Influence of chemical modifications of the crystallophore on protein nucleating properties and supramolecular interactions network.

Amandine Roux†,*[a,c] Zaynab Alsalman†,[b] Tao Jiang,[a] Jean-Christophe Mulatier,[a] Delphine Pitrat,[a] Elise Dumont,[e,f] François Riobé,[a,d] Natacha Gillet,*[a] Eric Girard*[b] and Olivier Maury*[a]

[a] Dr. A. Roux, T. Jiang, J.-C. Mulatier, D. Pitrat, Dr. N. Gillet, Dr. F. Riobé, Dr. O. Maury, Univ. Lyon, École Normale Supérieure de Lyon, CNRS UMR 5182, Laboratoire de Chimie, 46 allée d'Italie, France.
E-Mail: amandine.roux-gossart@ens-lyon.fr; natacha.gillet@ens-lyon.fr; olivier.maury@ens-lyon.fr
[b] Dr. Z. Alsalman, Dr. E. Girard, Univ Grenoble Alpes, CEA, CNRS, IBS, F-38000 Grenoble, France.
E-mail: eric.girard@ibs.fr
[c] Polyvalan SAS, Lyon, France
[d] Univ Bordeaux, Bordeaux INP, CNRS, Institut de Chimie de la Matière Condensée de Bordeaux, 33608 Pessac, France.
[e] Univ. Côte d'Azur, CNRS, Institut de Chimie de Nice, UMR 7272, 06108 Nice, France
[f] Institut Universitaire de France, 5 rue Descartes, 75005 Paris, France
† These authors contributed equally to the experimental work.

Supporting information for this article is given via a link at the end of the document.

Abstract: Crystallophores are lanthanide complexes that have demonstrated outstanding induction of crystallization for various proteins. This article explores the effect of tailored modifications of the crystallophore first generation and studies their impact on the nucleating properties, and protein crystal structures. Through high-throughput crystallization experiments and dataset analysis, we evaluated the effectiveness of these variants, in comparison to the first crystallophore generation **G₁**. In particular, the **V₁** variant, featuring a propyl-3-ol pendant arm, demonstrated the ability to produce new crystallization conditions for the proteins tested (hen-egg white lysozyme, proteinase K and thaumatin). Structural analysis performed in the case of hen egg-white lysozyme along with Molecular Dynamics simulations, highlights **V₁**'s unique behavior, taking advantage of the flexibility of its propyl-3-ol arm to explore different protein surfaces and form versatile supramolecular interactions.

Introduction

The determination of the 3D structure of biological macromolecules and particularly proteins at the atomic resolution is a major step to identify key structural features related to their biological function.[1] This research endeavor holds significant societal importance since understanding the structure and function of proteins offers the potential to design inhibitors or drugs for combatting dysregulation processes, pathogenic bacteria, and viruses responsible for various diseases. X-ray crystallography (XRD) has remained the cornerstone of structural biology for half a century with 86% of protein structures deposited in the Protein Data Bank (https://www.rcsb.org/). Despite the emergence of alternative techniques, like NMR or cryo-electron microscopy, XRD remains the experimental technique of choice for structure guided drug discovery and atomic resolution understanding of macromolecules. The two major bottlenecks of XRD are the production of high-quality diffracting crystals and the solving of the phase problem to compute the electron density map. This considerably limits the overall success of crystal structure determination to less than 15% according to structural genomics statistics.[2]

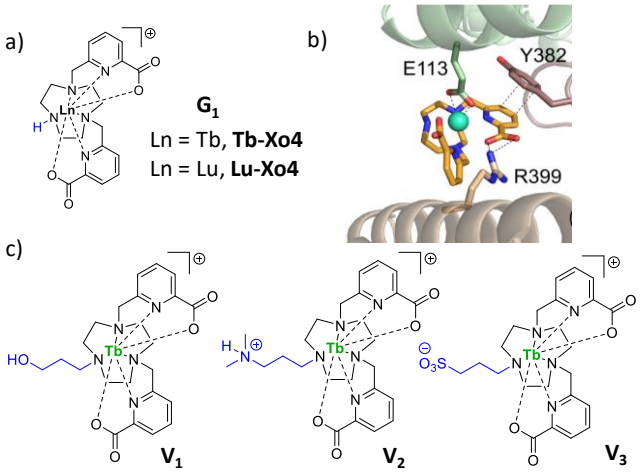


Figure 1. a) Molecular structure of the reference crystallophore family (**G₁**); b) typical interactions network between Tb-Xo4 and a protein, here FprA from Methanothermococcus thermolithotrophicus. c) Molecular structures of the variants **V₁₋₃**. In blue are indicated the structural modifications.

The emergence of artificial intelligence and deep learning software namely AlphaFold2 or RoseTTAFold, for structure prediction, and the development of fragment-based molecular replacement have provided a robust solution to the phase problem in crystallography.[3] Consequently, phasing, although remaining challenging, is no longer seen as a major issue.
On the other hand, the modern approach towards overcoming limitations in protein crystallization has mainly focused on technological achievements. In this context, the development of even more intense radiation sources, such as XFEL, coupled with highly sensitive detectors enables to address smaller-size crystals expanding the scope of structural analysis. In addition, the use of high-throughput crystallization platforms has become a

standard practice to screen an even larger number of conditions with a reduced amount of precious biological sample.[4] Comparatively, methodological developments aiming to design new additives to facilitate protein crystallization remains a largely underexplored high potential domain.

Indeed, the design of additives to control the protein crystallization process and ultimately improve the crucial nucleation step is a research domain still in its infancy. Beneath the empirical approaches involving the addition of heterogeneous "impurities" to promote nucleation,[5] a more rational approach has emerged over the last decade, based on supramolecular chemistry and centered on the understanding of the interactions between protein and additives. Three main classes of molecular soluble crystallization additives have been developed, including polyoxometalates,[6] anionic macrocycles (phosphonated or sulfonato-calix[4,6]arenes, cucurbituryl…),[7] and lanthanide complexes[8] among which the recently reported crystallophore, Ln-Xo4 (Ln = Tb or Lu, Figure 1).[9]

All these additives are strategically positioned at the interface between neighboring protein molecules within the crystal without disrupting the protein native fold and thus they act as "molecular glue", facilitating the crystal contact between proteins molecules.[7c, 7e, 9c, 10] The crystallophore family presents additional phasing ability due to the exceptional anomalous properties of f-block elements such as in the L_{III} absorption edge of Tb(III) and Lu(III) at 1.65 Å and at 1.34 Å, respectively.[9b] Consequently, they can be used as "*all-in-one*" additives for protein X-ray crystallography. Their versatile interactions in protein crystals have been elucidated using an integrated methodology combining XRD and molecular simulations (Figure 1). A mixed interaction network has been revealed including the direct coordination of carboxylate amino-acid (AA) (aspartate and glutamate), the formation of H-bonds network with protic AA (arginine, lysine..) and hydrophobic interaction with aromatic AA (π-π, CH-π with phenylalanine, tryptophan, tyrosine...).[9c] It is worth noting that the particular chemical structure of the Ln-Xo4 family is necessary to induce such nucleating properties with proteins since, in our hands, related complexes based on other macrocycles such as cyclam, cyclen, pyclen and many others have not shown similar behavior. This discovery indicates that minor structural modifications of the lanthanide complex may have substantial consequences on the overall protein crystallization process.

Herein, we explored the significant impact of modifications of the crystallophore structure, especially the introduction of various pendant arms at the free amino position of the TACN macrocycle (**V₁₋₃**, Figure 1), carrying different charges, on the overall protein crystallization process. The three variants have been involved in automated high-throughput crystallization experiments with three model proteins (lysozyme (HEWL), proteinase K (protK) and thaumatin (TdThau)) and the outcomes were subsequently analyzed using an extended version of our recently reported dataset analysis.[9b] The XRD structures of three {**Vᵢ**/HEWL} co-crystals (i =1-3)), as well as the one of {Tb-Xo4/HEWL} for unbiased comparison, have been determined and each variant interaction network has been thoroughly investigated using molecular dynamics simulations. Finally, compared to the reference Tb-Xo4 molecule (denoted as **G₁** for generation 1), the variant **V₁** featuring propyl-3-ol pendant arm exhibits the best results and is proposed as the second generation of crystallophore, complementary to the first one.

Results and Discussion

Synthesis and characterization. The three target complexes, called variants **V₁₋₃**, conserve the fundamental TACN-bis-picolinate structure of the crystallophore but present an additional substitution at the free amino position of the macrocycle by hydroxyl (**V₁**), ammonium (**V₂**) or sulfonate (**V₃**) propyl group. This peripheral substitution results in a variation of the overall complex charge shifting from +2 for **V₂** to 0 for the zwitterionic **V₃**. The synthesis of the ligands follows classical procedures and involved the N-alkylation of the free triazacyclononane amine (Scheme S1). After saponification of the ester groups, the reaction with $TbCl_3.6H_2O$ led to the desired complexes after dialysis and/or HPLC purification. Experimental protocols, characterizations and purification procedures are detailed in the Supporting information (Figures S1-S11).

Evaluation of the nucleating properties. The nucleating ability of each variant **V₁₋₃** was evaluated using automated crystallization experiments, conducted at the HTXlab (EMBL, Grenoble). These experiments were carried out with 6 standard commercial crystallization kits, corresponding to a total of 576 crystallization conditions (Table S1) and with three commercial proteins of known structure: HEWL, protK and TdThau.

We first investigated the behavior of the variants in the crystallization kit in the absence of any protein. At optimal concentration for inducing crystallization (*i.e*. 10 mM) [9a], Tb-Xo4 did not present any self-crystallization.[11] All variants displayed similar behavior with only rare self-crystallization or precipitation events (less than 10 occurrences per variant over 576 crystallization conditions) that generally disappear in the presence of the protein (see SI for details, Figures S12-14).

In our comparative HTX experiments, the crystallization of a given protein was carried out in three distinct conditions: in native conditions (*i.e.* protein alone), with Tb-Xo4 (**G₁** for first generation of crystallophore) and with one of the evaluated variant (**V₁**, **V₂** or **V₃**) taking advantage of the possibility to set up three hanging drops per well at the same time (Crystal Direct plates, sitting drop setup[12], see SI for details). After 30 days, drop inspections were conducted to detect crystallization events (referred to as "hits"). Each drop was attributed a score ranging from 1 to 6 generating a dataset of 576 numbers, one for each crystallization condition. In this study, only scores 4, 5, 6 were considered as "hits", with score 6 indicating the presence of distinct single crystals, 5 for needle-like crystals and 4 for the presence of urchins or micro-crystals (see SI for examples, Figures S15-17).

Given that the three variants were evaluated in similar conditions, the HTX experiments generated three (or four) datasets for both native and Tb-Xo4 (**G₁**) conditions. This large amount of data allows us to evaluate the reproducibility of the results (Figure S18 and Table S2). A simple count of the hits for each experiment suggests an apparent reproducibility of the HTX crystallization experiments. In order to quantify the dispersion of these results, we calculated the relative standard deviation (RSD), a classical statistic parameter defined by the ratio between the standard deviation to the mean and expressed in % (see SI for detail). For the HEWL and protK proteins, the RSD is estimated to be 27 and 25%, respectively, that significantly decreases in the presence of crystallophore **G₁** to 11 and 7%, respectively. For the TdThau protein, the RSD remains constant with or without **G₁** to about 34% (Table S2). Nevertheless, it seems that the presence of crystallophore has a positive impact on the reproducibility of the

protein crystallization process. However, this apparent reproducibility masks a more nuanced reality when we analyzed not only the overall number of hits but their consistency across the crystallization kit using a '*by plate*' analysis of the dataset.[9b] For example, considering the crystallization of HEWL in the presence of Tb-Xo4, the four different datasets (Figure S18 and Table S2) yield around 200 hits (score 4, 5, 6) for each experiment. It is worth noting that only 101 hits were consistently observed in all 4 datasets, 64 appear three times and 70 twice. In addition, 66 hits have been detected only once (Figure S21). Similar observations are made if we consider only the presence of single crystal (score 6) and can be generalized for all proteins both in native conditions or in the presence of the crystallophore (Figures S20–S25). The observed dispersion of the hits can be attributed to the intrinsic stochastic character of the protein crystallization process associated with an important redundancy of certain conditions in the commercial crystallization screens. As a consequence, to provide a better description of the crystallization space (including the dispersion), we introduced a *cumulated* value encompassing all the different hits collected from the 3 or 4 different crystallization assays (Figures S18 and S19). These cumulated datasets then serve as the reference for the evaluation of the different variants (**V**$_{1\text{-}3}$).

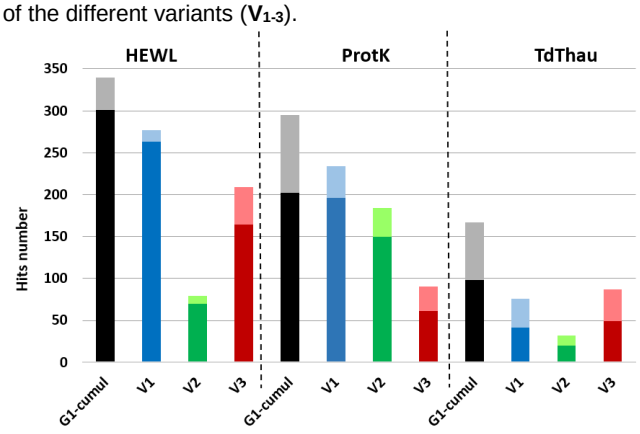


Figure 2. Analysis of the hit number of the automated crystallization experiments performed with the different additive and the three model proteins HEWL, protK and TdThau. The dark colored part indicates the 6 score (mono-crystal) and the light colored one, the additional 4 and 5 scores (needle, urchin and micro-crystals).

Figure 2 represents the number of hits obtained for the three model proteins during co-crystallization with the cumulated data of the first generation of crystallophore Tb-Xo4 and the three variants (**V**$_{1\text{-}3}$). (Data from each separated run of the first generation of crystallophore Tb-Xo4 are shown in Figure S18). At first glance, the crystallization outcomes vary depending on the considered crystallophores (**G**$_1$ and **V**$_{1\text{-}3}$) highlighting the significant influence of the nature of the lateral substituents. The nucleation efficiency of the three variants is of course influenced by the protein under consideration and the following trends are observed: **V**$_1$>**V**$_3$>**V**$_2$ for HEWL, **V**$_1$>**V**$_2$>**V**$_3$ for protK and **V**$_3$≈**V**$_1$>**V**$_2$ for TdThau (Figure 2). In all cases, **V**$_1$, featuring the propyl-3-ol pendant arm exhibits the highest hit number that compares favorably with the Tb-Xo4 reference (**G**$_1$) even using less favorable data, such as the cumulated dataset. Furthermore, it is important to emphasis that **V**$_1$ also presents a very large proportion of score 6 indicating its ability to induce the formation of well-defined single crystals ready-to-use for XRD analysis. This proportion is particularly significant in the cases of HEWL (95%) and protK (80%). Some selected and representative crystals are depicted in Figure 3a-c. Of course, the other variants were also able to promote the formation of single crystals (Figure 3d-i) but in lower amounts.

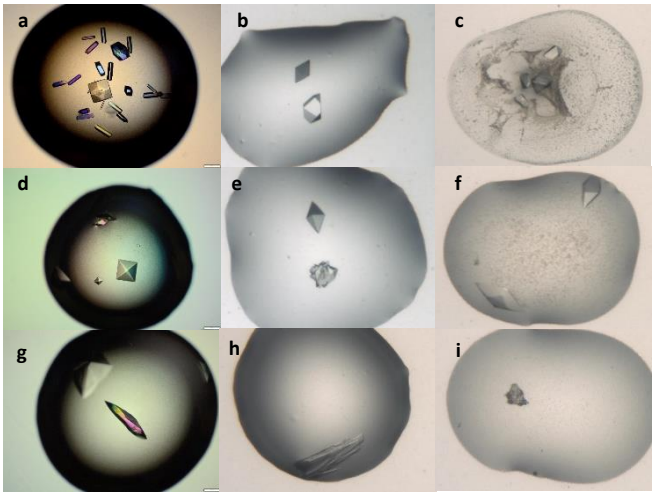


Figure 3. Pictures of {protein/**V**$_i$} co-crystals using either Olympus microscope under polarized light or from the HTX imaging system. In parenthesis are indicated the name of the commercial kit and the crystallization condition. a) {HEWL-**V**$_1$} (PACT premier: F02); b) {protK-**V**$_1$} (PEGs-I: F05); c) {TdThau-**V**$_1$} (Wizard I & II: A01); d) {HEWL-**V**$_2$} (PACT premier: D07); e) {protK-**V**$_2$} (PEGs-I: H11); f) {TdThau-**V**$_2$} (Wizard I & II: H01); g) {HEWL-**V**$_3$} (PACT premier: G12); h) {protK-**V**$_3$} (The Classics Suite: F06); i) {TdThau-**V**$_3$} (The Classics Suite: G12).

We have demonstrated that the new variants, especially **V**$_1$, exhibit interesting nucleating properties using our set of model proteins and are able to produce high-quality single crystals. The main question now concerns the ability of the variants to induce additional unique crystallization conditions when compared to both the native protein and the reference Tb-Xo4 (**G**$_1$).

To that end, we must simultaneously compare, for each variant and each protein, three different crystallization datasets namely the cumulated native hits (dataset(**N**^{cum})), the cumulated Tb-Xo4 hits (dataset(**G**$_1^{cum}$)) and the variant hits (dataset(**V**$_i$) with i =1, 2 or 3). The substantial volume of experimental data generated by these crystallizations assays has prompted the development of a database.[9b] We have previously described the '*subset-of-interest*' analysis method, noted SOI, for the comparison of two crystallization experiments enabling to rapidly isolate and analyze the unique crystallization conditions.[9b] Here we extend this approach for the comparison of three HTX datasets (**N**^{cum}, **G**$_1^{cum}$ and **V**$_i$). As previously outlined,[9b] the principle of the SOI analysis is to break down the by-plate representation and to represent each crystallization condition of a HTX experiment by a square (576 conditions). The results are now distributed depending on the presence of hit over the three datasets in eight distinct sub-sets: *zero hit* for any of the three datasets (noted **z**), *unique hits* for the native protein, for the protein with Tb-Xo4 or the protein with variant (three possibilities noted **n**, **g**$_1$ and **v**$_i$ respectively), *double hits* (i.e. common hits between two datasets: (i) native protein and protein with Tb-Xo4 (**ng**$_1$), (ii) native protein and protein with variant (**nv**$_i$) and (iii) protein with Tb-Xo4 and with variant (**g**$_1$**v**$_i$)) and a *triple hit* (common hits between the three datasets noted **ng**$_1$**v**$_i$). Collectively, these eight categories encompass the total of 576 experimental conditions. This SOI analysis is illustrated Figure in the case of HEWL with **V**$_1$ and

summarized in Table 1. In a similar manner, the results for all the other cases are compiled in Table 1.

Datasets	z	n	g_1	v_i	ng_1	nv_i	g_1v_i	ng_1v_i
HEWL-V_1	276 (220)	4 (22)	42 (48)	34 (29)	5 (23)	0 (4)	177 (164)	38 (66)
HEWL-V_2	306 (242)	4 (26)	183 (176)	4 (3)	22 (62)	0 (0)	36 (36)	21 (31)
HEWL-V_3	288 (226)	3 (24)	142 (108)	22 (23)	24 (54)	1 (2)	77 (104)	19 (35)
protK-V_1	344 (256)	14 (31)	42 (50)	103 (63)	18 (43)	6 (24)	21 (36)	28 (73)
protK-V_2	377 (270)	14 (44)	48 (58)	70 (49)	21 (54)	6 (11)	15 (28)	25 (62)
protK-V_3	441 (314)	20 (51)	58 (75)	6 (5)	25 (75)	0 (4)	5 (11)	21 (41)
TdThau-V_1	517 (400)	27 (57)	13 (37)	2 (17)	13 (41)	1 (4)	2 (7)	1 (13)
TdThau-V_2	519 (415)	28 (61)	14 (43)	0 (2)	7 (37)	0 (0)	1 (1)	7 (17)
TdThau-V_3	514 (390)	27 (58)	14 (42)	5 (27)	10 (37)	1 (3)	1 (2)	4 (17)

Table 1. Detailed SOI analysis obtained for each protein (HEWL, ProtK and TdThau) for the comparison between three datasets: dataset (N^{cum}), dataset (G_1^{cum}) and dataset(V_i).top number = score 6 (all).

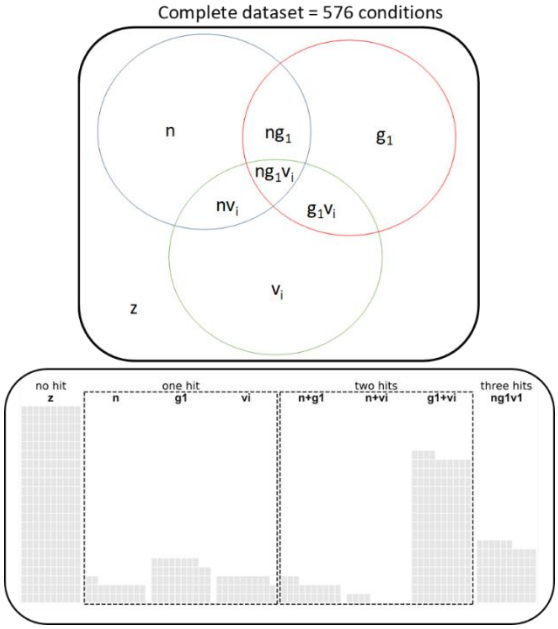


Figure 4. (top) Schematic representation of the comparison of three datasets: Dataset(N^{cum}) in blue, dataset(G_1^{cum}) in red and Dataset(V_i) in green defining the 8 sub-categories (z, **n**, g_1, v_i, ng_1, g_1v_i, nv_i and ng_1v_i). (bottom) Example of the SOI analysis ($v_i = v_1$) generated by our home-made database, enabling the comparison of cumulated HEWL native (N^{cum}), cumulated HEWL-Tb-Xo4 (G_1^{cum}) and HEWL-variant V_1 datasets.

For each variant, we can define three parameters specific to the protein used:

1°) **Productivity vs the native protein** P(V_i, protein): This parameter represents the number of additional hits induced by the variant compared to the native protein and is defined as P(V_i, protein) = v_i + g_1v_i. Note that all common hits corresponding to nv_i and ng_1v_i are excluded from this definition because in these cases, it is impossible to only ascribe the crystallization to the presence of the additive. The productivity allows us to compare the productivity of each complex including the original Tb-Xo4 (G_1^{cum}). By analogy, P(G_1^{cum}, protein) = g_1 + g_1v_i.

2°) **Complementarity vs the reference Tb-Xo4** (G_1) C(V_i, protein): This parameter represents the additional hits generated by the V_i variant that are not present either in the native (N^{cum}) or the Tb-Xo4 (G_1^{cum}) HTX experiment. It is defined as C(V_i, protein) = v_i and measures the complementarity of the variant compared to (G_1^{cum}).

3°) **Deleterious effect vs the reference G_1^{cum}** D(V_i, protein): It is worth mentioning that all additives present also, in some conditions, a negative effect on the crystallization defined by the number of native hits that disappear in the presence of the additive. This parameter is called **deleterious effect** and noted D(V_i, protein) = **n** + ng_1. By analogy, D(G_1^{cum}, protein) = **n** + nv_i). The values are compiled in Table 2. It is first worth noting that all variants exhibit productivity and generate additional hits when compared to the native but their effect varies significantly. Only V_1 presents a very high productivity for the three proteins, comparable and sometimes higher to that of the reference G_1^{cum} additive. The two other variants present a much more contrasted productivity, strongly dependent on the protein. As an example, P(V_2, protK) and P(V_3, TdThau) are significant but their productivity in the other cases is much lower. As example, P(V_2, TdThau) is rather small indicating that this variant has a negligible effect on the crystallization process of TdThau. V_2 and V_3 are therefore more selective of one protein compared to V_1 that is more universal (as G_1).

	G_1^{cum}	V_1	V_2	V_3
		Productivity vs **N**		
HEWL	219 (212)	211 (193)	40 (39)	99 (127)
protK	63 (86)	124 (99)	85 (77)	11 (16)
TdThau	15 (44)	4 (24)	1 (3)	6 (29)
		Complementarity vs G_1		
HEWL	-	34 (29)	4 (3)	22 (33)
protK	-	103 (63)	70 (49)	6 (5)
TdThau	-	2 (17)	0 (2)	5 (27)
		Deleterious effect vs G_1		
HEWL	4 (26)	9 (45)	26 (84)	27 (78)
protK	20 (55)	32 (74)	35 (98)	45 (126)
TdThau	28 (61)	40 (98)	35 (98)	37 (95)

Table 2. Productivity, complementarity vs G_1^{cum} and deleterious effect (score 6 (all)) for the variants V_i (i =1-3) and the three model protein (HEWL, ProtK, TdThau).

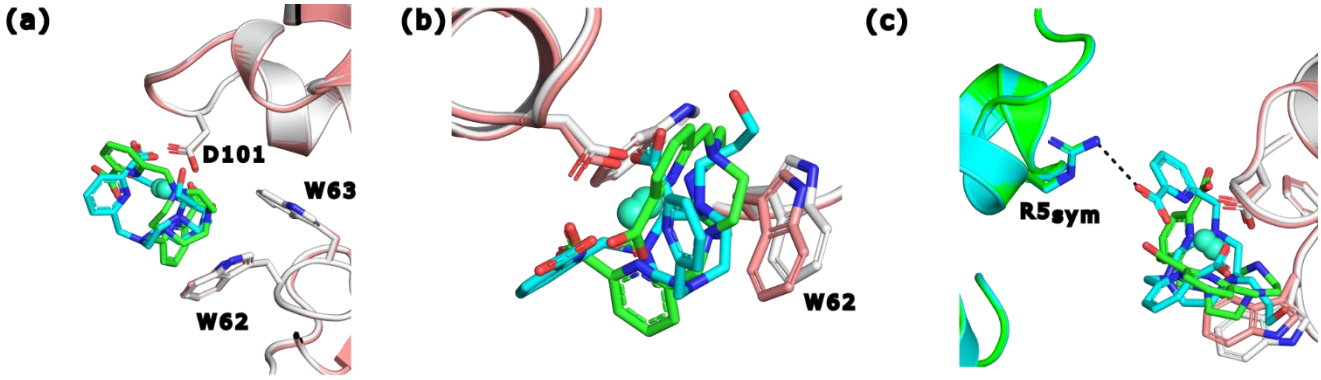


Figure 5. Binding site comparison between Tb-Xo4 (**G₁**) and the variant with propyl-3-ol pendant arm (**V₁**). (a) Superimposition of interaction sites in {**G₁**/HEWL} (in white) and in {**V₁**/HEWL} (in salmon). **G₁** and **V₁** variant molecules are colored in green and cyan, respectively. (b) Different view of (a) showing the change in W62 position. (c) Extra view of the superposition showing the interaction with R5 from a symmetrical molecule. The crystal-symmetry related molecules are colored similarly to their respective lanthanide complexes (green and cyan for {**G₁**/HEWL} and {**V₁**/HEWL}, respectively). The protein molecules are depicted in cartoon mode with the exception of interacting residues represented as sticks.

It is now interesting to compare the complementarity of **V₁** vs **G₁** that presents similar productivity. Importantly, for all proteins, **V₁** shows a significant amount of unique additional hits not present neither in the native nor in the **G₁^{cum}** datasets. The complementarity vs **G₁^{cum}** is particularly pronounced in the case of C(**V₁**, protK) and C(**V₁,** TdThau) where respectively 80% and 50% of the productivity is composed of additional hits compared to Tb-Xo4 (**G₁^{cum}**). It is therefore possible to conclude that **V₁** is the most promising variant since it combines the highest productivity and the best complementarity compared to the Tb-Xo4 additive. The other two variants seem to be more selective for a given protein. These heterogeneous results should be correlated to the overall charge of the complex, **V₁** being mono-cationic (as **G₁**) whereas **V₂** and **V₃** are dicationic and neutral respectively.

The deleterious effect of all variants is slightly higher than that of **G₁** and its origin is still under investigation. This also indicates that **G₁**, presents the highest number of common hits with the native protein as previously observed for the initial protein panel tested.[9a]

Finally, the comparison between the interaction of HEWL and the different variants vs. **G₁** was investigated using a combined XRD-molecular simulation approach. The binding pocket of {**G₁**, HEWL} has been previously described and revealed the formation of a dimer linked to the protein via the aspartate D101 with a calculated stabilization energy of -15.1 kcal.mol^{-1}.[9c] However, this description relied on a structure (PDB Id: 6F2I) obtained from a {**G₁**/HEWL} crystal in the presence of 100 mM Tb-Xo4.

The three structures of the {**V$_i$**/HEWL} adducts (i =1-3) have been solved along with the structure of {**G₁**/HEWL} at the same 10 mM concentration, to minimize bias in structure comparisons (PDB 8OWC for Tb-Xo4 and PDB = 8P2Q, 8PIW, 8POB for each **V$_i$** with i = 1-3 respectively, Table S3 and S4).

The structure of {**G₁**/HEWL} no longer indicated the presence of a dimer. Only one Tb-Xo4 molecule can be modeled bound to D101 with a small signal in the anomalous Fourier synthesis indicative of a potential second molecule (Figure S26). Moreover, the presence of residual electron density in the (Fo-Fc) Fourier map close to the modeled conformation of the Xo4 ligand may reveals ligand dynamics and the possible presence of at least a second Xo4 conformation. Thus, it is likely that the binding of a second Tb-Xo4 molecule through coordination of the second Tb^{3+} by one of the carboxylate moieties of the first Tb-Xo4 as observed in the dimer[9c] locks the conformation of the first molecule.

In the cases of both {**V₁**/HEWL} and {**V₂**/HEWL} crystal structures, a single binding site is detected from the anomalous Fourier map, while in {**V₃**/HEWL} crystal structure, the binding site presents an other small anomalous signal as observed in {**G₁**/HEWL} structure (Figure S27(a)). All three variants share similar binding modes with Tb-Xo4 including the direct metal coordination by D101 and the hydrophobic interaction between the Xo4 macrocycle and tryptophan W62 (Figure S27(a)). However, the addition of a pendant arm induces a change in the conformation of the Xo4 ligand. This is exemplified by the structural comparison of **V₁** with Tb-Xo4 which shows a substantial rotation relative to the initial Tb-Xo4 conformation (Figure 5(a)). While W63's position remains strictly similar, W62 adopts a slightly different position in the structure of {**V₁**/HEWL} thereby maintaining the hydrophobic interaction with the macrocycle (Figure 5(b)). Associated with the conformation adopted by **V₁**, a hydrogen bond is formed between a picolinate moiety and arginine R5 side chain from a symmetry-related HEWL molecule (Figure 5(c)). Since the R5 conformation is similar in {**G₁**/HEWL} structure and no interaction is observed with the Xo4 ligand, it is likely that the interaction with R5 in {**V₁**/HEWL} results from the final crystal packing and cannot be related to the nucleating properties of **V₁**. The same conformational modification is observed for **V₂** and **V₃** (Figure S27(a)) clearly indicating that the presence of a pendant arm modifies the crystallophore conformation at HEWL surface. However, the absence of a clear electron density for the pendant arm in all variants (Figure S27 (b)) indicates that the added moieties are mobile and do not participate in strong interactions with the protein molecules, at least within HEWL crystals.

To gain insights into the role of the pendant arm, we performed Molecular Dynamics (MD) simulations by considering the crystal packing. For this purpose, four HEWL molecules (forming a tetramer noted HEWL A-D) were taken into account from the crystallographic structures to reconstruct the binding pocket in the presence of **G₁**, **V₁**, **V₂** or **V₃** additives. Each variant was reconstructed within the different pockets and modeled with their dedicated force field parameters. Our analysis primarily focused on a Tb-Xo4 located at the interface of the four HEWL molecules. The {tetramer-complex} model is stable during the 200 ns MD regardless of the variant, allowing the characterization of the binding pocket (Figure).

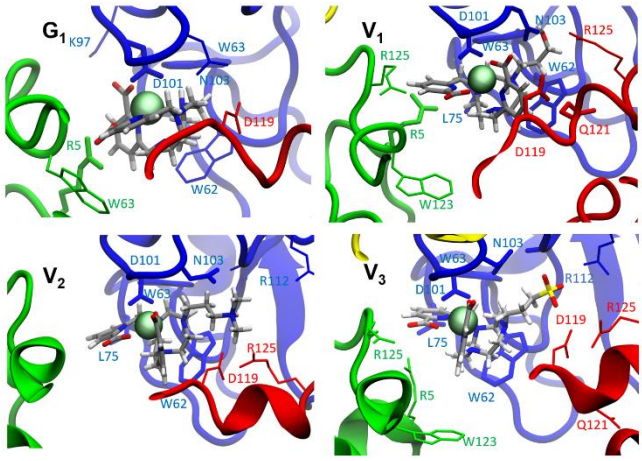


Figure 6. Binding pocket of **G₁** and the three variants **V₁**, **V₂** and **V₃** at the interface of the four HEWL (blue: HEWL A; red: HEWL B; green: HEWL C; yellow: HEWL D). Yttrium cation is represented by the green sphere, ligand and interacting residues (Occurrence rate > 0.5) are draw with ball and sticks representations.

First, in line with crystallographic structures, the crystallophore conformation in the binding cavity is conserved during the simulation for **G₁** and the three variants. We use the ProLIF tools [13] to characterize the contribution of the different residues to the binding pocket of each variant. Based on the residue-ligand distance, this approach allows us to discriminate the various interactions involved and their occurrence, with regard to the pendant arm (see Table S5 and Figure S28). The rotation of the complex subtly modifies the interaction with HEWL A: the interaction with $W62_A$ and $W63_A$ presents a stronger contribution with the variants than with **G₁**, and the interaction between $W62_A$ and the macrocycle is maintained during the 200 ns. Moreover, new contacts appear between the pendant arm and residues Asparagine $N106_A$, Alanine $A107_A$ and Arginine $R112_A$. The pendant arm also interacts with the C-term loop of a second lysozyme molecule (HEWL B) with respect to its terminal chemical group. Indeed, according to Table S5, the positively charged **V₂** pendant arm transiently interacts with $D119_B$ or is surrounded by water molecules, while the negatively charged sulfonate of **V₃** prefers arginines ($R125_B$ and also $R112_A$). The pendant arm of **V₁** presents a more versatile behavior and explores various conformations and interaction patterns, with $W63_A$, $N103_A$, Glutamine $Q121_B$ and $R125_B$. The N-term α-helix (R5) or C-term W123 and R125 of a third HEWL (HEWL C) can interact with one picolinate of the ligand, regardless of its nature. Indeed, in the **G₁** simulation, $R5_C$ residue, which is not interacting with the crystallophore in the crystallographic structure, rotates after few nanoseconds of simulation to form a salt bridge with the picolinate group, and then moves back and forth between the picolinate and the backbone of the C-terminal coil of HEWL C. On the contrary, the hydrogen bond between this arginine and the variants observed in the crystallographic structure is not maintained all along the simulations, where the structures are slightly relaxed compared with the crystal environment. *In silico* results corroborate the idea that this interaction is due to crystal packing. No direct interaction with the fourth lysozyme molecule is identified. Moreover, a first estimate of the binding energies can be obtained through MM-PBSA free energy post-analysis. In our simulations, **G₁** and **V₁** have the same interaction energy with the tetramer. This energy is increased by 20 kcal·mol^{-1} for **V₂** and **V₃** (see Table S6), suggesting a less efficient binding of these latter variants.

Our simulations show that the variants can be strongly bound to one lysozyme molecule through the carboxylate–metal interaction (such as Tb-Xo4 (**G₁**)), while their pendant arms are free to explore the surface of other protein molecules. The salt bridges formed between the ammonium or sulfonate moieties and acidic or basic residues respectively can be relatively strong whereas the alcohol group can be involved in weaker hydrogen bonds, as both H donor or acceptor. Consequently, the **V₁** variant takes more advantage of the flexibility of the pendant arm to explore the different protein surfaces and can create weak links between them like a molecular lasso.

Conclusion

In this article we scrutinized the impact of secondary modifications of the crystallophore structure, especially the introduction of various pendant arms at the free amino position of the TACN macrocycle (**V₁₋₃**,) on the overall protein crystallization process. Based on high-throughput crystallization experiments and dataset analysis, we estimated the productivity and the complementarity of the different variants versus the first crystallophore generation **G₁**. This methodology clearly evidenced that **V₁** featuring propyl-3-ol pendant arm exhibits the most promising variant in terms of productivity and complementarity vs **G₁** for the three proteins tested (hen-egg white lysozyme, proteinase K and thaumatin). To gain insights into the role of the pendant arm, three structures of the {**V**$_i$/HEWL} adducts (i =1-3) have been solved along with the structure of {**G₁**/HEWL} and Molecular Dynamics simulations considering the crystal packing were achieved. Here again, **V₁** variant has a special behavior taking more advantage of the flexibility of the propyl-3-ol arm to explore the different protein surfaces and can create weak versatile links with them. These findings underscore the significance of fine-tuning the crystallophore structure to optimize its performance in protein crystallization on the way to a universal crystallization additive.

Funding sources

This work was supported by the French Agence Nationale de la Recherche (ANR) through the Project Ln23 (ANR-13BS07-0007-01) and by the Région Auvergne-Rhône-Alpes through the project Xo4-2.0 (Pack Ambition Recherche 2017). Polyvalan also acknowledged the French ministry of higher education and research innovation competition i-Lab operated by BPI France for

support. Financial support was also provided by Instruct-ERIC (PID: 15980).

Acknowledgments

The authors acknowledge Soleil synchrotron for beamtime provision and thank Proxima1 beamline staff for their help. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble-Alpes graduate school CBHEUR-GS (ANR-17-EURE-0003).

Notes

The authors declare the following competing financial interest(s): OM, FR and EG declare a potential conflict of interest since they are co-founders of the Polyvalan company that commercializes the crystallophore. AR is hired by the Polyvalan company.

References

[1] A. McPherson, J. A. Gavira, *Acta Crystallogr. F Struct. Biol. Commun.* **2014**, *70*, 2-20.

[2] a) S. Khurshid, E. Saridakis, L. Govada, N. E. Chayen, *Nat. Protoc.* **2014**, *9*, 1621-1633; b) T. C. Terwilliger, D. Stuart, S. Yokoyama, *Ann. Rev. Biophys.* **2009**, *38*, 371-383.

[3] a) T. C. Terwilliger, D. Liebschner, T. I. Croll, C. J. Williams, A. J. McCoy, B. K. Poon, P. V. Afonine, R. D. Oeffner, J. S. Richardson, R. J. Read, P. D. Adams, *bioRxiv* **2022**, 2022.2011.2021.517405; b) C. Millan, E. Jimenez, A. Schuster, K. Diederichs, I. Uson, *Acta Crystallogr. D Struct. Biol.* **2020**, *76*, 209-220.

[4] J. Brown, T. S. Walter, L. Carter, N. G. A. Abrescia, A. R. Aricescu, T. D. Batuwangala, L. E. Bird, N. Brown, P. P. Chamberlain, S. J. Davis, E. Dubinina, J. Endicott, J. A. Fennelly, R. J. C. Gilbert, M. Harkiolaki, W. C. Hon, F. Kimberley, C. A. Love, E. J. Mancini, R. Manso-Sancho, C. E. Nichols, R. A. Robinson, G. C. Sutton, N. Schueller, M. C. Sleeman, G. B. Stewart-Jones, M. Vuong, J. Welburn, Z. Zhang, D. K. Stammers, R. J. Owens, E. Y. Jones, K. Harlos, D. I. Stuart, *J. Appl. Crystallogr.* **2003**, *36*, 315-318.

[5] a) A. McPherson, P. Shlichta, *Science* **1988**, *239*, 385; b) N. E. Chayen, E. Saridakis, R. El-Bahar, Y. Nemirovsky, *J. Mol. Biol.* **2001**, *312*, 591-595; c) A. D'Arcy, A. Mac Sweeney, A. Haber, *Acta Crystallogr. D Biol. Cryst.* **2003**, *59*, 1343-1346; d) E. Saridakis, S. Khurshid, L. Govada, Q. Phan, D. Hawkins, G. V. Crichlow, E. Lolis, S. M. Reddy, N. E. Chayen, *Proc. Nat. Ac. Sci.* **2011**, *108*, 11081.

[6] a) A. Bijelic, A. Rompel, *Coord. Chem. Rev.* **2015**, *299*, 22-38; b) J. Breibeck, A. Bijelic, A. Rompel, *Chem. Commun.* **2019**, *55*, 11519-11522; c) A. Mac Sweeney, A. Chambovey, M. Wicki, M. Muller, N. Artico, R. Lange, A. Bijelic, J. Breibeck, A. Rompel, *PLoS One* **2018**, *13*, e0199639; d) A. Bijelic, A. Rompel, *ChemTexts* **2018**, *4*, 10.

[7] a) R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power, P. B. Crowley, *Nat. Chem.* **2012**, *4*, 527-533; b) R. E. McGovern, S. C. Feifel, F. Lisdat, P. B. Crowley, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 6356-6359; c) F. Guagnini, P. M. Antonik, M. L. Rennie, P. O'Byrne, A. R. Khan, R. Pinalli, E. Dalcanale, P. B. Crowley, *Angew. Chem., Int. Ed.* **2018**, *57*, 7126-7130; d) S. Engilberge, M. L. Rennie, E. Dumont, P. B. Crowley, *ACS Nano* **2019**, *13*, 10343-10350; e) K. O. Ramberg, S. Engilberge, T. Skorek, P. B. Crowley, *J. Am. Chem. Soc.* **2021**, *143*, 1896-1907.

[8] a) G. Pompidor, A. D'Aleo, J. Vicat, L. Toupet, N. Giraud, R. Kahn, O. Maury, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 3388-3391; b) R. Talon, R. Kahn, M. A. Dura, O. Maury, F. M. Vellieux, B. Franzetti, E. Girard, *J. Synchrotron Radiat.* **2011**, *18*, 74-78.

[9] a) S. Engilberge, F. Riobe, S. Di Pietro, L. Lassalle, N. Coquelle, C. A. Arnaud, D. Pitrat, J. C. Mulatier, D. Madern, C. Breyton, O. Maury, E. Girard, *Chem. Sci.* **2017**, *8*, 5909-5917; b) T. Jiang, A. Roux, S. Engilberge, Z. Alsalman, S. Di Pietro, B. Franzetti, F. Riobé, O. Maury, E. Girard, *Cryst. Growth Des.* **2020**, *20*, 5322-5329; c) S. Engilberge, F. Riobe, T. Wagner, S. Di Pietro, C. Breyton, B. Franzetti, S. Shima, E. Girard, E. Dumont, O. Maury, *Chem. Eur. J.* **2018**, *24*, 9739-9746; d) I. Bernhardsgrütter, B. Vögeli, T. Wagner, D. M. Peter, N. S. Cortina, J. Kahnt, G. Bange, S. Engilberge, E. Girard, F. Riobé, O. Maury, S. Shima, J. Zarzycki, T. J. Erb, *Nat. Chem. Biol.* **2018**, *14*, 1127-1132; e) S. Engilberge, T. Wagner, G. Santoni, C. Breyton, S. Shima, B. Franzetti, F. Riobe, O. Maury, E. Girard, *J. Appl. Crystallogr.* **2019**, *52*, 722-731.

[10] M. L. Rennie, G. C. Fox, J. Perez, P. B. Crowley, *Angew. Chem., Int. Ed.* **2018**, *57*, 13764-13769.

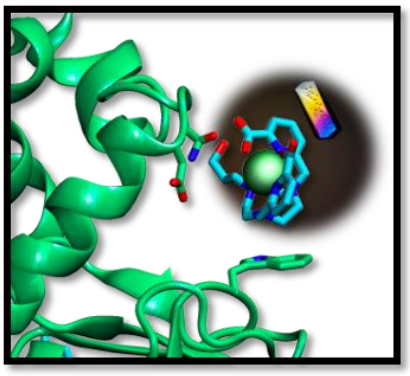
[11] A. Roux, R. Talon, Z. Alsalman, S. Engilberge, A. D'Aleo, S. Di Pietro, A. Robin, A. Bartocci, G. Pilet, E. Dumont, T. Wagner, S. Shima, F. Riobe, E. Girard, O. Maury, *Inorg. Chem.* **2021**, *60*, 15208-15214.

[12] N. Dimasi, D. Flot, F. Dupeux, J. A. Marquez, *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **2007**, *63*, 204-208.

[13] C. Bouysset, S. Fiorucci, *J. Cheminformatics* **2021**, *13*, 72.

Keywords: protein structure determination • crystallophore • nucleating effect • supramolecular chemistry • lanthanides • molecular dynamics

Entry for the Table of Contents



Crystallophores are lanthanide complexes known to favor protein crystallization. Herein, we explore the effect of the introduction of pendant arms at the free amino position of the TACN macrocycle on the nucleating properties, and protein crystal structures. Through high-throughput crystallization experiments and dataset analysis, we evaluated the effectiveness of these variants, in comparison to the first crystallophore generation, and highlighted the potential of the $\mathbf{V_1}$ variant, featuring a propyl-3-ol pendant arm.

Institute and/or researcher Twitter usernames: @LcEnsl