STRUCTURAL SCIENCE
CRYSTAL ENGINEERING
MATERIALS

Volume 72 (2016)

Supporting information for article:

An insight into real and average structure from diffuse X-ray scattering - a case study

Michał Leszek Chodkiewicz, Anna Makal, Roman Gajda, Dragoslav Vidovic and Krzysztof Woźniak

S1. 100K Temperature Data Collection

A prism of $[C_{21}H_{27}N_2]^+Cl^-$ of the size of 0.32 x 0.20 x 0.013mm was mounted on a kapton loop with Paratone-N oil. The X-ray intensity data were measured. A total of 903 frames were collected. The total exposure time was 2.51 hours. The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. The integration of the data using a tetragonal unit cell yielded a total of 9109 reflections to a maximum θ angle of 29.90° (0.71 Å resolution), of which 1611 were independent (average redundancy 5.654, completeness = 99.8%, R_{int} = 4.07%, R_{sig} = 2.64%) and 1391 (86.34%) were greater than $2\sigma(F^2)$. The final cell constants of a = 19.924(10) Å, b = 19.924(10) Å, c = 5.088(3) Å, volume = 2020.(2) $Å^3$, are based upon the refinement of the XYZ-centroids of 3578 reflections above 20 σ(I) with 5.773° < 2θ < 59.68°. Data were corrected for absorption effects using the multi-scan method (SADABS). The ratio of minimum to maximum apparent transmission was 0.897.

The structure was solved and refined using the Bruker SHELXTL Software Package, using the space group I 4/m, with Z = 4 for the formula unit, $C_{21}H_{27}ClN_2$. The anisotropic full-matrix least-squares refinement on F^2 with 88 variables converged at R1 = 4.00%, for the observed data and wR2 = 13.18% for all data. The goodness-of-fit was 1.124. The largest peak in the final difference electron density synthesis was 0.421 $e^-/Å^3$ and the largest hole was -0.217 $e^-/Å^3$ with an RMS deviation of 0.048 $e^-/Å^3$. On the basis of the final model, the calculated density was 1.126 g/cm^3 and F(000), 734e^-.

The improved model, with di-methyl-phenyl moieties tilted from the crystallographic $m_{[001]}$ plane was obtained as follows: (1) the di-methyl-phenyl moiety with the adjacent nitrogen atom was defined as a rigid fragment (2) the tilt was introduced with the OLEX2 interactive tool, analogous to the tilt present in the ordered polymorph; the atomic occupancies were adjusted and such a model was refined until convergence, (3) the rigid body restraints were removed; the FLAT restraint, ensuring the planarity of the phenyl rings and similarity restraints DELU were introduced for the atoms belonging to the phenyl rings and this model was refined until convergence.

The final anisotropic full-matrix least-squares refinement on F^2 with 120 variables converged at R1 = 3.78%, for the observed data and wR2 = 12.49% for all data. The goodness-of-fit was 1.127. The

largest peak in the final difference electron density synthesis was 0.470 $e^-/Å^3$ and the largest hole was -0.160 $e^-/Å^3$ with an RMS deviation of 0.046 $e^-/Å^3$.

The structure was also refined in subgroups of I4/m space group down to P1, introducing appropriate number of additional (disordered) molecules. None of these models lead to an R factor lower than 11%, and the very high correlations between the refined parameters, as well as intensity statistics suggested, that lower-symmetry space group cannot yield a better description of the average structure.

S2. Room temperature data collection

A total of 1430 frames were collected. The total exposure time was 3.97 hours. The frames were integrated with the Bruker SAINT software package using a narrow-frame algorithm. The integration of the data using a tetragonal unit cell yielded a total of 8122 reflections to a maximum θ angle of 22.02° (0.95 Å resolution), of which 737 were independent (average redundancy 11.020, completeness = 100.0%, R_{int} = 3.69%, R_{sig} = 1.57%) and 615 (83.45%) were greater than $2\sigma(F^2)$. The final cell constants of a = 20.2750(8) Å, b = 20.2750(8) Å, c = 5.0989(2) Å, volume = 2096.03(18) $Å^3$, are based upon the refinement of the XYZ-centroids of 2668 reflections above 20 σ(I) with 5.681° < 2θ < 43.15°. Data were corrected for absorption effects using the multi-scan method (SADABS). The ratio of minimum to maximum apparent transmission was 0.933.

The structure was solved and refined using the Bruker SHELXTL Software Package, using the space group I 4/m, with Z = 4 for the formula unit, $C_{21}H_{27}ClN_2$. The anisotropic full-matrix least-squares refinement on F^2 with 88 variables converged at R1 = 4.11%, for the observed data and wR2 = 15.61% for all data. The goodness-of-fit was 1.155. The largest peak in the final difference electron density synthesis was 0.120 $e^-/Å^3$ and the largest hole was -0.216 $e^-/Å^3$ with an RMS deviation of 0.034 $e^-/Å^3$. On the basis of the final model, the calculated density was 1.087 g/cm^3 and F(000), 734e^-.

The improved model was obtained by the same procedure as the low-temperature structure. The fully unrestrained refinement tended to converge toward the structure with untilted dimethyl-phenyl rings and nitrogen atoms located on the $m_{[001]}$ plane.

The anisotropic full-matrix least-squares refinement on F^2 with 118 variables converged at R1 = 5.40%, for the observed data and wR2 = 19.60% for all data. The goodness-of-fit was 1.127. The largest peak in the final difference electron density synthesis was 0.325 $e^-/Å^3$ and the largest hole was -0.234 $e^-/Å^3$ with an RMS deviation of 0.043 $e^-/Å^3$.

S3. Electrostatic potential around the chain in OP

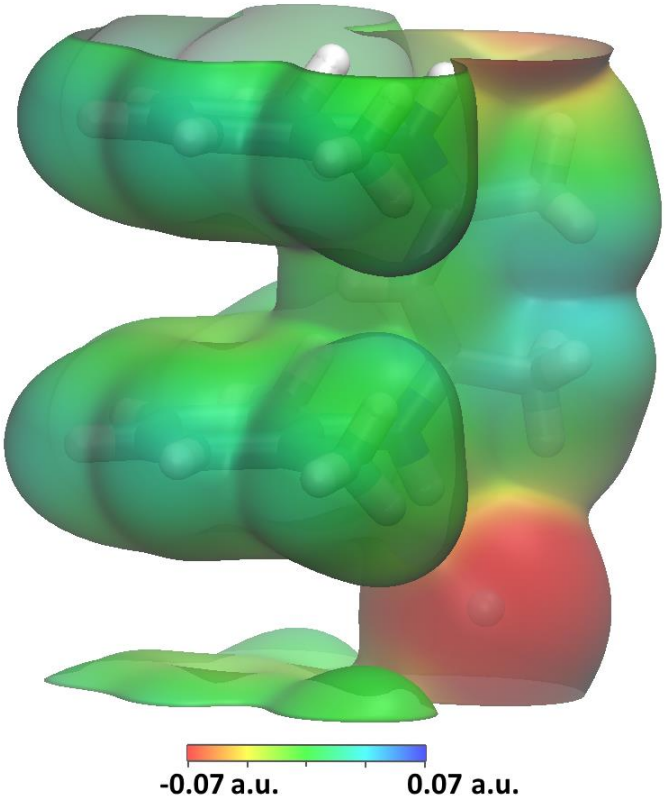

Figure S1 Electrostatic potential map on electron density isosurface (0.002 au) for the chain in **OP**, calculated at B3LYP/6-31g** level with CRYSTAL09 (Dovesi *et al.*, 2009). The chain has been treated as a 1D periodic system, with period equal to the **c** length in **OP**. Calculations have been carried out using resources provided by Wroclaw Centre for Networking and Supercomputing (http://wcss.pl), grant No. 115. The image was made with VMD software support. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

S4. Systematic extinction for hkl (l=n) planes

It will be shown that if sites $s_1, s_2, .., s_n$ can be occupied by one of the two alternative, building blocks ($s_{k\uparrow}$ or $s_{k\downarrow}$) and if the building blocks are periodic along the **c** direction, and related by translation by

one period along that direction (in terms of the average structure lattice) then the diffuse scattering is absent at *h k l*=n layers. Neglecting displacive disorder diffuse scattering can be expressed in the following way (see e.g. Weber & Simonov, 2012 Eq. (7) after grouping atoms into building blocks, see also a comment to Eq. (6)):

$$I_{\mathrm{diff}}(\mathbf{h}) = \sum_{k,k'} \sum_{v,v'} F_{s_k,v}(\mathbf{h}) F^*_{s_k',v'}(\mathbf{h}) \sum_{\mathbf{u}} \left[\mathrm{P}_{s_k,v,s_{k'},v',\mathbf{u}} - \mathrm{P}_{s_k,v} \mathrm{P}_{s_{k'},v'} \right] \mathrm{e}^{2\pi \mathrm{i} \mathbf{h}^{\mathrm{T}} \mathbf{u}}$$

The first sum runs over sites *s* in the unit cell and the second sum over the two possible variants *v* of the building blocks. The last sum runs over unit cell separations **u** = *m***a**+*n***b**. $\mathrm{F}_{s,v}$ is a structure factor of a variant *v* of a column at site *s* calculated with respect to the origin of the unit cell it belongs to, $\mathrm{P}_{s,v,s',v',\mathbf{u}}$ is a joint probability of finding a variant *v* of column at site *s* and variant *v'* of column at site *s'* with the site's unit cells separated by vector **u** and $\mathrm{P}_{s,v}$ is a probability (occupancy) for a variant *v* of the column at site *s*. $F_{s_k,v}(\mathbf{h})$ are structure factors of the building blocks. It is convenient to use rearranged form of the above equation:

$$I_{\mathrm{diff}}(\mathbf{h}) = \sum_{\mathbf{u}} \mathrm{e}^{2\pi \mathrm{i} \mathbf{h}^{\mathrm{T}} \mathbf{u}} \sum_{k,k'} \sum_{v,v'} F_{s_k,v}(\mathbf{h}) F^*_{s_{k'},v'}(\mathbf{h}) \left[\mathrm{P}_{s_k,v,s_{k'},v',\mathbf{u}} - \mathrm{P}_{s_k,v} \mathrm{P}_{s_{k'},v'} \right] \qquad \text{(S1)}$$

Because of the translational relation between the same site variants of the building blocks the following relation holds for their structure factors for $l = n$: $F_{s_k\uparrow} = F_{s_k\downarrow}$, which let us write the last sum in Eq. (S1) for arbitrary pair of sites $s_k, s_{k'}$ and arbitrary **u** as:

$$\sum_{\mathrm{v},\mathrm{v}'} F_{\mathrm{s_k,v}} F^*_{s_{k'},\mathrm{v}'} \left[\mathrm{P}_{\mathrm{s_k,v},s_{k'},\mathrm{v}',\mathbf{u}} - 1/4 \right]$$
$$= F_{\mathrm{s_k},\uparrow}(\mathbf{h}) F^*_{s_{k'},\uparrow}(\mathbf{h}) \left[\mathrm{P}_{\mathrm{s_k},\uparrow,s_{k'},\uparrow,\mathbf{u}} + \mathrm{P}_{\mathrm{s_k},\downarrow,s_{k'},\downarrow,\mathbf{u}} + \mathrm{P}_{\mathrm{s_k},\uparrow,\mathrm{s}',\downarrow,\mathbf{u}} + \mathrm{P}_{\mathrm{s_k},\downarrow,s_{k'},\uparrow,\mathbf{u}} - 1 \right]$$

The sum in the square bracket on the right hand side of the above equation is equal to zero. It can be shown by inserting the following relations between the joint probabilities into the expression in brackets:

$$\mathrm{P}_{\mathrm{s_k},\uparrow,s_{k'},\uparrow,\mathbf{u}} + \mathrm{P}_{\mathrm{s_k},\uparrow,\mathrm{s}',\downarrow,\mathbf{u}} = \mathrm{P}_{\mathrm{s_k},\uparrow}$$
$$\mathrm{P}_{\mathrm{s_k},\downarrow,s_{k'},\downarrow,\mathbf{u}} + \mathrm{P}_{\mathrm{s_k},\downarrow,s_{k'},\uparrow,\mathbf{u}} = \mathrm{P}_{\mathrm{s_k},\downarrow}$$
$$\mathrm{P}_{\mathrm{s_k},\uparrow} + \mathrm{P}_{\mathrm{s_k},\downarrow} = 1$$

Therefore the total diffuse scattering for the layers described is also zero.

S5. Conditional probabilities

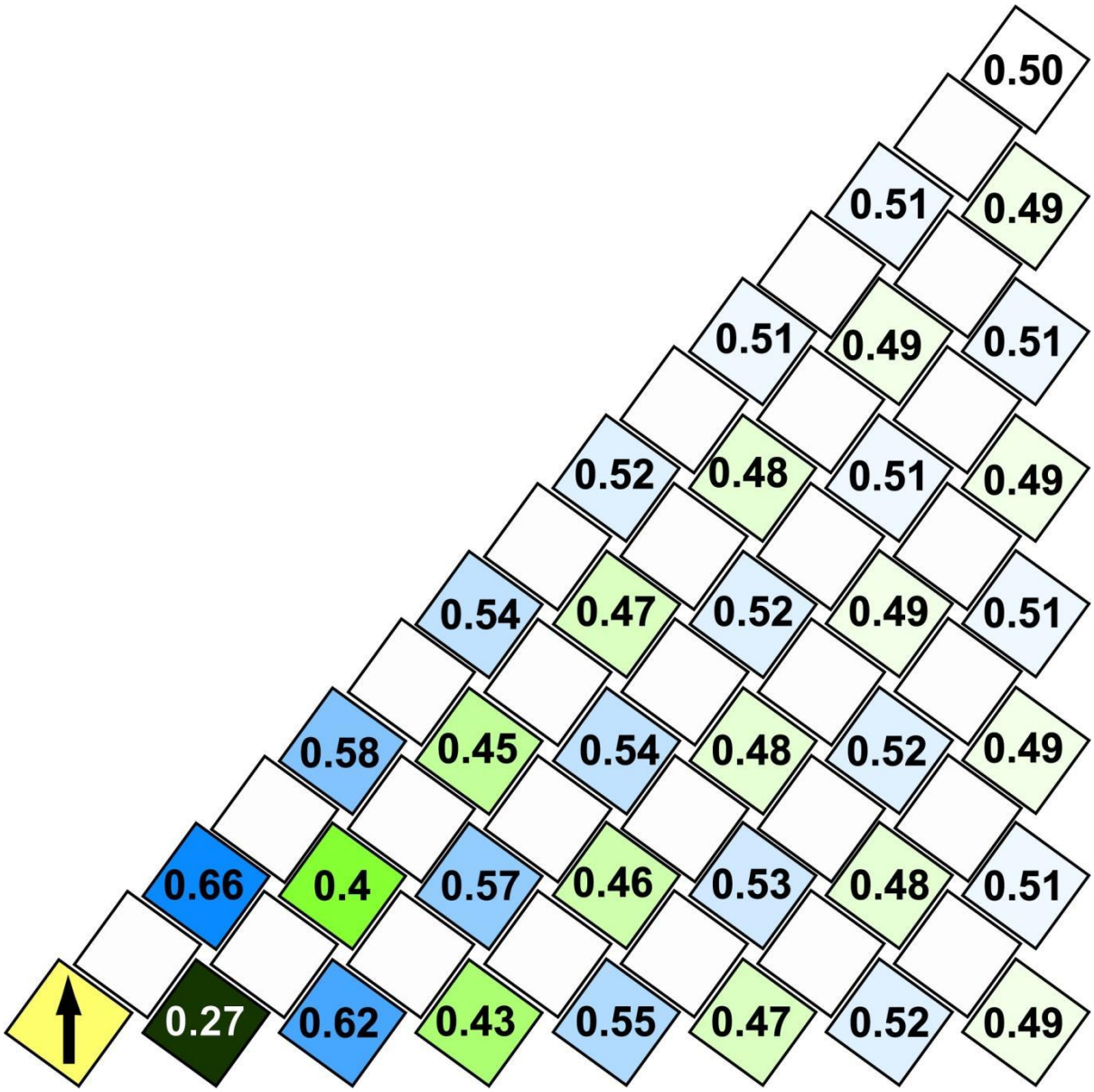

Figure S2 Conditional probabilities $P(\uparrow \mid \uparrow, \boldsymbol{u})$ of finding column of the same variant. It is related to joint probabilities and the $c_{m,n}$ coefficients in the following way: $P(\uparrow \mid \uparrow, \boldsymbol{u}) = 2P_{\uparrow,\uparrow,\mathbf{u}} = c_{\mathbf{u}}/2 + 1/2$.

S6. Diffuse scattering and column scattering

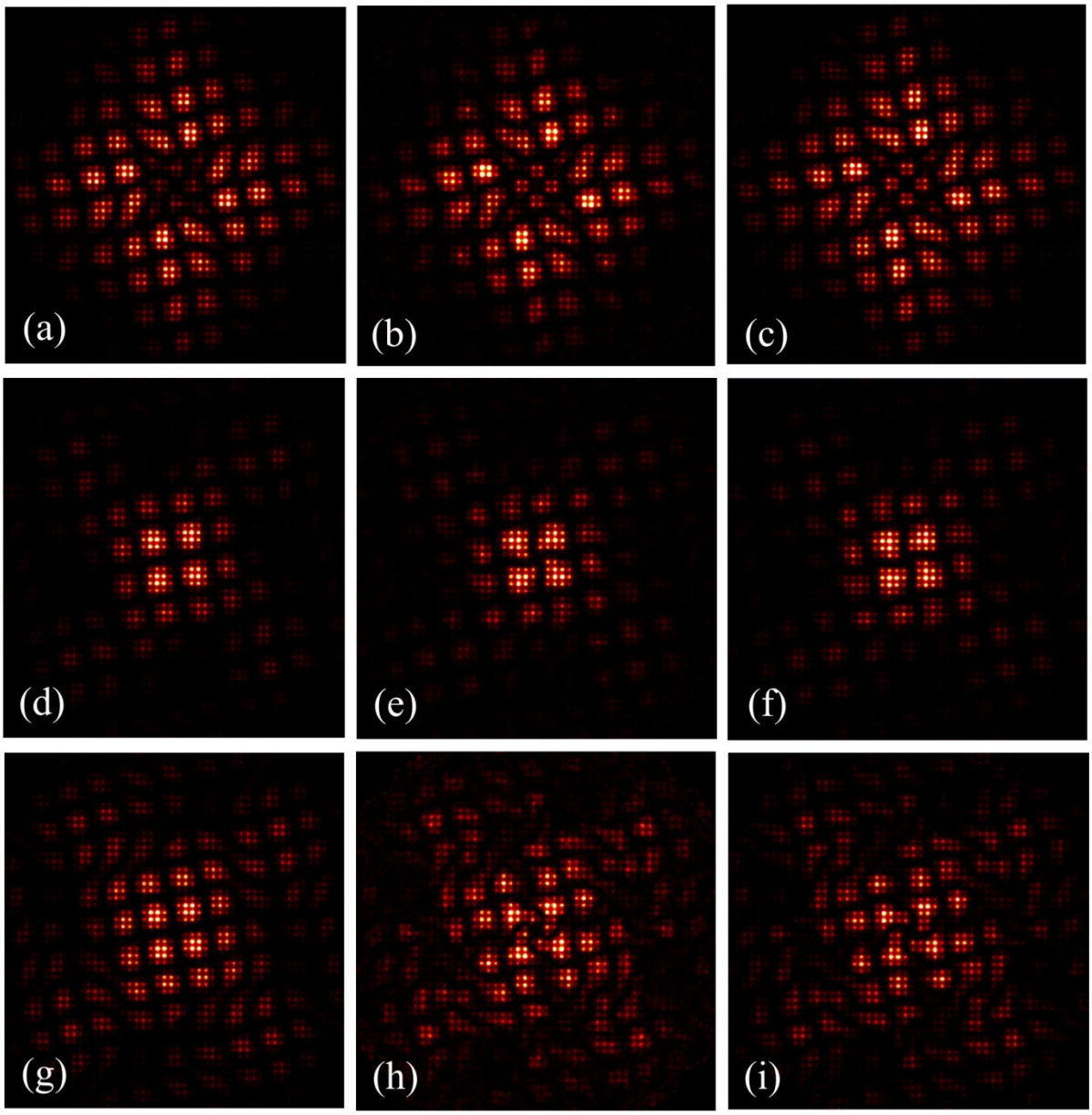

Figure S3 Diffuse scattering measured at 100 K and calculated one: (*h k* 0.5) (a) calculated for the initial average structure model with parallel phenyl rings (discussed in paragraph 4) (b) experiment, (c) calculated for the final average structure model with tilted phenyl rings (the same model was used to generate Fig. 9) (d)-(f) (*h k* 1.5) in the same order as for (*h k* 0.5), (g)-(i) (*h k* 2.5) in the same order as for (*h k* 0.5).

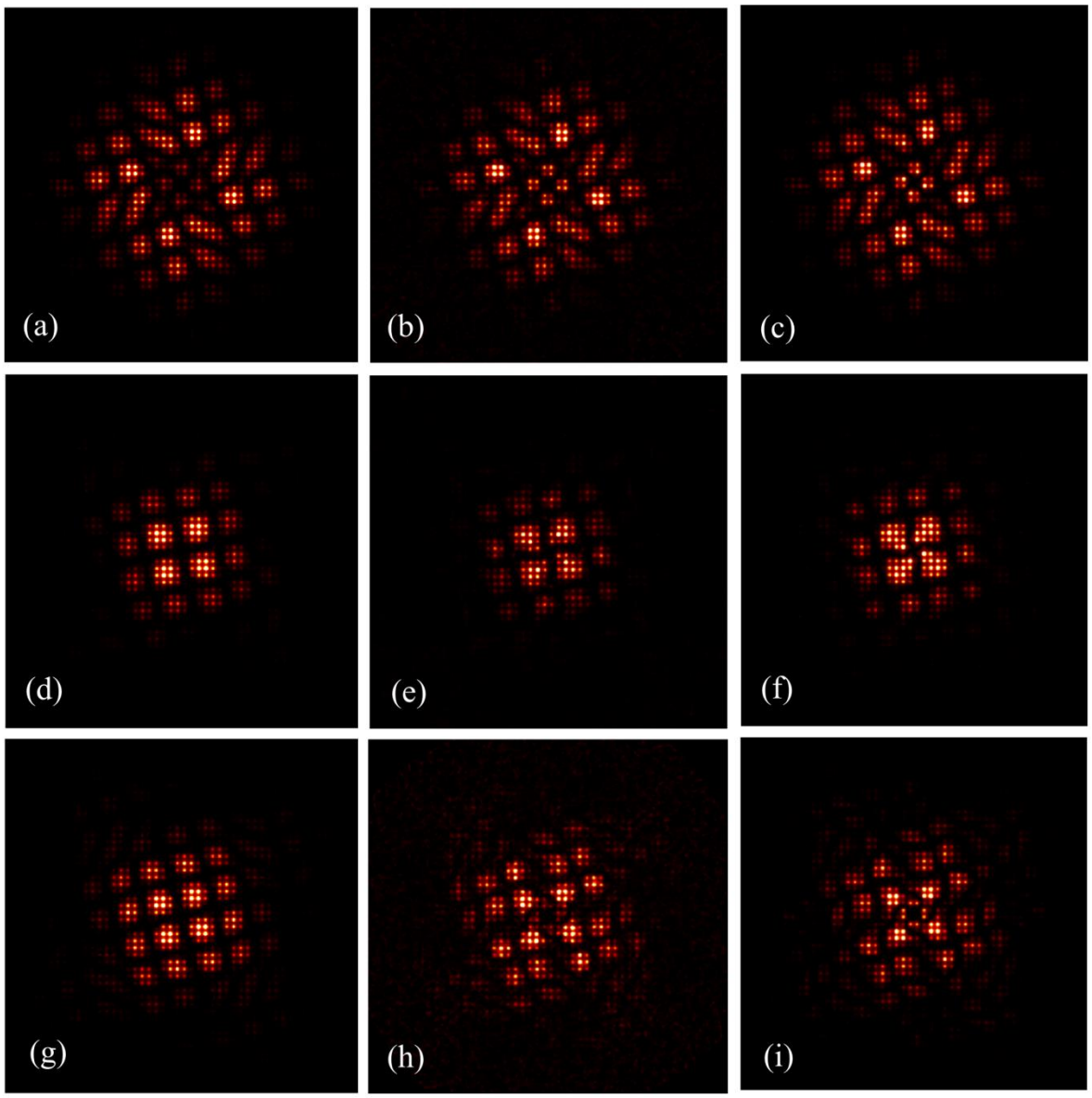

Figure S4 Diffuse scattering measured at 298 K: *(h k* 0.5) (a) calculated for the initial average structure model with parallel phenyl rings (discussed in paragraph 4) (b) experiment, (c) calculated for the final average structure model with tilted phenyl rings (the same model was used to generate Fig. 9) (d)-(f) *(h k* 1.5) in the same order as for *(h k* 0.5), (g)-(i) *(h k* 2.5) in the same order as for *(h k* 0.5).

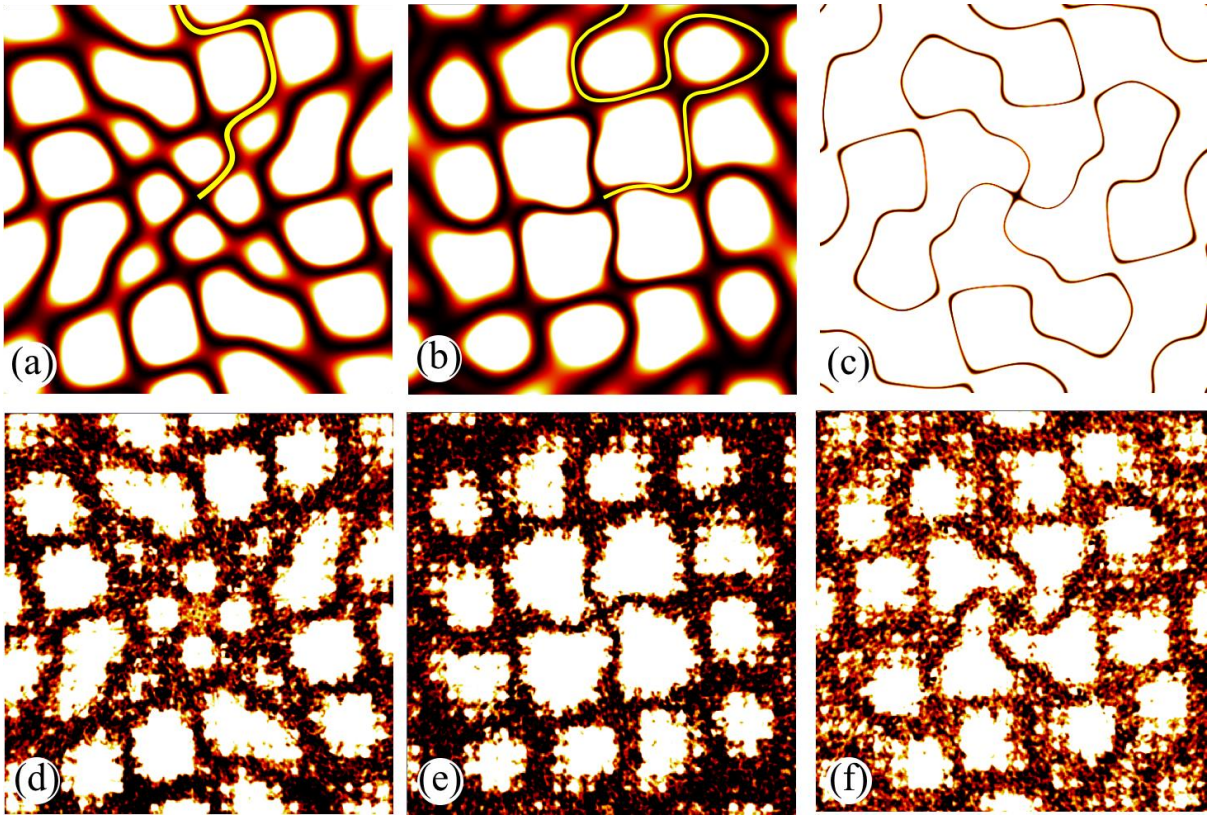

Figure S5 Single column scattering intensity $I_{\uparrow}(\boldsymbol{h})$ in plane (a) (*h k* 1/2), (b) (*h k* 3/2) and (c) (*h k* 5/2) - areas colored with white correspond to $I_{\uparrow}(\boldsymbol{h})$ values higher than a threshold (different for different planes), in (a) and (b) the extinction lines marked with yellow line, in (c) significantly lower threshold was used to show the lines. (d)-(f) Corresponding planes for experimental diffuse scattering I_{diff} ((d) (*h k* 1/2), (e) (*h k* 3/2) and (f) (*h k* 5/2)) - areas where $I_{diff}(\mathbf{h})$ values are larger than 0.01 of the maximal value of I_{diff} are colored with white.

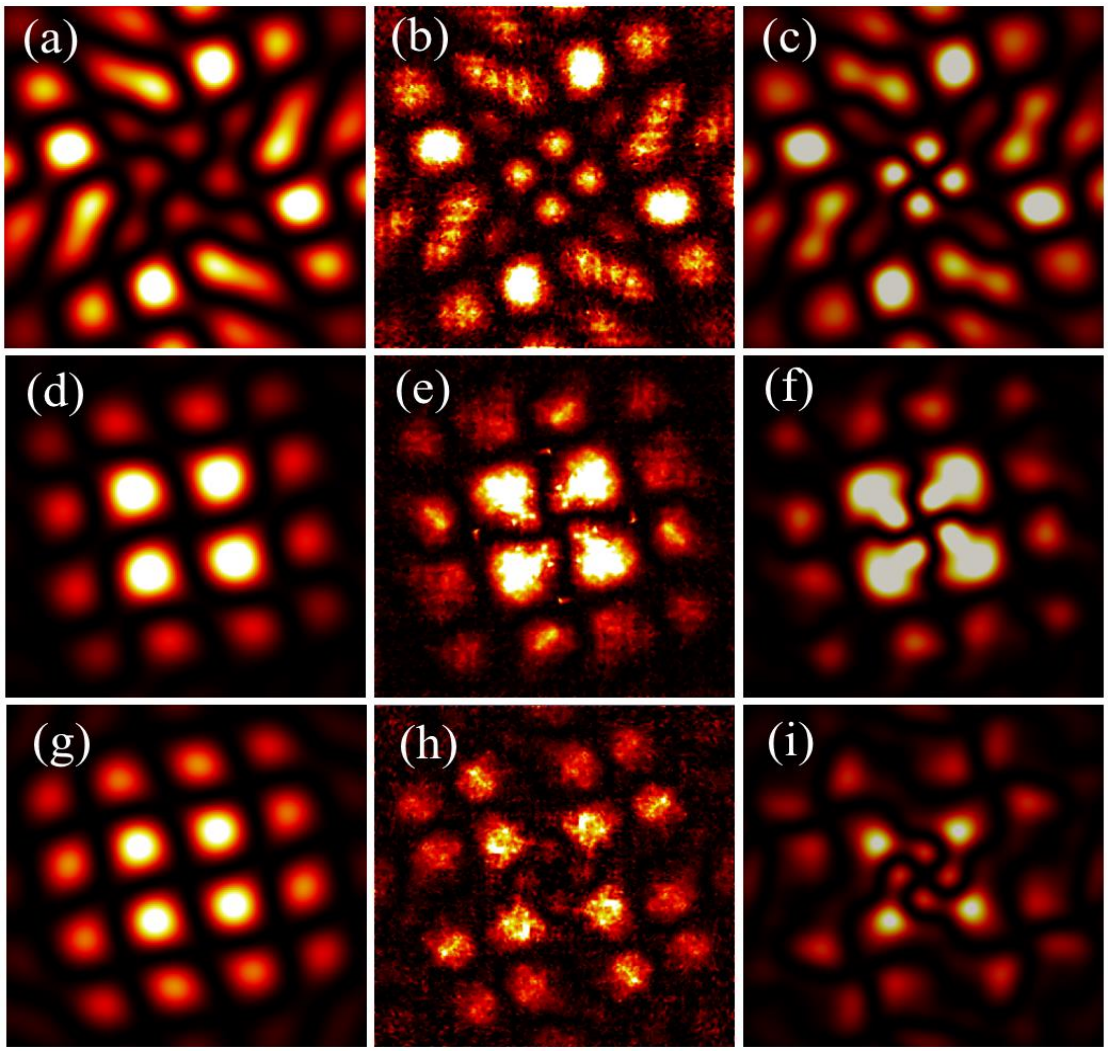

Figure S6 Scattering intensities from the column for room temperature structure at (*h k* 0.5): (a) model with parallel phenyl rings, (b) experiment, (c) model with tilted phenyl rings, (d)-(f) for (*h k* 1.5) in the same order as for (*h k* 0.5), (g)-(i) for (*h k* 2.5) also in the same order.

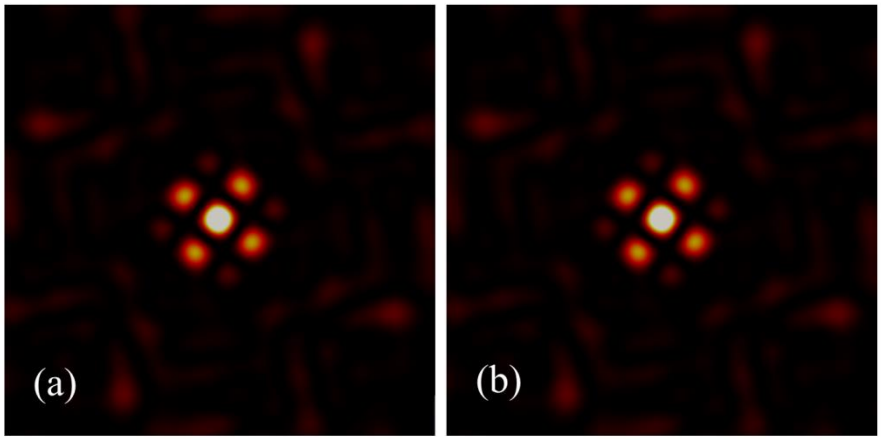

Figure S7 An example of column scattering intensity at integer *l* (*l*=2) for the structural models with (a) tilted and (b) parallel rings

S7. Ambiguity in determination of occupations correlations between columns of different types

Occupational correlations which do not contribute directly to DS cannot be retrieved from the DS alone since there is no guarantee that the results are unambiguous. Here we present simulations illustrating such ambiguities in the context of the studied system. The interactions model used previously was extended to include also first and third neighbor interactions:

$$E = -\sum_{t}\sum_{m,n}\left[J\left(\sigma^{t}_{m,n}\sigma^{t}_{m+1,n} + \sigma^{t}_{m,n}\sigma^{t}_{m,n+1}\right) + J_3\left(\sigma^{t}_{m,n}\sigma^{t}_{m+1,n+1} + \sigma^{t}_{m,n}\sigma^{t}_{m-1,n+1}\right)\right]$$
$$-J_1\sum_{m,n}\left(\sigma^{G}_{m,n}\sigma^{Y}_{m,n} - \sigma^{G}_{m,n}\sigma^{Y}_{m-1,n} + \sigma^{G}_{m,n}\sigma^{Y}_{m-1,n-1} - \sigma^{G}_{m,n}\sigma^{Y}_{m,n-1}\right)$$

J_1 and J_3 describe respectively first and third neighbor interactions, column types in the last sum were denoted as G and Y (corresponding to grey and yellow symbols of the columns in Fig. S8

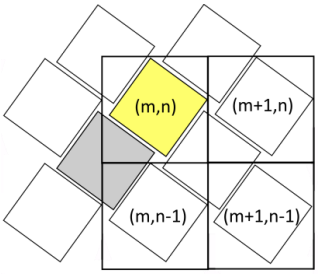

Figure S8 Unit cells and associated column sites arrangement as used in the reported MC simulations (the sites corresponding to the cell with indices (m,n) are colored).

MC simulations were performed to test if it were possible to reproduce the function $S(h,k)$ (and therefore also the corresponding DS) with different parameterizations of the model resulting in different occupational correlations between columns of different types. The model with only second neighbor interactions was used as a reference (with J= -0.36 k_BT). For the extended model two values of J were chosen: -0.4 k_BT and -0.36 k_BT and MC simulations were performed for a range of J_1 and J_3 values. For each set of parameters the discrepancy measure M_2 (Eq. (3)) between the reference $S(h,k)$ and the one corresponding to the simulation was calculated (Fig. S9(a),(b)). Also conditional joint probabilities (**cjp**) of finding a variant of 'grey' column provided that the variant of 'yellow' column present in the same unit cell is denoted with the same value of σ were calculated. For uncorrelated occupancies of the columns in the reference model such **cjp** is equal to 0.5. It was possible to reproduce the reference $S(h,k)$ with the value of the agreement factor (M_2) below 0.05 and the value of the **cjp** ranging from ~0.25 to ~0.75 ((Fig. S9(c),(d))). This shows the source of ambiguity in determination of the **cjp** - large changes in the **cjp** can correspond to only slight changes in the agreement factors. There is a possibility that for more complex models of interactions the changes in **cjp** can correspond to virtually no change in agreement factor. The values of M_2 are overestimated because of limited sampling in MC simulation (simulation with the same parameters as in the

reference system results in M_2=0.014). Using finer grid in the parameters space would result in finding points with the same **cjp** and lower M_2. Extending the model with further energy terms would most probably allow for further reduction of the error while keeping discrepancy in the **cjp** at the same level. Since different parameters of the model can lead to similar DS, values of the parameters optimized against experimental data would be highly dependent on the experimental errors and affected by influence of instrumental resolution function on DS.

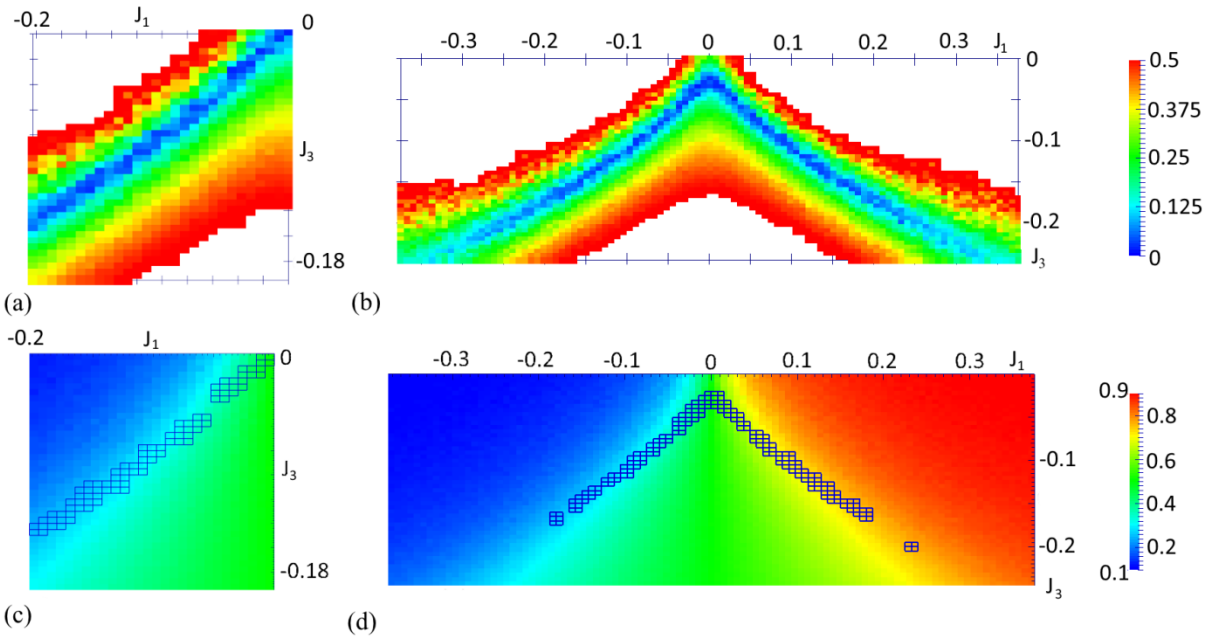

Figure S9 (a) and (b) - Estimates of discrepancy measure M_2 between $S(h,k)$ calculated for the reference model and models parameterized with different J_1 and J_3 parameters and (a) with J= -0.36 k_BT (as in the reference system) and (b) J= -0.4 k_BT. (c) and (d) - Conditional joint probabilities of finding specific variant of 'grey' column in unit cell provided that a variant of 'yellow' column such that $\sigma^G_{m,n} = \sigma^Y_{m,n}$ is present in the unit cell. Pixels corresponding to $M_2 < 0.05$ are outlined.

S7.1. Details of the calculations

Disordered crystals of size 100 by 100 unit cells were simulated for each set of parameters. During the simulations joint probabilities were sampled and the estimates of the probabilities were used to calculate $c_{m,n}$ coefficients (for $m,n \leq 10$). Neglecting the remaining $c_{m,n}$ introduces only small error to estimates of M_2 (about 0.0015 for the reference model). The coefficients are used for estimation of the agreement factor M_2. Summation in the expression for M_2 (Eq. (3)) is replaced by integration over period of identity of $S(h,k)$ and the integration is performed analytically.