering the salting-out effects induced by ammonium sulfate; these effects include preferential hydration, salt exclusion in the vicinity of the surface[20], and increased costs of solvent cavity formation[21]. Sulfate ions are excluded with varying degrees from a macromolecular surface[22], with local negative charges contributing strongly to the preferential expulsion[23]. This in turn leads to a net attraction when two macromolecules are close to each other, owing to an imbalance in the osmotic pressure[24]. Thus the strength of the C_1 contact is probably a direct consequence of the 16 negative charges that are buried upon formation, compared with the five such buried negative charges in C_2. Such symmetry breaking will be less pronounced when the precipitant is PEG, which induces a more isotropic attraction, leading to rhombic nuclei, at low concentrations. On the other hand, PEG-induced depletion attraction is not likely to be perfectly uniform for anisotropic particles, as it will favour protein–protein interactions that maximize the overlap volume[25] and, by proxy, the total buried surface area upon complexation. The C_1 contact has the largest difference in accessible surface area (ΔASA), followed by C_2 and C_3 (Extended Data Table 2). We argue that this contributes to the emergence of $P2_12_12$ crystals in intermediate PEG concentrations, where the differences between the contacts become amplified.

An additional level of polymorph control can be gained by means of site-directed mutagenesis: knowing the amino-acid composition of specific lattice contacts allows one to tune their strength. We designed three classes of glucose isomerase mutant that are selectively perturbed in the C_1, C_2 or C_3 modes of interaction. We predict that mutant proteins with impaired C_1 or C_2 contacts will not form $P2_12_12$ crystals; mutants with

altered C_3 contacts should be $I222$ 'knockouts'. We used crystallization screening of said mutants to investigate the strategy of polymorph control by mutagenesis. As predicted, mutants with defective C_1 contacts (in these S171W mutants, the amino acid at position 171, serine, was mutated to tryptophan) or defective C_2 contacts (R387A mutants, with arginine 387 mutated to alanine; and GI_His mutants, in which the protein's carboxy terminus was tagged with a run of histidine residues) no longer produced $P2_12_12$ crystals in the tested conditions, but still nucleated into the $I222$ space group (Table 1 and Extended Data Fig. 8a). The opposite was true of C_3 mutants (R331A plus R340D, where D is aspartate). Seeding experiments using wild-type glucose isomerase microcrystals complement the results of the nucleation trials: wild-type $P2_12_12$ crystals exhibited no growth when transferred to solutions containing C_1 or C_2 mutants; similarly, wild-type $I222$ crystals exhibited no growth in solutions of the C_3 mutant (Table 1). Notably, cryoTEM images of S171W, R387A and GI_His crystals in high concentrations of ammonium sulfate reveal the presence of amorphous aggregates, rather than the nanorod or fibre assemblies seen with the wild-type protein (Extended Data Fig. 8b).

To summarize, when in the presence of a high concentration of ammonium sulfate, the glucose isomerase solution undergoes a rapid decomposition into nanorods that have a quaternary structure similar to the molecular arrangement along the c-axis of the $P2_12_12$ space group. At later time points, fibres are formed (at either high ammonium sulfate or intermediate PEG concentrations) that again have identical intermolecular distances to each other along their long axis. Such high-aspect-ratio assemblies are not observed under conditions that exclusively yield the $I222$ polymorph, nor do they have a structure that is compatible with the crystal lattice of the latter. The fibres are therefore exclusively precursors to the prismatic $P2_12_12$ polymorph. For the ammonium sulfate pathway, our observations suggest that nanorods are the primary building blocks of a next-level self-assembly process that leads to the formation of nanorod oligomers, and subsequently to fibre-like assemblies. Having said that, the data obtained with intermediate PEG concentrations show that fibres can also be formed in the absence of a nanorod phase. Fibres increase in width by lateral attachment, which involves the formation of a large number of interprotein bonds—a complex process that can lead to kinetic traps, as shown by the disorder seen in many fibres (Fig. 1f). Thus, assembly size and crystallinity are order parameters that can evolve independently of each other. We hypothesize that local relaxation from a strained, more disordered state—as seen in many fibrous assemblies—into the crystalline arrangement is associated with an activation barrier that is prohibitively large, yielding disordered fibre assemblies that represent a metastable trap in the protein-assembly pathway. Samples in low PEG concentrations show a total absence of nanorods, higher-order assemblies thereof, or any disordered, liquid-like phases. We detect only faceted crystalline objects, suggesting that $I222$ crystals follow a direct nucleation pathway with monomers as their principal building blocks.

The various pathways seen during the crystallization of glucose isomerase reveal a mechanism of protein polymorph selection that takes place at the earliest measurable stages (20 seconds) of self-assembly (Fig. 6). The primary multimers that are formed have an

Table 1 | Nucleation trials of glucose isomerase mutants and crystal growth tests using wild-type seeds

Mutant	Altered contact	$I222$ nucleation*	$P2_12_12$ nucleation†	$I222$ seed growth*	$P2_12_12$ seed growth†
S171W	C_1	✓	✗	—	✗
R387A	C_2	✓	✗	—	✗
GI_His	C_2	✓	✗	✗	✗
R331A/R340D	C_3	✗	✓	-	-

*50 mM HEPES 7.0 buffer, 100 mM MgCl$_2$, 4% (w/v) PEG$_{1000}$.
†50 mM HEPES 7.0 buffer, 100 mM MgCl$_2$, 1.55 M ammonium sulfate.
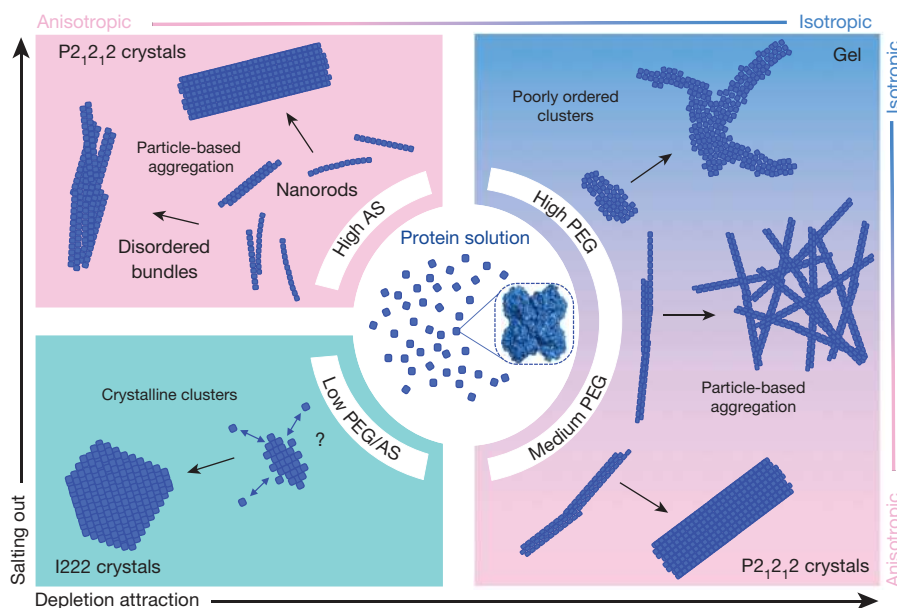
Figure 6 | Proposed model for glucose isomerase crystallization. Top left, high concentrations of ammonium sulfate (AS) induce a strong anisotropic interaction between glucose isomerase molecules, which leads to nanorod formation. These nanorods self-assemble into either disordered fibre bundles or $P2_12_12$ crystals. Bottom left, at low PEG or AS concentrations, glucose isomerase follows a direct nucleation pathway towards $I222$ crystals, owing to a balance between isotropic repulsion and anisotropic attraction. Bottom right, at medium PEG concentrations, fibres are formed that serve as the precursors to either $P2_12_12$ crystals or gel fibre networks. Top right, at high PEG concentrations, a strong isotropic depletion attraction promotes aggregation into ramified aggregates, resulting in dynamic arrest.

architecture that already resembles the crystalline state. This direct nucleation mechanism can be attributed to the mode of interaction between the glucose isomerase molecules, which is a combination of isotropic repulsion and anisotropic attraction[13]. Such interaction potentials affect the emergent nucleation pathway, as they disfavour disordered dense states[26,27]. Self-organization of monomers into (pre) critical clusters with a pronounced symmetry determines their subsequent assembly path at their points of inception[28]. Most unexpectedly, the rod-shaped cluster nucleation pathway for glucose isomerase diverges from the two-step nucleation model for proteins[29] that has gained traction recently, but is perhaps more reminiscent of the cluster–cluster interaction at high supersaturation that is described by classical nucleation theory.

To date, control over emerging polymorphs has been based mostly on detailed knowledge of phase diagrams, and has focused predominantly on solubility differences between polymorphs. By contrast, our insights into the mechanism of polymorphism could inspire selection strategies that are geared towards controlling the modes of interaction, including directionality and kinetics. By (de)stabilizing the modes of interaction that are specific to each polymorph, one can control the throughput of the various nucleation pathways, and ultimately influence the yield of the desired polymorph. Such an approach could aid in the development of hydrogel- and crystal-based biotherapeutic agents that require precise control over the outcome of macromolecular phase transitions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Siezen, R. J., Fisch, M. R., Slingsby, C. & Benedek, G. B. Opacification of gamma-crystallin solutions from calf lens in relation to cold cataract formation. *Proc. Natl Acad. Sci. USA* **82,** 1701–1705 (1985).
2. Patel, A. *et al.* A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162,** 1066–1077 (2015).
3. Eaton, W. A. & Hofrichter, J. in *Advances in Protein Chemistry* (eds Anfinsen, C. B., Edsal, J. T., Richards, F. M. & Eisenberg, D. S.) 63–279 (Academic Press, 1990).
4. Ghiso, J. & Frangione, B. Amyloidosis and Alzheimer's disease. *Adv. Drug Deliv. Rev.* **54,** 1539–1551 (2002).
5. Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **35,** D301–D303 (2007).
6. Basu, S. K., Govardhan, C. P., Jung, C. W. & Margolin, A. L. Protein crystals for the delivery of biopharmaceuticals. *Expert Opin. Biol. Ther.* **4,** 301–317 (2004).
7. ten Wolde, P. R. & Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **277,** 1975–1978 (1997).
8. Pan, W., Galkin, O., Filobelo, L., Nagel, R. L. & Vekilov, P. G. Metastable mesoscopic clusters in solutions of sickle-cell hemoglobin. *Biophys. J.* **92,** 267–277 (2007).
9. Sleutel, M. & Van Driessche, A. E. S. Role of clusters in nonclassical nucleation and growth of protein crystals. *Proc. Natl Acad. Sci. USA* **111,** E546–E553 (2014)
10. Sauter, A. *et al.* Real-time observation of nonclassical protein crystallization kinetics. *J. Am. Chem. Soc.* **137,** 1485–1491 (2015).
11. Sauter, A. *et al.* Nonclassical pathways of protein crystallization in the presence of multivalent metal ions. *Cryst. Growth Des.* **14,** 6357–6366 (2014).
12. Gliko, O. *et al.* Metastable Liquid Clusters in Super- and Undersaturated Protein Solutions. *J. Phys. Chem. B* **111,** 3106–3114 (2007).
13. Sleutel, M., Lutsko, J., Van Driessche, A. E. S., Durán-Olivencia, M. A. & Maes, D. Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nat. Commun.* **5,** 5598 (2014).
14. Chung, S., Shin, S.-H., Bertozzi, C. R. & Yoreo, J. J. D. Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proc. Natl Acad. Sci. USA* **107,** 16536–16541 (2010).
15. Yau, S.-T. & Vekilov, P. G. Quasi-planar nucleus structure in apoferritin crystallization. *Nature* **406,** 494–497 (2000).
16. Nielsen, M. H., Aloni, S. & Yoreo, J. J. D. In situ TEM imaging of $CaCO_3$ nucleation reveals coexistence of direct and indirect pathways. *Science* **345,** 1158–1162 (2014).
17. Yamazaki, T. *et al.* Two types of amorphous protein particles facilitate crystal nucleation. *Proc. Natl Acad. Sci. USA* **114,** 2154–2159 (2017).
18. Bhosale, S. H., Rao, M. B. & Deshpande, V. V. Molecular and industrial aspects of glucose isomerase. *Microbiol. Rev.* **60,** 280–300 (1996).
19. Gillespie, C. M., Asthagiri, D. & Lenhoff, A. M. Polymorphic protein crystal growth: influence of hydration and ions in glucose isomerase. *Cryst. Growth Des.* **14,** 46–57 (2014).
20. Arakawa, T. & Timasheff, S. N. Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry* **21,** 6545–6552 (1982).
21. Melander, W. & Horváth, C. Salt effect on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. *Arch. Biochem. Biophys.* **183,** 200–215 (1977).
22. Fudo, S., Qi, F., Nukaga, M. & Hoshino, T. Influence of precipitants on molecular arrangement and space group of protein crystals. *Cryst. Growth Des.* **17,** 534–542 (2017).
23. Paterová, J. *et al.* Reversal of the Hofmeister series: specific ion effects on peptides. *J. Phys. Chem. B* **117,** 8150–8158 (2013).
24. Asakura, S. & Oosawa, F. On interaction between two bodies immersed in a solution of macromolecules. *J. Chem. Phys.* **22,** 1255–1256 (1954).

25. Kraft, D. J. *et al.* Surface roughness directed self-assembly of patchy particles into colloidal micelles. *Proc. Natl Acad. Sci. USA* **109,** 10787–10792 (2012).
26. Whitelam, S. Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions. *Phys. Rev. Lett.* **105,** 088102 (2010).
27. Hedges, L. O. & Whitelam, S. Limit of validity of Ostwald's rule of stages in a statistical mechanical model of crystallization. *J. Chem. Phys.* **135,** 164902 (2011).
28. Russo, J. & Tanaka, H. Crystal nucleation as the ordering of multiple order parameters. *J. Chem. Phys.* **145,** 211801 (2016).
29. Vekilov, P. G. Dense liquid precursor for the nucleation of ordered solid phases from solution. *Cryst. Growth Des.* **4,** 671–685 (2004).

METHODS

Protein production and purification. Glucose isomerase was obtained from Hampton Research (from wild-type *Streptomyces rubiginosus*) and received as a crystalline slurry. Small aliquots were dialysed overnight (using Spectra/Por standard regenerated cellulose (RC) tubing, molecular weight cut-off (MWCO) 12–14 kDa; SpectrumLabs) against 10 mM HEPES pH 7.0 buffer plus 1 mM MgCl$_2$ at 4 °C. The protein solution was concentrated using a centrifugal filter with a MWCO of 100 kDa (Amicon Ultra-15 Cellulose, Milipore) to a typical concentration of 200–250 mg ml^{-1} and stored at 4 °C. Concentrations were determined by measuring the absorbance at 280 nm using an extinction coefficient ε_{280} of 1.074 mg^{-1} ml cm^{-1}.

Synthetic DNA of full-length wild-type glucose isomerase (UniProt, *P24300*) with a carboxy-terminal His$_6$ tag (GI_His), and mutants (S171W, R387A, R331A/R340D) with no carboxy-terminal His$_6$ tag cloned into plasmid pET22b via *Nde*I and *Nco*I restriction sites, was ordered at GenScript. Recombinant proteins were expressed in *Escherichia coli* strain BL21(DE3) after induction at an optical density (OD)$_{600nm}$ of 0.7 with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h at 37 °C. Cells were harvested by centrifugation at 6,238g for 15 min and resuspended in 100 mM Tris-HCl pH 7.3, 1 mM ethylenediaminetetra acetic acid (EDTA, 4 ml per gram of wet cells) supplemented with 5 μM leupeptin, 1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), 100 μg ml^{-1} lysozyme and 20 μg ml^{-1} DNase I, and incubated for 30 min at 4 °C. Subsequently, MgCl$_2$ was added to a final concentration of 10 mM and cells were lysed by two passages in a Constant System Cell Cracker at 20 kilopounds per square inch (kpsi) at 4 °C; cell debris was removed by centrifugation at 48,400g for 45 min at 4 °C. The cytoplasmic extract was incubated for 10 min at 65 °C and the insoluble fraction was removed by centrifugation at 48,400g for 45 min at 4 °C.

For the non-His-tagged constructs, the supernatant was filtrated through a 0.22 μm pore filter and loaded on a 5 ml pre-packed Hitrap Q FF column (GE Healthcare) equilibrated with buffer A (50 mM bis-tris-HCl pH 6.0, 10 mM NaCl). The column was then washed with 40 bed volumes of 20% buffer B (50 mM bis-tris-HCl pH 6.0, 1 M NaCl) and bound proteins were eluted with a linear gradient of 20–50% buffer B over 10 bed volumes. Fractions containing wild-type or mutant glucose isomerase—as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE)—were pooled and supplemented with ammonium sulfate to a final concentration of 1.5 M and loaded on a 5 ml pre-packed HiTrap Phenyl HP column (GE Healthcare) equilibrated with buffer A (100 mM Tris pH 7.3, 1.5 M ammonium sulfate). The column was then washed with 40 bed volumes of 25% buffer B (100 mM Tris pH 7.3) and bound proteins were eluted with a linear gradient of 25–85% buffer B over 15 bed volumes. Fractions containing wild-type or mutant glucose isomerase, as determined by SDS–PAGE, were pooled and dialysed (Spectra/Por standard RC tubing, MWCO 12–14 kDa; SpectrumLabs) against 10 mM HEPES pH 7.0 plus 1 mM MgCl$_2$ overnight at 4 °C (the buffer was replaced twice), and concentrated in a MWCO 100 kDa spin concentrator (Amicon Ultra-15 Cellulose; Milipore) to a typical final concentration of 30 mg ml^{-1}. (For the S171W mutant, $\varepsilon_{280} = 1.198$ mg^{-1} ml cm^{-1}). Cleared cytoplasmic extracts of GI_His were loaded on a 5 ml pre-packed Histrap Ni-NTA column (GE Healthcare) equilibrated with buffer A (50 mM Tris-HCl pH 7.3, 500 mM NaCl and 20 mM imidazole). The column was then washed with 40 bed volumes of buffer A, and bound proteins were eluted with a linear gradient of 0–75% buffer B (50 mM Tris-HCl pH 7.3, 500 mM NaCl and 500 mM imidazole) over 15 bed volumes. Fractions containing GI_His were pooled, dialysed and concentrated as indicated above.

Glucose isomerase crystallization. For a typical crystallization experiment with wild-type glucose isomerase, the protein stock solution was first diluted to a concentration of 75 mg ml^{-1} in 50 mM HEPES pH 7.0 and 100 mM MgCl$_2$, and then mixed at 22 °C with an equal volume of a buffered ammonium sulfate, PEG$_{1000}$ or PEG$_{1500}$ solution that was at a concentration twice that desired after mixing. Final concentrations ranged from 0.5 M to 1.75 M ammonium sulfate, and from 3% to 7% (w/v) PEG$_{1000}$ or PEG$_{1500}$. Space-group determination was based on the distinct crystal morphologies of both space groups (Extended Data Fig. 1a), using a wide-field optical microscope. The phase diagrams in Extended Data Fig. 1b, c were determined by setting up triplicate crystallization tests using the micro-batch-under-oil method (with Nunc 72 microwell minitrays; Sigma Aldrich) at 22 °C, with 10 μl drops of mother liquor.

Precipitant dependence of glucose isomerase crystallization. We began by mapping out the concentration dependence of glucose isomerase polymorphism for the precipitants used here. Our starting point was 50 mM HEPES pH 7.0 plus 100 mM MgCl$_2$, which yields exclusively rhombic (I222) crystals for a broad range of protein concentrations (20–75 mg ml^{-1}). By supplementing that condition with ammonium sulfate in 100–200 mM increments, from 0.5 M to an upper limit of 1.75 M final concentration, we saw a gradual shift from rhombic to prismatic (P2$_1$2$_1$2)

crystals (Extended Data Fig. 1b). Similarly, by supplementing the base condition with either PEG$_{1000}$ or PEG$_{1500}$ to a final concentration of 3% to 8.5% (w/v), we recorded a gradual shift from I222 to P2$_1$2$_1$2 crystals; at the higher PEG concentrations, a dense gel phase was formed. We note that there was a narrow PEG concentration range (for PEG$_{1500}$, from 4.5% to 5.5%) where I222 crystals, P2$_1$2$_1$2 crystals and gels were observed simultaneously (see Extended Data Fig. 5 for a detailed microscopic record). Gelation depended only weakly on glucose isomerase concentration: the gelation line occurs as a vertical in Extended Data Fig. 1.

Induction-time measurements. We determined induction times (t_{ind}) for glucose isomerase crystallization as a function of ammonium sulfate concentration (1.2–1.65 M) by following the change in absorbance of freshly prepared supersaturated solutions. We monitored the increase of absorbance in the reacting solutions by *in situ* time-resolved ultraviolet-visible spectroscopy (Agilent Cary 300E spectrophotometer). Measurements were carried out at a wavelength of 500 nm (absorbance of individual glucose isomerase molecules is minimal at this wavelength) and performed in poly (methyl methacrylate) (PMMA) cuvettes located inside a peltier thermostated cell module at 20 °C. The time elapsed between the mixing of protein and salt solutions and the first observed change in turbidity was taken as the induction time. This point was determined through linear fitting of the sigmoidal absorbance curve near the inflection point and determining the intersection with the *x*-axis.

To obtain a general idea about the nucleation kinetics and to estimate the nucleation induction time, we monitored the turbidity of the crystallizing solution between 1.2 M and 1.65 M ammonium sulfate, and obtained an exponential dependence of t_{ind} on ammonium sulfate concentration (Extended Data Fig. 2). We later used these simple estimates of t_{ind} as a guide for the preparation of cryo-TEM samples, to determine the desired number of samples and time intervals for each condition. We also note that, under high ammonium sulfate concentrations, the system undergoes near-spinodal decomposition with respect to the crystalline phase. Wide-field light microscopy imaging confirms that, even under these conditions, glucose isomerase solidifies exclusively into the (P2$_1$2$_1$2) crystalline state. We find no evidence that any amorphous solid states are formed; nor is there any indication that a liquid–liquid phase boundary is crossed.

Time-resolved dynamic light scattering. We collected intensity correlation functions of mixed protein–precipitant solutions at 20 °C, using 10 mm cylindrical cuvettes at an angle of 90°, and employing an ALV-CGS-3 static and dynamic light scattering (DLS) device with a 22 mW helium–neon laser (wavelength 632.8 nm). We collected data in a pseudo cross-correlation set-up in order to minimize the contribution of dead time effects and photomultiplier-tube after-pulsing to the recorded signal. The intensity autocorrelation function $g_2(\tau) - 1$, with τ being the delay time, is connected to the electric-field correlation function $g_1(\tau) - 1$ through the Siegert relation[30]:

$$g_2(\tau) = B(1 + \beta \left| g_1(\tau) \right|^2)$$

where B is the baseline of the correlation function at infinite delay, and β is the function value at zero delay. For samples containing PEG$_{1000}$, we used the following double-exponential function (equation (1) is used to fit $g_1(\tau)$ at time points before kinetic arrest, and equation (2) is a stretched exponential used after gelation has occurred):

$$g_1(\tau) = A_1 e^{-\Gamma_1 t} + A_2 e^{-\Gamma_2 t} \tag{1}$$

$$g_1(\tau) = e^{-\Gamma t^p} \tag{2}$$

where p is a fitting parameter, and $\Gamma = Dq^2$ is the decay rate defined by the diffusion coefficient D of the particles and the magnitude of the scattering vector $q = \frac{4\pi n}{\lambda} \sin(\theta/2)$ at the scattering angle θ.

We collected time-lapse DLS acquisitions to follow, in real-time, the crystallization of filtered glucose isomerase solutions in 50 mM HEPES pH 7.0 plus 100 mM MgCl$_2$, using a 1.5 M ammonium sulfate solution and 48 mg ml^{-1} glucose isomerase. The time evolution of the intensity correlation function is shown in Extended Data Fig. 6a. Processing of the raw curves with the ALV-correlator software (ALV-7004 v3.0.5.1), using the regularization method, yields hydrodynamic radii of the various glucose isomerase populations that form in solution. Thirty seconds after glucose isomerase/ammonium sulfate mixing, light scattered by the glucose isomerase monomers only was collected (measured hydrodynamic radius $R_h = 8$ nm). Over the course of minutes, a second shoulder started to form in the correlogram, with the earliest measurable species (at 4 min) corresponding to micrometre-sized particles (denoted as 'clusters'). These species rapidly grew until they completely dominated the recorded signal (by 14 min). Visual inspection at this point showed that the sample had become opaque. *Ex situ* wide-field microscopy analysis after 30 min confirmed the presence of P2$_1$2$_1$2 nanocrystals with a

small minority of $I222$ crystals. On the basis of the typical nanorod dimensions determined by cryoTEM (length 100 nm; aspect ratio 6), we predict—following the corrections described in ref. 31—that they would have an apparent hydrodynamic radius of ± 45 nm. Given the results discussed above, we conclude that we could not detect any light scattered by particles in this size range.

We also tested conditions that do not yield an ordered solid, but instead lead to a kinetically trapped gel state. Time-resolved DLS of a solution of 50 mM HEPES pH 7.0, 100 mM $MgCl_2$, 7% (w/v) PEG_{1000} and 25 mg ml^{-1} glucose isomerase showed that the intensity auto-correlation function (ACF) could be fit at early time points with a double-exponential decay (with a fast-diffusing population corresponding to monomers, and a slowly diffusing population corresponding to clusters that grow as a function of time). At later stages a stretched exponential was required to reproduce the ACF (Extended Data Fig. 6b, c). Stretched exponentials indicate a hierarchy of fluctuations on all length scales and are a well known characteristic of gels[32]. Using optical microscopy, we obtained a visual confirmation of the gelled state. The inset of Extended Data Fig. 6c clearly resolves the pores that are present in the mesh of fibres in the kinetically arrested state.

Seeding experiments using glucose isomerase hydrogels. Crystallization trials using PEG as a precipitant showed that glucose isomerase exhibits $I222/P2_12_12$ polymorphism over a narrow concentration range (Extended Data Figs 1, 5). For concentrations lower than or equal to 4% (w/v) PEG_{1500}, we observed only $I222$ crystals. Conversely, for concentrations higher than or equal to 6% (w/v) PEG_{1500}, we obtained opaque glucose isomerase hydrogels. At 4.5% (w/v) PEG_{1500}, $I222/P2_12_12$ polymorphism occurred, with strongly varying nucleation densities for the $P2_12_12$ space group; in some cases a gelatinous phase also formed that seemed to enter into competition with the crystalline phases. We observed a similar transition regime for PEG_{1000}, but shifted towards higher PEG concentrations (Extended Data Fig. 1). We transferred a small gel fragment grown at 7% (w/v) PEG_{1500} (Fig. 5a) to a freshly prepared solution that was identical in composition but of a lower PEG_{1500} concentration ((w/v) 4%; Fig. 5b). Time-lapse imaging of the gel–solution interface revealed the rapid and exclusive formation of $P2_12_12$ crystals on, or protruding from, the gel phase (Fig. 5c–e). Transferring a gel fragment to 3% (w/v) PEG_{1500}, however, led to the gradual dissolution of the gel phase as $P2_12_12$ crystals emerged over time (Fig. 5f, g).

Crystallization of GI_His and mutant proteins. To gain more control over the polymorph selection process, we designed and produced glucose isomerase mutants that we predicted to affect space-group-specific intermolecular contacts while leaving all other contacts unchanged. We had three different types of mutant, impairing C_1 contacts (the S171W mutant, with steric inhibition), C_2 contacts (the R387A mutant, with a salt bridge removed, and GI_His, with steric inhibition) or C_3 contacts (the R331A/R340D mutant, with charge inversion). We predicted that mutants with defective C_1 and/or C_2 interactions would form exclusively $I222$ crystals, whereas impaired C_3 constructs would form just $P2_12_12$ crystals.

We gauged our ability to control polymorph selection through site-directed mutagenesis by setting up crystallization trials for the new constructs, using conditions that lead (almost) exclusively to either $I222$ or $P2_12_12$ crystals with wild-type glucose isomerase. Thus, 50 mM HEPES pH 7.0, 100 mM $MgCl_2$ and 1.55 M ammonium sulfate leads predominantly to $P2_12_12$ crystals, whereas 50 mM HEPES pH 7.0, 100 mM $MgCl_2$ and 4% (w/v) PEG_{1000} favours the nucleation of the $I222$ space group. If no crystallization (of either space group) could be induced with the selected mutant under these conditions, we set up grid screens by varying the precipitant concentration. If only one space group could be obtained after such a screening, we classified the tested mutant as $I222$-negative or $P2_12_12$-negative (Table 1 and Extended Data Fig. 8a). We note that any crystallization screening is inherently finite, and therefore cannot be used to conclusively rule out the absence of a particular polymorph throughout all of chemical space. Hence, as an auxiliary method, we set up seeded crystallization tests using pre-grown wild-type $I222$ or $P2_12_12$ glucose isomerase crystals, which we then washed in their corresponding mother liquors to remove any soluble glucose isomerase species, and transferred to an identical mother liquor solution supplemented with 10 mg ml^{-1} of the respective mutant. We monitored the growth of these seed crystals over time using wide-field microscopy (Table 1).

Cryo-TEM. For cryo-TEM, we used 200-mesh copper grids with Quantifoil R 2/2 holey carbon films (Quantifoil Micro Tools GmbH). We prepared samples using an automated vitrification robot (FEI Vitrobot Mark III) for plunging in liquid ethane[33]. Before use, all TEM grids were surface plasma treated for 40 seconds using a Cressington 208 carbon coater. We studied the samples with the Technische Universiteit Eindhoven/FEI cryoTITAN (www.cryotem.nl) operated at 300 kV, equipped with a field emission gun (FEG), a post-column Gatan energy filter (GIF) and a post-GIF 2k × 2k Gatan charge-coupled-device camera. We choose t_0 as the moment at which we induced supersaturation with respect to the crystalline phase (that is, when we mixed the protein with the precipitant solution)

and t_{end} as the time at which crystals became detectable using light microscopy. The exact time point of the samples as indicated in the main text was defined as the moment (after blotting excess liquid) when the electron-microscopy grid was plunged into the liquid ethane. The selected solution conditions represent a compromise between the nucleation density and the overall rate of transformation—that is, for TEM one needs on the one hand a high enough particle density, and on the other slow enough kinetics to manage the cryogenic-quenching at constant time intervals (roughly 2 minutes). We acquired images in low-dose mode at a magnification of either ×24,000 with a nominal defocus of $-5\,\mu m$, or ×11,500 and $-10\,\mu m$ defocus.

Single-particle data processing and projection approximation. We determined the defocus of the micrographs by using a script developed in-house (written by R. Efremov). We manually picked and stacked 1,240 particles in E2BOXER. A low-pass Gaussian filter was applied to remove excessive high-frequency noise, and the contrast was inverted before classification. We carried out two-dimensional class averaging by K-means classification using a soft circular mask, and then performed a multi-reference alignment using SPARX[34].

To approximate a low-dose cryoTEM projection of the rod assemblies (Extended Data Fig. 8), we used a Protein DataBank (PDB) model. From the PDB model (containing all atom coordinates), we created a three-dimensional density map of the rod via Chimera 1.12 (using the molmap command). Each atom is described as a three-dimensional Gaussian distribution of width proportional to the resolution (3 nm at $-5\,\mu m$ defocus) and amplitude proportional to the atomic number. The pixel size was set to 0.4 nm, which is close to the pixel size of acquisition (0.38 nm). We summed this three-dimensional intensity map in Matlab along the y-direction perpendicular to the rod length, creating a density projection of the structure. The TEM image was approximated by subtracting the density projection from a flat background image containing Poisson noise (mean intensity = 100 electrons per pixel, as the beam intensity during cryoTEM imaging). Fresnel fringes (white lines surrounding glucose isomerase) arising from the applied underfocus during imaging were not included in the simulation.

Distances along nanorods, fibres, crystals and gel fibres. To obtain estimates of the intermolecular distances within glucose isomerase's nanorod structures, we plotted the power spectrum of the greyscale values along the long axis of a single nanorod, and identified two dominant frequencies that correspond to the characteristic intermolecule and intramolecule distances (Extended Data Fig. 3a–d). In a second approach, we calculated 2D-FFTs using ImageJ 1.50i (ref. 35) from an entire TEM image containing numerous nanorods lying in random orientations (Extended Data Fig. 3a). This orientational averaging yielded an FFT image containing two concentric circles, whose radii again corresponded to the intermolecule and intramolecule distances (Extended Data Fig. 3e). We applied orientational averaging to the 2D-FFT (Extended Data Fig. 3f) and took the inverse frequencies of the two maxima. Applying this second approach to 51 images, we obtained a value of 8.2 ± 0.1 nm for the intermolecular distance along the long nanorod axis (Extended Data Fig. 3g). We used a similar method to determine the characteristic distance within the fibre structures and for the nanocrystals, but used selections corresponding to specific regions of interest (fibre or crystal outline). Also, in the orientational averaging step of the 2D-FFT, we calculated the radial profile over a range of 20° and 5° for fibres and crystals, respectively, instead of 180° for the nanorods. Using 27 fibres, we obtained a characteristic distance of 8.0 ± 0.1 nm along the long axis, and using 29 crystals, we measured 8.1 ± 0.1 nm in the c-direction. For the gel fibres, we integrated over 20° and obtained a spacing of 8.0 ± 0.2 nm (4% (w/v) PEG_{1500}) and 7.0 ± 0.2 nm (7% (w/v) PEG_{1500}), on the basis of 24 and 16 measurements, respectively.

Crystallographic analysis. As starting models for the atomic structures of glucose isomerase in the $P2_12_12$ and $I222$ space groups, we used the biological assembly models of entries 1OAD and 9XIA in the RCSB PDB (https://www.rcsb.org). We generated nearest crystallographic neighbours of the glucose isomerase molecule for both space groups using Chimera 1.11.2rc (Fig. 4). We identified residues that partake in lattice contacts by calculating the accessible surface area (ASA) on a per-residue level, using AREAIMOL of the CCP4 software suite[36]. We determined ASAs for both of the starting models, and for the models consisting of glucose isomerase and its nearest neighbour, by using a probe radius of 1.4 Å. Residues with a non-zero ΔASA are (partially) buried in the bound complex and therefore considered to be part of the lattice-contact patch. We identified hydrogen-bond pairs with the FindHBond tool in Chimera 1.11.2rc using default settings, and salt bridges using the PDBePISA (http://www.ebi.ac.uk/pdbe/pisa/) and the 2P2I (http://2p2idb.cnrs-mrs.fr/2p2i_inspector.html) protein-interaction webservers.

Given these two models (1OAD and 9XIA) of glucose isomerase for both space groups, we used crystallographic symmetry operations to generate a number of plausible candidates for the experimentally observed nanorods. For this, we

identified the various classes of lattice contact (see below) that exist in both space groups, and applied translational and rotational symmetry operations using the Supercell plugin (https://pymolwiki.org/index.php/Supercell) to Pymol to construct linear glucose isomerase sequences in space. A basic requirement that we set is that adjoining glucose isomerase molecules must be in direct contact with each other, be it through van der Waals, hydrogen-bond or electrostatic interactions. We discarded loose packing structures with water-mediated contacts only (Extended Data Fig. 8). Next, we compared the intermolecule distances and particle silhouettes to identify a potential match with the nanorods that were imaged using cryoTEM. Nanorods constructed along the (100), (010), (001) and (011) directions of the $I222$ spacegroup, and along the (100), (010) and (110) directions for the $P2_12_12$ spacegroup, all have a helical ultrastructure that has a pitch defined by the respective unit-cell dimensions. The only linear array of glucose isomerase molecules that could be generated is that in the (001) direction for the $P2_12_12$ space group, and the (111) direction for $I222$. Careful comparison of the orientations of the molecules with respect to the nanorod axis led us to conclude that the $P2_12_12$ (001) model is the most plausible. Indeed, juxtaposing the cryoTEM image of a single nanorod with the van der Waals surface representation of our $P2_12_12$ (001) nanorod model showed good agreement between the two. Also, the $P2_12_12$ (001) nanorod had an intermolecular distance of 7.83 nm when using PDB 1OAD as a reference structure. The nanocrystals that we grew, however, were less compact. The intermolecular distance (along the c-axis; based on the fringe pattern in the cryoTEM images) is 8.1 ± 0.1 nm—a good match to the experimental intermolecular distance within the nanorods (8.1 ± 0.2 nm). We therefore conclude that the nanorods that are formed in high ammonium sulfate concentrations are linear molecular arrays that can also be found along the c-axis of a mature $P2_12_12$ glucose isomerase crystal.

Lattice contacts. For the $P2_12_12$ space group, we identified two types of lattice contact that involve three different surface patches, designated P1, P2a and P2b (Extended Data Table 2). Contact 1 (C_1) is made in the (001) direction by the self-recognition of patch P1, and is duplicated owing to the non-crystallographic twofold symmetry of the glucose isomerase tetramer. The total contact therefore includes the formation of 2×3 hydrogen bonds and has a ΔASA of 844 Å^2. Contact 2 (C_2) along the (110) direction is formed by the binding of P2a with P2b, involving the formation of seven hydrogen bonds and two salt bridges, and encompassing a total ΔASA of 622 Å^2. For the $I222$ space group, there is just one lattice-contact type (contact 3, C_3), which involves two surface patches, Ia and Ib, making three hydrogen bonds and resulting in a net ΔASA of 372 Å^2. Note that these patches are unique to their respective space groups, although P2a and Ia share one amino acid (D81).

Data availability. We declare that the data supporting the findings of this study are available within the paper and the Extended Data figures and tables. Further data are available from the corresponding authors upon request.

30. Berne, B. J. & Pecora, R. *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics* (Courier Dover Publications, 1976).
31. Ortega, A. & García de la Torre, J. Hydrodynamic properties of rodlike and disklike particles in dilute solution. *J. Chem. Phys.* **119,** 9914–9919 (2003).
32. Krall, A. H. & Weitz, D. A. Internal dynamics and elasticity of fractal colloidal gels. *Phys. Rev. Lett.* **80,** 778–781 (1998).
33. Friedrich, H., Frederik, P. M., de With, G. & Sommerdijk, N. A. J. M. Imaging of self-assembled structures: interpretation of TEM and cryo-TEM images. *Angew. Chem. Int. Edn* **49,** 7850–7858 (2010).
34. Hohn, M. *et al.* SPARX, a new environment for cryo-EM image processing. *J. Struct. Biol.* **157,** 47–55 (2007).
35. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
36. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59,** 1131–1137 (2003).
37. Vuolanto, A., Uotila, S., Leisola, M. & Visuri, K. Solubility and crystallization of xylose isomerase from *Streptomyces rubiginosus*. *J. Cryst. Growth* **257,** 403–411 (2003).
38. Sleutel, M., Willaert, R., Wyns, L. & Maes, D. Kinetics and thermodynamics of glucose isomerase crystallization. *Cryst. Growth Des.* **9,** 497–504 (2009).
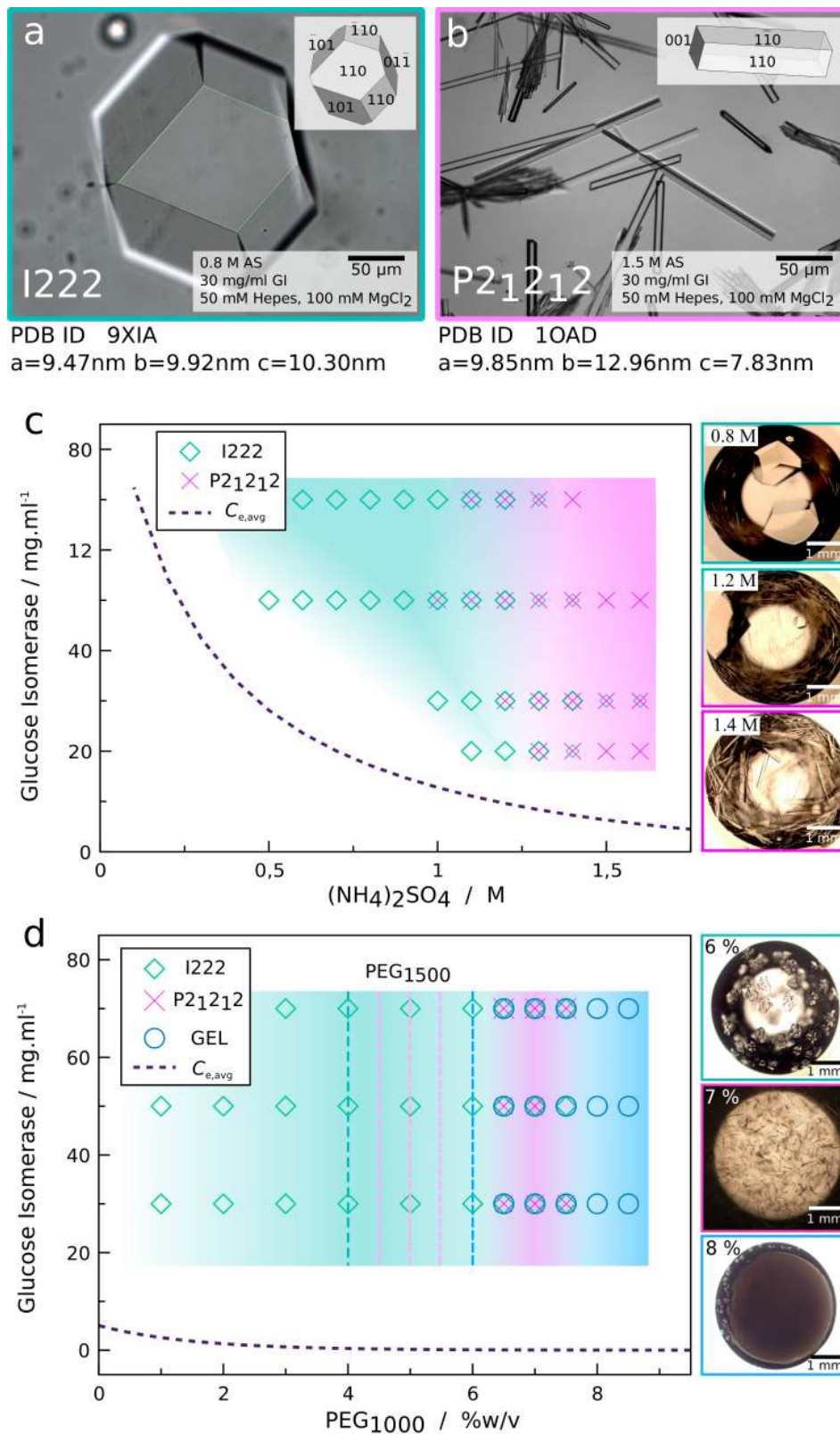
Extended Data Figure 1 | Phase diagrams for glucose isomerase.
a, **b**, Wide-field microscopic images of the *I222* (**a**) and *P2₁2₁2* (**b**) glucose isomerase (GI) polymorphs obtained with 0.8 M and 1.5 M ammonium sulfate (AS). **c**, Glucose isomerase phase diagram in ammonium sulfate ($(NH_4)_2SO_4$) at 22 °C (single points represent triplicate measurements), showing the solubility line $C_{e,avg}$ (dashed line). Smaller diamonds and crosses denote smaller numbers of crystals than larger symbols. $C_{e,avg}$ is a mathematical average that we calculated by using the solubilities at 19 °C and 25 °C from ref. 9. **d**, Glucose isomerase phase diagram in PEG_{1000} at 22 °C, with the $C_{e,avg}$ solubility line taken from ref. 38. The dotted lines, following the same colour code as the single points, indicate the phase boundaries in PEG_{1500}. The photographs to the right are representative microscopy images at the indicated precipitant concentrations.
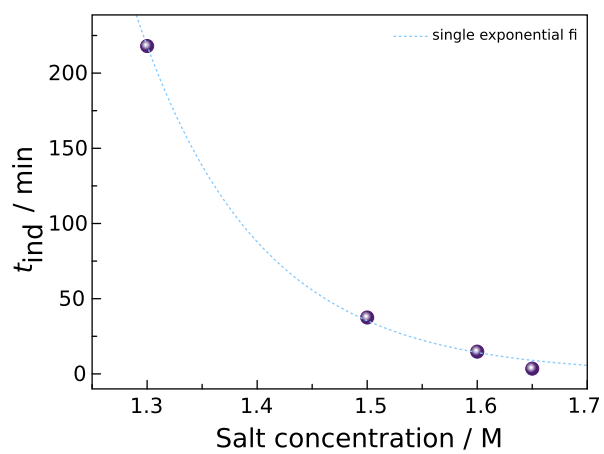
Extended Data Figure 2 | Induction time measurements. Induction time, t_{ind}, as a function of ammonium sulfate concentration. Values next to data points correspond to calculated supersaturation ($\ln C/C_e$) values, according to ref. 37.
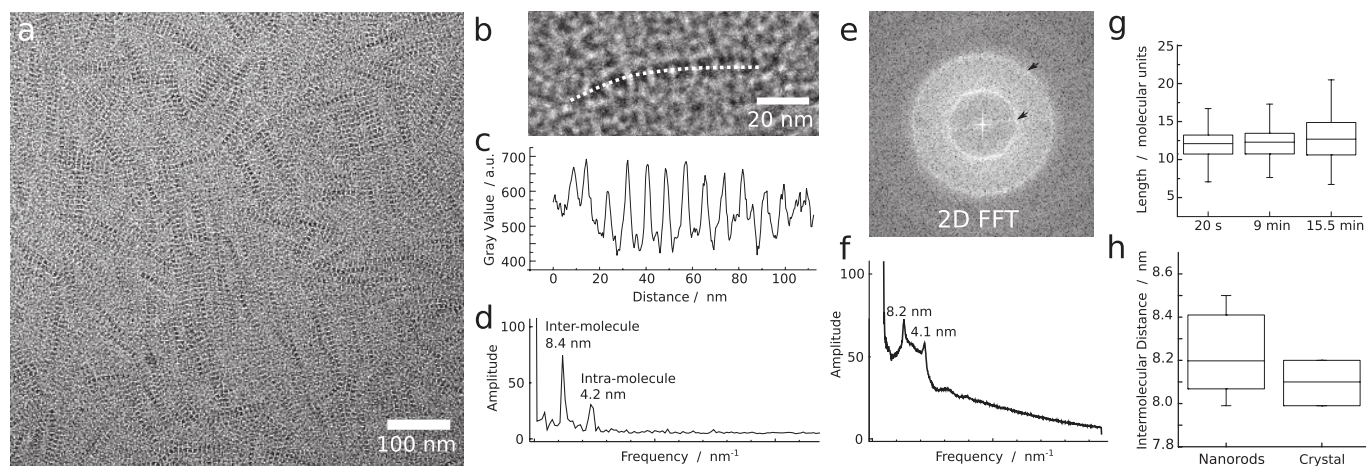
Extended Data Figure 3 | Determination of the intermolecular distance along the nanorod axis. a, Complete image acquired at ×24,000 magnification. **b**, CryoTEM image of single nanorods. **c**, Greyscale values along the length of the dotted line in panel **a**. **d**, 1D-FFT of panel **c**, calculated using OriginPro 8.6.0. **e**, 2D-FFT image calculated using ImageJ 1.50i. **f**, Radial average of panel **e**. **g**, Nanorod length expressed in molecular units as a function of time in 1.5 M ammonium sulfate. **h**, Intermolecular distance along the nanorod axis compared with the crystallographic distance along the *c*-axis of the prismatic crystals. The box range shows the 25th to 75th percentiles; the horizontal line is the median; error bars highlight the 10th and 90th percentiles.
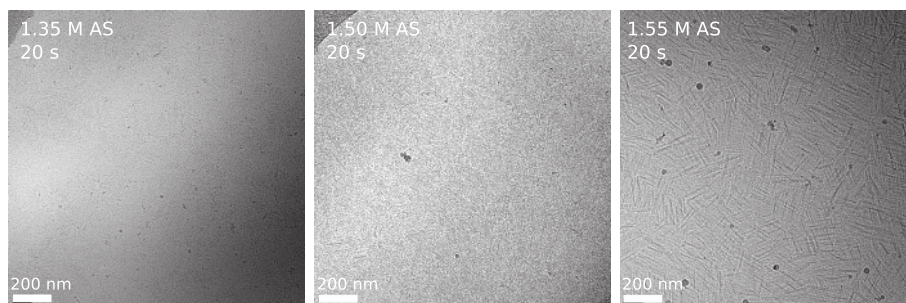
Extended Data Figure 4 | Nanorod formation at early time points. CryoTEM images of crystallizing glucose isomerase solutions 20 seconds after protein–precipitant mixing with 1.35 M, 1.50 M or 1.55 M ammonium sulfate.

4% PEG1500   4.5% PEG1500   5% PEG1500
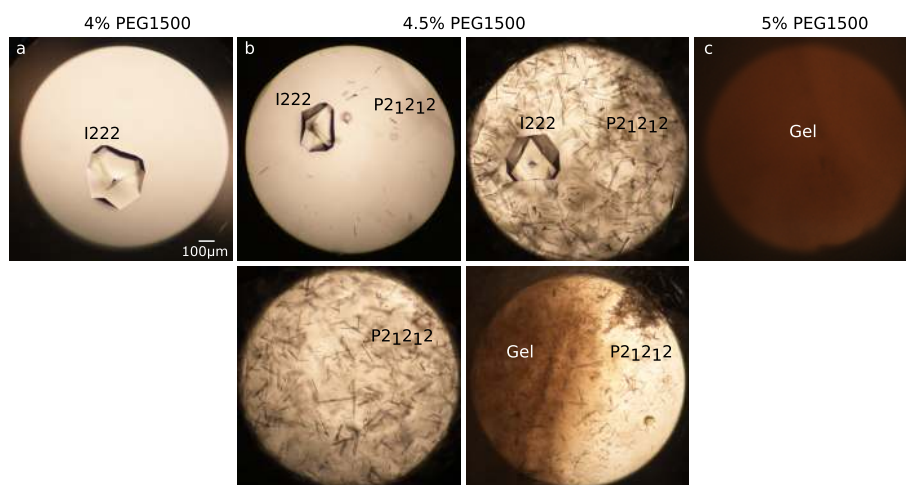
Extended Data Figure 5 | $I222/P2_12_12$/gel coexistence point. a–c, Crystallization experiments using the microbatch-under-oil set-up, with $86\,mg\,ml^{-1}$ glucose isomerase, $50\,mM$ HEPES pH 7.0 and $100\,mM$ MgCl$_2$, and 4% (**a**), 4.5% (**b**) or 5% (**c**) (w/v) PEG$_{1500}$. **a**, $I222$ crystals; **b**, $I222 + P2_12_12$, $P2_12_12$, $P2_12_12$ and gel; **c**, gel.
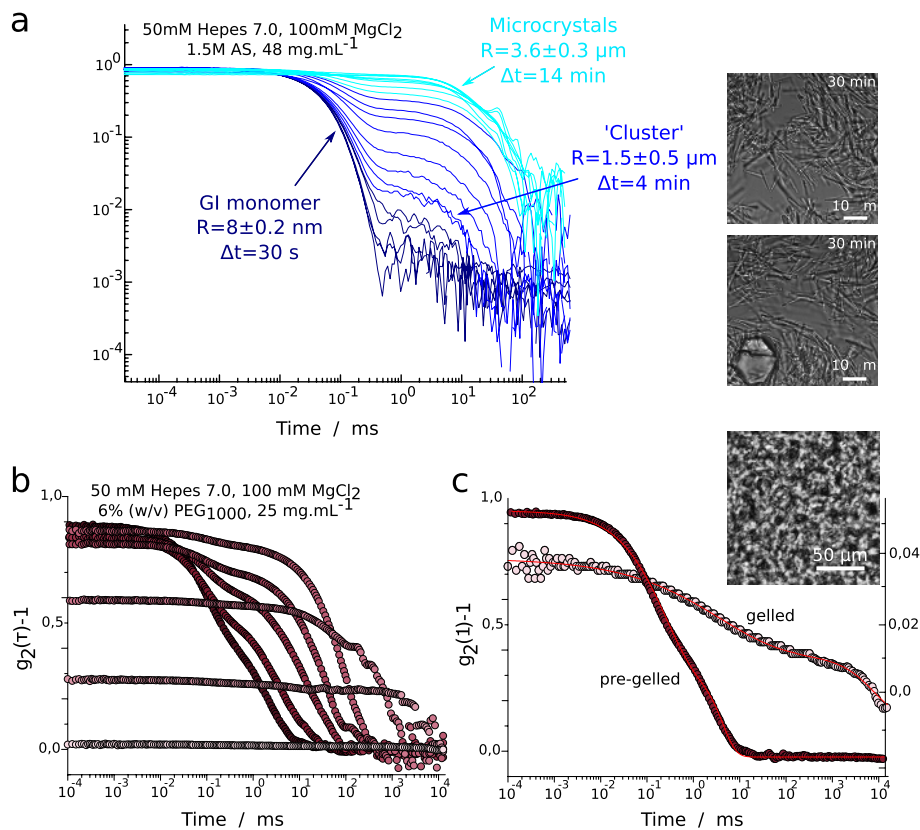
Extended Data Figure 6 | Time-resolved DLS of crystallizing glucose isomerase solutions. a, DLS time series of a crystallizing 48 mg ml^{-1} glucose isomerase solution with 50 mM HEPES pH 7.0, 100 mM MgCl$_2$, 1.5 M ammonium sulfate, collected at an angle of 90°, ranging from 30 seconds to 14 minutes after protein/precipitant mixing. R, particle radius. Microscopy snapshots at the right were taken *ex situ* after 30 minutes. **b**, Time evolution (from dark to light) of the intensity correlation function of a 50 mM HEPES pH 7.0, 100 mM MgCl$_2$, 6% PEG$_{1000}$ (w/v) solution collected at an angle of 90°. **c**, Fitting of a pre-gelled (20 seconds; left-hand y-axis) and gelled (30 minutes; right-hand y-axis) sample using equations (1) and (2) respectively. Inset, wide-field microscopy image of the gelled state.
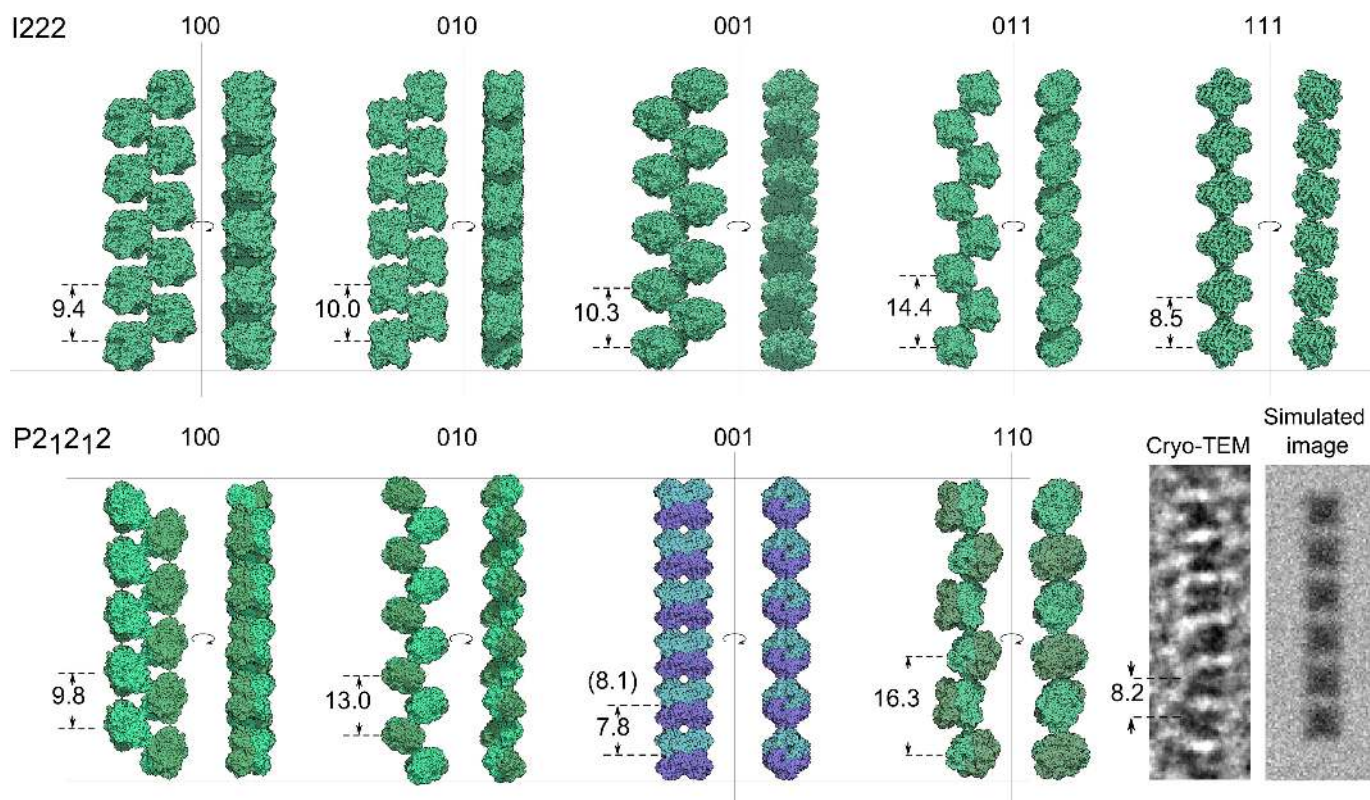
Extended Data Figure 7 | Crystallographic modelling of the nanorods. Models of glucose isomerase nanorods in various directions, based on the unit-cell dimensions of the PDB entries 9XIA and 1OAD, and the crystallographic symmetry elements of space groups I222 and $P2_12_12$. The numbers designating intermolecular distances are in nanometres. The number in brackets for $P2_12_12$ (001) is the value that we obtained experimentally. For reference, we compare a magnified cryoTEM image of a single nanorod and a simulated TEM projection based on the $P2_12_12$ (001) nanorod model.
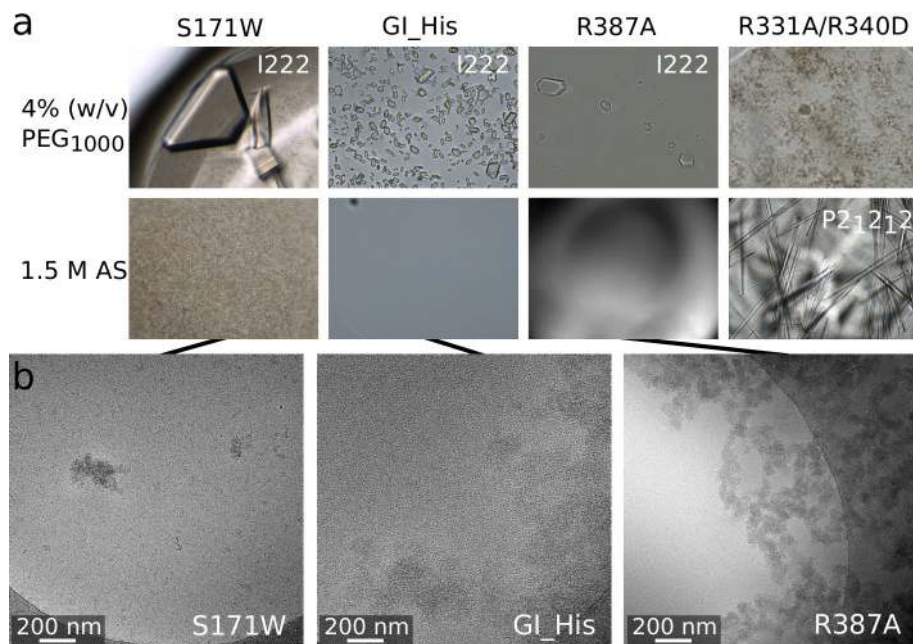
Extended Data Figure 8 | Crystallization screening of glucose isomerase mutants with perturbed lattice contacts. a, Initial crystallization screening of mutants in 50 mM HEPES pH 7.0, 100 mM $MgCl_2$, 15 mg ml^{-1} of glucose isomerase mutant and 4% (w/v) PEG_{1000} or 1.5 M ammonium sulfate. The mutants are S171W (with perturbed C_1 interactions), GI_His (perturbed C_1), R387A (perturbed C_2) and R331A/R340D (perturbed C_3). **b**, Cryo-TEM images of various mutants in 50 mM HEPES pH 7.0, 100 mM $MgCl_2$, 15 mg ml^{-1} mutant protein and 1.5 M ammonium sulfate, 2 minutes after protein/precipitant mixing.

Extended Data Table 1 | Comparison of experimental and theoretical distances for both space groups

I222	Face	$d_{02\bar{2}}$	d_{200}		$P2_12_12$	Face	d_{001}	$d_{\bar{2}20}$
Theoretical	(011)	6,975	4,961		Theoretical	(110)	7.833	8.137
	(101)	7,151	4,703			$(1\bar{1}0)$	7.833	8.137
	(110)	6,836	5,151					
Experimental	AS	6.59 $\pm$0.15	4.72 $\pm$0.10		Experimental	AS	8.07 $\pm$0.12	8.36 $\pm$0.10
	PEG	5.02	7.08					

Distances (d) are in nanometres. AS, ammonium sulfate.

Extended Data Table 2 | Lattice-contact analysis of both space groups

Space Group	Contact	H-bonds	Salt-Bridges	Patch	ΔASA	Charges	Residues
P2₁2₁2	C1	6	0	P1	422	(1−,4−)	**E167**, **T170**, S171, **Q172**, G173, E207, R208, E210, A344, D345, G346, L347, **Q348**, A349
P2₁2₁2	C2	7	2	P2a	320	(3−,3−)	V37, E38, **R41**, E70, K73, **R74**, **Q77**, <u>D81</u>
				P2b	320	(2 ,2−)	A319, A322, D323, P324, **E325**, R374, **G385**, A386, **R387**, G388
I222	C3	3	0	I1a	177	(2−,2−)	R76, **D80**, <u>D81</u>, T82, G83, K85
				I1b	195	(2−,1−)	**R331**, D336, **R340**, P341

The residues listed in the last column have non-zero ΔASA (Å²) values and are therefore considered to be (partially) buried. Residues that are involved in a hydrogen (H) bond or salt bridge with a nearest-neighbour residue are shown in bold and red, respectively. The underlined residue D81 is common to patches P2a and Ia.