

Image Annotation and Database Mining to Create a Novel Screen for the Chemotype-Dependent Crystallization of HCV NS3 Protease

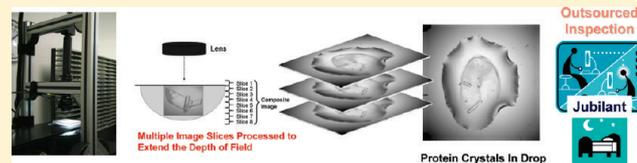
Published as part of the *Crystal Growth & Design* virtual special issue on the 13th International Conference on the Crystallization of Biological Macromolecules (ICCBM13)

Herbert E. Klei, Kevin Kish, Mark F. Russo, Steve J. Michalczyk, Matthew H. Cahn, Jeffrey Tredup, ChiehYing Chang, Javed Khan, and Eric T. Baldwin*

Bristol-Myers Squibb Company, Research and Development, Applied Biotechnology, Princeton, New Jersey, United States

S Supporting Information

ABSTRACT: An effective process for screening, imaging, and optimizing crystallization trials using a combination of external and internal hardware and software has been deployed. The combination of this infrastructure with a vast annotated crystallization database enables the creation of custom crystallization screening strategies. Because of the strong chemotype-dependent crystallization observed with HCV NS3 protease (HCVPr), this strategy was applied to a chemotype resistant to all



prior crystallization efforts. The crystallization database was mined for ingredients used to generate earlier HCVPr/inhibitor co-crystals. A random screen was created from the most prolific ingredients. A previously untested combination of proven ingredients was identified that led to a successful crystallization condition for the resistant chemotype.

INTRODUCTION

Over the past 10 or more years, government and privately funded structural genomics initiatives have introduced higher throughput methods to the X-ray structure determination process. These efforts have advanced the maturation of robotic/automated methods for protein purification, crystallization experimental setup, crystal image capture, data collection, and structure determination (i.e., the gene-to-structure process). The goal of these efforts has been to greatly extend the numbers of interesting structures publically available in several broad categories including, but not limited to, the expansion of the known protein fold-space, human and pathogen metabolic pathway enzymes, and other medically important target proteins.^{1–11} In parallel to these efforts, drug discovery organizations have applied similar principles to fine-tune high throughput methods to the unique needs of the pharmaceutical drug discovery environment.^{12–15} The first difference between these two approaches stems from the number of targets considered attractive for structural work. In the pharmaceutical environment, targets are almost always limited to those with some degree of drug target validation. The number of targets actively investigated at any given time is further limited to ensure that appropriate resources can be leveraged to drive the discovery efforts quickly to the clinic. Thus, target selection can be influenced, but rarely dominated, by structural considerations. Furthermore, stakeholders are reluctant to accept alternative structures to the specific target that is the focus of the discovery effort. Hence, these limitations must be reflected in the prioritization of

structural projects. The second major difference from the genomics initiatives is the need to provide structures with key ligands in order to aid in the atomic understanding of the structure–activity relationships (SAR). For any given project, it is common for hundreds of ligands to be studied over the course of the drug discovery process. Therefore, methods need to be developed that support the successful and timely crystallization of proteins with diverse compounds. The reduction of structure determination of protein/ligand complexes to routine practice is critical to the structure-based drug design process and a central focus of structural efforts in the pharmaceutical industry.

While the need for a robust gene-to-structure process is self-evident, the difficulty in developing a routine process for the crystallization of protein/ligand complexes may not be apparent. The recent success of a number of biotech companies and the subsequent adoption of fragment-based methods in pharmaceutical companies may suggest that achieving routine crystallization is straightforward.^{16–20} In fragment-amenable crystallization systems, compounds can be soaked into existing apo crystals and structures can be obtained of small ($M_w < 400$ Da), diverse compounds bound to the target of interest. Structures can be determined for very weakly binding compounds with a limited investment of effort. In our experience, this kind of structural work is the exception to the norm. More often, the identification

Received: October 13, 2010

Revised: February 1, 2011

Published: March 21, 2011

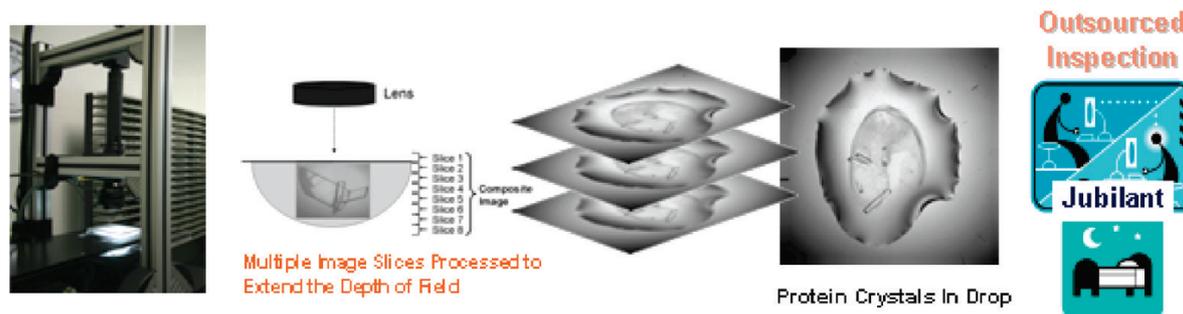


Figure 1. The crystal imaging through inspection/annotation process is outlined. The digital microscope (far left image) is suspended on a vibration dampening platform. The crystallization tray is illuminated from below during the imaging process. Behind the camera are nests for tray storage. Eight to ten images of each drop are taken in successive focal planes. The series of images is computationally combined to produce a single extended depth-of-field image. The combined image is sent to scientists at Jubilant Biosys, Bangalore, India, where each image is inspected and scored. U.S.-based researchers receive scoring updates before each business day.

of a crystallization condition that yields apo crystals, or even produces a successful co-crystallization complex, does not automatically produce a tight set of crystallization conditions that effectively leads to routine success. In a survey of three kinase projects, we observe an average success rate (structures delivered/compounds attempted) of about 35%, 40%, and 90% [kinase A: 10 chemotypes; success rate range 0–100%, average 35%; kinase B: 5 chemotypes success rate range 0–100%, average 40%; kinase C: 4 chemotypes success rate range 67–100%, average 90%; ChiehYing Chang, unpublished]. We and others (A. Hassell, personal communication) have called this phenomenon “chemotype-dependent crystallization.” Some compounds work well in crystallization while other compounds, which can be very similar, are much more resistant to co-crystallization with the target of interest. Repetitive broad screening, expansion of the set of compounds attempted, exploration of alternative compound co-solvents, and changing the protein construct are employed to move a project toward the goal of a routine process.

In this report, we summarize our implementation of a high-throughput crystallization infrastructure and the utilization of that infrastructure to address the chemotype-dependent crystallization of HCV NS3 protease (HCVPr). HCVPr has an average success rate of about 40% with prior chemotypes. However, an important chemotype failed to produce leads of any sort when more than 10 closely related compounds were attempted in co-crystallization across a standard set of broad screens by three different experimenters. We mined one year of annotated HCVPr crystallization data from successful experiments with the prior chemotypes and used this knowledge to create a new 96-well random screen that allowed us to determine the structure of the resistant chemotype.

■ EXPERIMENTAL SECTION

Overview of Crystallization Infrastructure. With the exception of the automated imager, the crystallization infrastructure described here consists of fixed stations of commercially available robotic components. Crystal drop setting is accomplished by an Innovadyne Screenmaker 96 + 8 in which drops ranging from 300 to 500 nL of protein + 300–500 nL well mixture are combined on Innovaplate SD2 plates (in either 1 or 2 drop mode). Additionally, the Innovadyne can dispense onto a Neuro Probe crystallization plate by using an adapter stand to hold the cover sheet. Well solutions are transferred from 96-well deep well blocks to the reservoir either by the Innovadyne directly or more typically manually with a RAININ Liquidator 96 channel pipet. Custom 96-well screens are formulated using a Tecan Genesis Freedom 200

fitted with a POSID (for vial barcode scanning), Thermo Seal-It 100 plate sealer, robotic arm for plate movement and a liquid handling arm. The Tecan is driven by RockMaker 2.0 software²¹ which allows the effective design of custom crystallization screens from our library of >100 of prebarcoded, deck-ready solutions (Emerald BioStructures, 40 mL vials with septum). Screens are dispensed directly in 96-well blocks or in the appropriate experimental tray type.

ASPECT Crystallization Database. An Oracle database which links crystal drop formulation information to the corresponding images and annotations was developed. This database includes the configuration of each experimental tray — the type of tray, number and formulation of protein drops, and the screening conditions in each well. Since RockMaker 2.0 software is used to design the crystallization screens, an XML-based connection was created between RockMaker and ASPECT that automatically uploads this information from RockMaker to the ASPECT database when experiments are marked as “dispensed” within RockMaker. The database includes the extended depth of field images of each drop after post processing and all time points. Annotations (scores) associated with each image are stored in the database.

Task Scheduling Software. A custom software application was written to allow users to instruct a robotic arm to feed crystallization trays from environmentally controlled storage to an automated imager, and return it when imaging is complete. The imager captures and stores images of each droplet in the experimental tray. Trays can be setup for imaging with fixed or custom schedules which can be modified as needed.

After placing experimental trays in an input nest outside of the storage system, the crystallographer submits a request for the robot to check for new trays to be moved into system storage. The robot senses the presence of a tray in an input nest, fetches it, reads its bar code, and transports it to an open storage nest. The system tracks the location of all trays in a local relational database using tray bar code as the primary identifier. A user can also request that one or more trays be scheduled for imaging at a specific time with a given set of imaging parameters. The system continuously looks for imaging tasks that are due to be performed. When ready, the robot moves a tray due for imaging to a nest in the imager and the system controller instructs the imager to begin collecting images of all drops in the experimental tray. Images are automatically saved, processed, and transferred to the database.

Operation of the robot transport, automated imager and other devices and sensors are coordinated through a custom system controller program which was designed using a Petri net model.^{22,23} Unique real-time control features of the integrated system are achieved by building an executable version of the Petri net model that is capable of controlling hardware.²⁴

Imaging Hardware. The automated imager is a custom instrument designed around a remotely controlled digital microscope. Software driving the microscope automatically finds the location of drops in

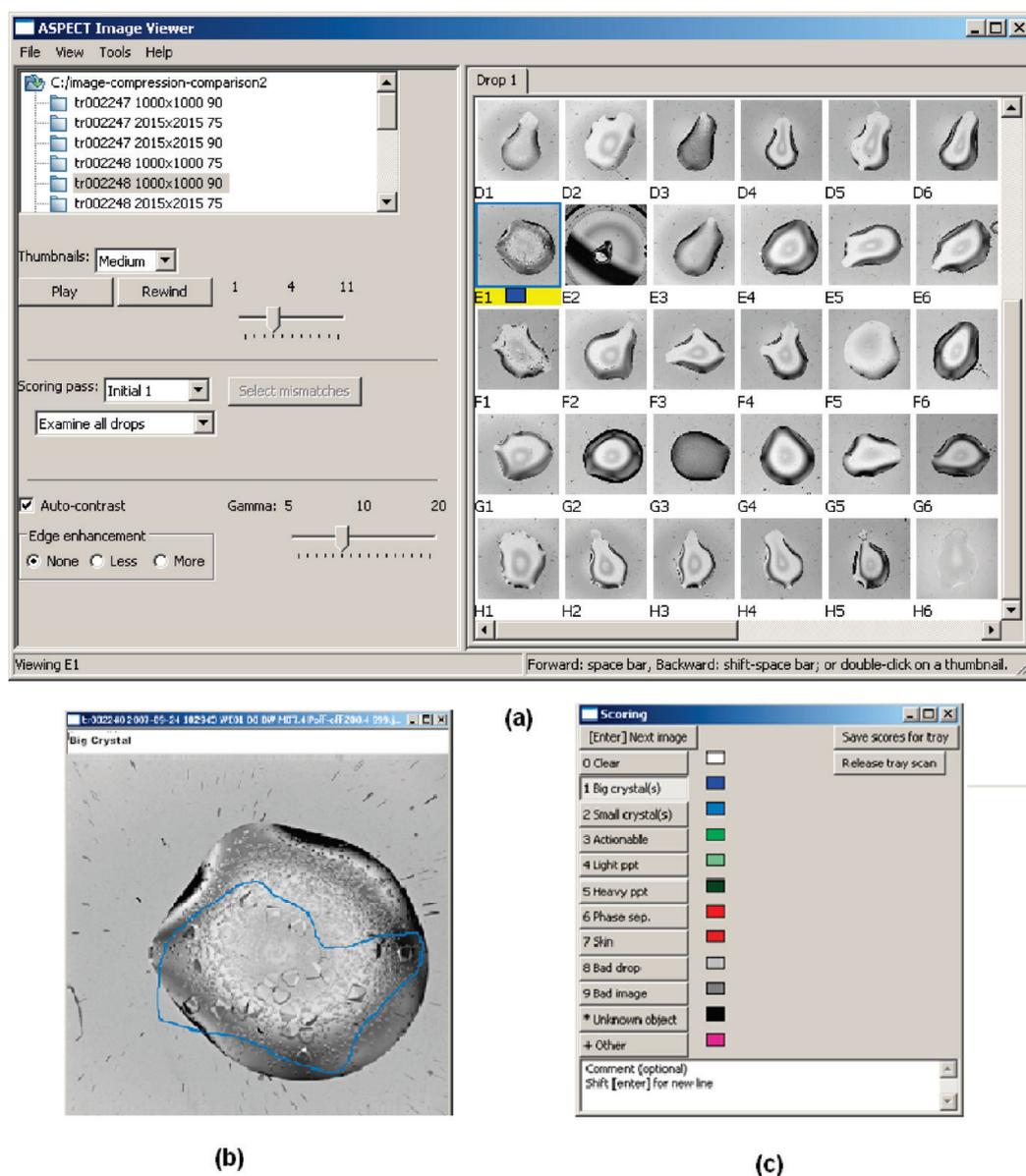


Figure 2. (a) The ASPECT image viewer presents thumbnail images of each drop from the selected crystallization tray. The image of E1 has a blue frame around the image and a blue box under the image to indicate that it was scored as a “big crystal”. (b) In this example of an annotated drop image the score “Big Crystals” is captured in the banner across the top of the image. The scorer used the mouse to indicate with a blue line what he was classifying as a big crystal. These annotations were all stored in the database upon scoring and are presented in ASPECT viewer when the image is recalled. (c) Scoring GUI from the ASPECT viewer that is used by internal and partner scientists to assign scores and provide additional written annotation for each drop.

each well of an experimental tray, zooms to fill the field of view with the drop, and then collects a stack of images with focal planes at successive depths that pass through the droplet. The imager is designed to accept a variety of sitting and hanging drop crystallization tray types. A LED dome light minimizes refraction from the drop edge during image acquisition, and an aperture is inserted into the light path to accentuate the drop edge when centering the drop. The imager uses a large format (2048×2048 pixel surface with each pixel $7.4 \mu\text{m}$ square) split CCD camera to minimize exposure time and to increase frame rate. An exposure time of less than 3 ms insures that vibration induced movement of the drop is frozen at high magnification.

In a postprocessing step, the images collected for one drop are aligned and then combined into a composite image with all detail in focus using image stacking software. The composite image is stored in the database

and made available to crystallographers for off-line inspection and annotation (Figure 1).

Image Stacking Software. The depth-of-field of an optical system is defined as the range of distance from the lens within which the subject is considered to be in-focus. As the magnification of an optical system increases, the depth-of-field necessarily decreases. The level of magnification required to capture details of a typical protein crystallization droplet is sufficiently large to make it impossible to capture a single in-focus image that spans the complete droplet depth.

This problem is solved using special software referred to as an “extended depth-of-field”, “image stacking”, or just a “stacking” software application. Stacking software works by assembling a single in-focus image from a stack of images captured at successive focal planes down through the subject’s depth. The depth-of-field of each image overlaps

with that of the image above it and below it in the image stack. Stacking software assembles a combined image by stitching together the parts of each image in the stack considered to be the most in-focus. We used an open source image stacker called CombineZ.²⁵

The ASPECT Image Viewer. A custom application written in Python and the wxPython user interface toolkit was designed to enable crystal drop image inspection. The desktop software presents the user with a grid of thumbnail images for each tray (Figure 2). The user can view the full image for selected drops, or proceed through the images for an entire tray. The user can also navigate forward and backward through the different time points for each drop in order to view the changes in the drop over time. Each image is annotated with the crystallization conditions for the drop. The user can assign one or more scores to each drop: “clear”, “small crystal(s)”, “big crystals(s)”, etc. The user may also draw lines on the image using the mouse to highlight areas or features of interest. The scores and written annotations are saved in the ASPECT database for later viewing and data mining (Figure 2b,c). The viewer includes a feature that allows the user to load crystallization conditions into the RockMaker application by dragging crystallization conditions from the viewer onto the open RockMaker application. This feature allows users to collect “hit conditions” and to rapidly design follow-up screens around those conditions.

Crystal Score Definitions. While some classifications are mutually exclusive (e.g., clear and small crystal), most classifications can be applied to the same drop (e.g., big crystal, light precipitate, skin). Examples of drop classifications are shown in Table 1.

CLEAR — Properly prepared and imaged but featureless drop.

PHASE SEPARATION — Visible separation of crystallization recipe into multiple phases (e.g., oil droplets in aqueous solution).

SKIN — Surface skin on drop formed over time. Often found on older drops. Usually indicated by rippled surface generated as the congealed material contracts to accommodate dehydrated drop.

BIG CRYSTAL — At least one crystal with one dimension of 0.05 mm or larger. We typically only score big crystal when large and small crystals are both present.

SMALL CRYSTAL — No crystal with one or more dimension of 0.05 mm or larger.

BAD DROP — Problematic drop usually due to some issue with the crystallization robotics (e.g., no drop, multiple small drops, drop placed on edge of coverslip).

BAD IMAGE — Problematic image (e.g., software failed to center the drop in field of view, out of focus) of properly prepared drop.

ACTIONABLE PRECIPITATE — Structured precipitate with some characteristics of small crystals (e.g., shiny, generates reflected glints of light when observed under microscope, some appearance of order or granularity) but no clear edges as with single crystals. This classification arose because the conditions of drops with these characteristics could often be optimized to generate crystals. It was intended to differentiate between drops with some promise and those with amorphous (usually brown) precipitate.

LIGHT PRECIPITATE — Amorphous precipitate. Often brown. Not heavy enough to obscure vision through drop.

HEAVY PRECIPITATE — Similar to light precipitate just more of it (enough to hinder observation of other features in drop). Can obscure crystals — especially small ones.

UNKNOWN OBJECT — Unidentified feature not crystal or precipitate (e.g., not sure if object is shard of glass or plastic from labware).

OTHER — Some feature or noteworthy observation (e.g., dust fiber present).

Image Scoring by Offshore Partner. As images are acquired by the imaging system and stored in the ASPECT database, the extended depth-of-field images are also transmitted to an off-shore partner for scoring. Transmission is accomplished with file transfer protocol (FTP) through a virtual private network (VPN) over the Internet. The partner

scores the images by viewing them in the same viewer application that is used in-house, but in a mode that reads the transmitted images locally from disk, rather than from the ASPECT database (Table 1). Two scorers each conduct an initial scoring pass, and an expert gives a final score to any images on which the initial scorers disagree. The viewer stores these scores locally, and the scores are transmitted back and stored in the ASPECT database. These scores annotate the images in the viewer, and the U.S.-based user can quickly view which conditions have produced crystals or other outcomes. Crystallizers also receive email notification when crystals are observed in their trays.

Data Mining of the Aspect Database. A series of SQL database queries were written to extract the crystallization conditions from all 2008 HCVPr trials that were annotated as “actionable precipitate” or “small/large crystal”. Each ingredient was classified as buffer, precipitant, or salt, and the frequency with which it was observed was tallied (Tables 2–4). Since this query polled all HCVPr crystallization trials, including a 3×96 custom screen that was the source of many crystals that were converted to structures of protein/ligand complexes, a second query was run that excluded custom screens (i.e., included only commercial screens). The rationale for this exclusion was that the comprehensive frequency distribution was too heavily weighted toward ingredients used in proven crystallization conditions. Furthermore, the purpose behind doing a random screen was to identify previously unsampled ingredient combinations. As expected, the frequency distribution without the custom screens was similar, but not identical, to the comprehensive distribution (Table 5, column 5). Only the ingredients that were selected as components in the new random screen are shown in Table 5.

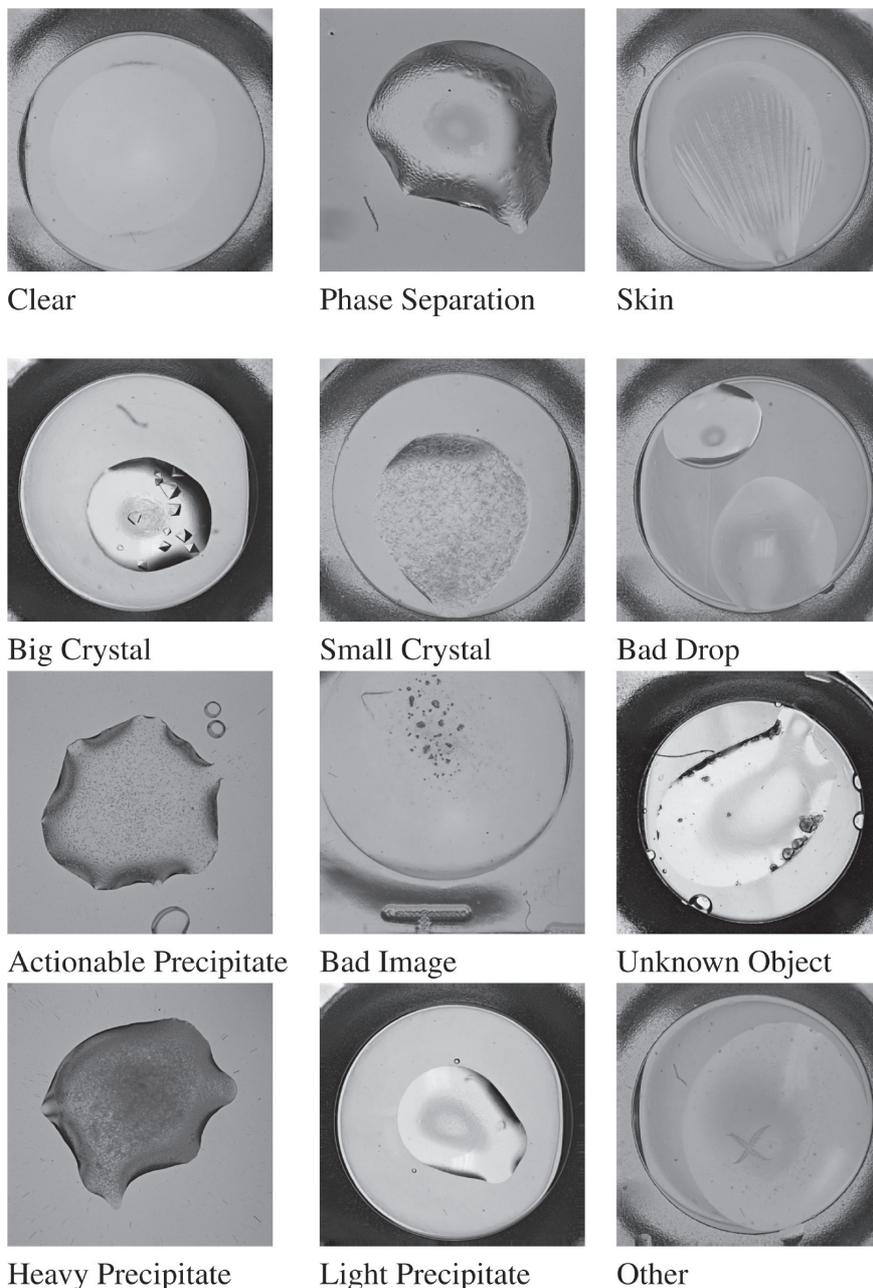
Another set of queries was used to extract pairwise ingredient combinations from these same data to assess the uniqueness of the new random screen created. From these data the PEG3350/phosphate pair was found as a hit seven times in the commercial screening data, while the PCB/phosphate and PCB/PEG3350 pairs were not previously associated with crystals or actionable precipitate.

Creation of the Random Screen with RockMaker 2.0. A random screen was generated using the RockMaker random screen interface, with the ingredients listed in Table 5. The software enables a user to group ingredients by type (buffer, precipitants, and salts), and each ingredient within a group was assigned a concentration range and pH. Each ingredient was assigned weights which the software would use to calculate the probability that each ingredient would be selected during the generation of the random screen. The weight for each ingredient was determined from the observed frequency of the ingredient found from data mining (see Table 5, column 5). The random selection process was forced to always choose one buffer, one salt, and one precipitant for all of the 96 experiments. The concentration and pH for ingredients were randomly selected from the available ranges during the screen creation. Once the parameters were set within the RockMaker, a random screen was generated upon saving. The required robot-ready solutions were placed on the Tecan deck, and RockMaker was used to drive the Tecan to create the screen in a deep well block.

All crystallization solutions were prepared by Emerald BioStructures and used the highest grade of each available reagent.

Crystallization of HCVPr with the Resistant Chemotype. The HCVPr construct is a fusion of the C-terminal 11 residues of the NS4a co-factor with NS3 protease (see full amino-acid sequence below). The residues are numbered to maintain the native designations for HCVPr. Non-native residues at the N-terminus, MKKK, the 11-residue NS4a fragment, GSVVIVGRINL, and the four-residue linker between NS4a and NS3, SGDT precede the HCVPr sequence. The protease domain sequence is the same as sequence 18 noted in Wittekind et al.^{26,27} with a C159S change to further improve crystallization behavior. The expressed sequence is extended at the C-terminus by a 22-residue read-through product, AIRAPSTSLR PHSSTTTTTT EI.

Table 1. Drop Classification Used for Scoring Were Provided to the Partner along with Several Examples of Drops Scored by the Authors^a



^a Single examples are shown in the table for illustration.

MKKKGSVVIV GRINLSGDTA YAQQTRGEEG CQETSQTGRD
KNQVEGEVQI VSTATQTFLA PCTCGSSDLY TSINGVLWTV
YHGAGTRTIA SPKGPVTQMY TNVDKDLVGW QAPQGSRLT
LVTRHADVIP VRRRGDSRGS LLSRPISYL KGSSGGPLLC
PAGHAVGIFR AAVCTRGVAK AVDFIPVESL ETTMRSPAIR
APSTSLRPHS STTTTTTEI

HCVP_r was cloned into a pET29a vector and expressed in BL21 (DE3) cells; the cells were grown at 37 °C after inoculation with 10 mL overnight liquid stocks into 1 L shaker flasks of M9 minimal media supplemented with 0.5% (w/v) Bacto-Casamino acids, 0.5% w/v glucose, 22.4 mM Na₂HPO₄, 17.2 mM KH₂PO₄, 8.6 mM NaCl, 0.74 μM vitamin B12, 3 μM

thiamine, 100 μM CaCl₂, 2 mM MgSO₄, and trace minerals (40 μM FeCl₃, 1.4 μM ZnCl₂, 2.9 μM CoCl₂, 3.2 μM CuSO₄, 3.2 μM H₃BO₃, 3.0 μM MnSO₄, 2.9 μM Na₂MoO₄). Just prior to induction 30 mM ZnCl₂ was added to the media (final concentration). Selection was maintained by 25 μg/L kanamycin. The cells were induced at OD₆₀₀ ~ 0.7 with 0.5 mM IPTG followed by culture for 20 h at 20 °C. Harvested cells were sonicated in a lysis buffer containing 25 mM MES (pH 6.5), 200 mM NaCl, 5% v/v glycerol, and 5 mM DTT. In addition, 1 μg/mL pepstatin, 0.2 mg/mL lysozyme, and 25 units/mL benzonase were added. The resulting cell lysate was clarified by centrifugation at 30 000 rpm (Ti45 rotor) for 30 min at 4 °C. The supernatant was extracted by cation exchange (40 mL SP sepharose fast flow column) with an AKTA Explorer-100 and eluted with a

Table 2. Buffer Ingredients from All of the HCVPr Crystallization Conditions in the ASPECT Database for Which a Score of Crystal or Actionable Precipitate Was Observed^a

buffer	low pH	high pH	frequency
Bis-Tris propane	7	7	1
CHES	9.5	9.5	1
sodium acetate Cl-free	4.6	4.6	1
sodium acetate anhydrous	4.6	4.6	1
sodium acetate trihydrate	4.6	4.6	1
sodium citrate dihydrate	5.6	5.6	1
trisodium citrate dihydrate	5.6	5.6	1
citric acid	4	5	2
SPG	5	7	3
sodium cacodylate	6.5	6.5	3
MMT	4	9	4
MIB	4	8	5
HEPES	7	7.5	6
PCB JCSG	4	9	7
Tris	8	8.5	7
MES	5.6	6.5	28
Bis-Tris	5.5	8.5	57

^a The pH range observed is indicated by the low and high pH (central columns), and the frequency of the ingredient is recorded in the right column.

Table 3. Precipitant Ingredients from All of the HCVPr Crystallization Conditions in the ASPECT Database for Which a Score of Crystal or Actionable Precipitate Was Observed^a

precipitant	concentration	frequency
2-methyl-2,4-pentanediol	40%	1
PEG 200	50% w/v	1
PEG 400	10% w/v	1
sodium malonate	0.5 M	1
ethanol	15%	1
ethylene glycol	50%	1
isopropanol	20%	1
Jeffamine ED-2001	0.5% w/v	1
sodium chloride	4.3 M	1
PEG 10000	17% w/v	2
PEG 5000 monomethyl ether	25% w/v	2
ammonium sulfate	2 M	4
glycerol anhydrous	20%	4
PEG 4000	27% w/v	8
PEG 2000 monomethyl ether	30% w/v	9
PEG 8000	20% w/v	15
PEG 6000	20% w/v	16
PEG 1500	25% w/v	19
ammonium sulfate	60%	52
PEG 3350	30% w/v	94

^a The concentration observed is in the central column, and the frequency of the ingredient is recorded in the right column.

200–800 mM NaCl gradient. The HCVPr fractions were concentrated by an Amicon 10 K/15 mL concentrator unit and then further purified and buffer exchanged by size exclusion chromatography (HiLoad 16/60

Table 4. Salt Ingredients from All of the HCVPr Crystallization Conditions in the ASPECT Database for Which a Score of Crystal or Actionable Precipitate Was Observed^a

salt	concentration (M)	frequency
calcium acetate hydrate	0.16	1
lithium nitrate	0.5	1
lithium sulfate monohydrate	0.5	1
magnesium sulfate heptahydrate	0.01	1
potassium acetate	0.4	1
potassium chloride	0.2	1
potassium formate	0.2	1
sodium fluoride	0.2	1
tacsimate	35	1
tripotassium citrate monohydrate	0.2	1
ammonium formate	1	2
ammonium sulfate	3.15	2
magnesium chloride	0.05	2
potassium bromide	0.5	2
DL-malic acid	2.1	3
lithium chloride anhydrous	0.2	3
sodium thiocyanate	0.2	3
sodium bromide	0.2	4
zinc chloride	0.01	4
calcium chloride dihydrate	0.2	4
sodium iodide	0.2	4
sodium nitrate	0.6	5
ammonium chloride	3.5	5
sodium acetate anhydrous	0.2	5
ammonium acetate	0.8	6
potassium sodium tartrate tetrahydrate	0.2	6
sodium chloride	0.2	6
sodium formate	0.8	6
trisodium citrate dihydrate	1.6	6
sodium malonate	2.4	7
sodium/potassium phosphate	0.2	7
sodium sulfate decahydrate	0.2	8
potassium thiocyanate	0.2	13
potassium dihydrogen phosphate	0.04	14
magnesium chloride hexahydrate	0.2	32

^a The concentration observed is in the central column, and the frequency of the ingredient is recorded in the right column.

Superdex 75) in a buffer containing 25 mM MES (pH 6.5), 500 mM NaCl, 5% v/v glycerol, and 5 mM DTT. The average yield was about 70 mg from 1 L of culture. The protein is monomeric by dynamic light scattering and had the expected molecular weight by mass spectrometry.

The HCVPr protein was concentrated to 16.4 mg/mL in 25 mM MES (pH 6.5), 500 mM NaCl, 5% v/v glycerol with fresh 5 mM DTT as above for crystallization experiments. A stock of 50 mM (dissolved in 100% DMSO) of the resistant chemotype was added to the protein to a final concentration of 1.4 mM. After gentle mixing, the sample was allowed to stand at room temperature for 1 h and was subsequently clarified by tabletop centrifugation at 4 °C. The protein/ligand complex was crystallized using the 96-well random screen using the infrastructure described above with 0.5 μ L + 0.5 μ L drops on Innovadyne SD2 trays. Large crystals were obtained from a solution containing 100 mM PCB²⁸ pH 7.33; 21.7% PEG3350; and 66.7 mM Na phosphate pH 8.8. The crystals were cryopreserved using a quick dip of loop mounted

Table 5. Design Parameters for the Random Screen Formulated from the Most Prolific Ingredients Used to Grow Earlier Co-Crystals^a

ingredient	name	[min]	[max]	count/(total)	probability
buffer	Bis-Tris	100 mM	100 mM	64/102	0.63
	MES	50	100	14/102	0.14
	Tris	100	100	10/102	0.10
	Na-HEPES	100	100	7/102	0.07
	PCB	100	100	7/102	0.07
precipitant	PEG 3350	20% w/v	25% w/v	89/142	0.63
	PEG 1500	25	25	19/142	0.13
	PEG 6000	20	20	17/142	0.12
	MME 2000	20	30	9/142	0.06
	PEG 4000	18	27	8/142	0.06
salt	potassium thiocyanate	0.1 M	0.2 M	12/64	0.19
	ammonium acetate	0.1	0.4	8/64	0.13
	sodium chloride	0.2	0.2	8/64	0.13
	sodium sulfate	0.2	0.2	8/64	0.13
	sodium malonate	0.2	1.2	7/64	0.11
	sodium nitrate	0.2	0.6	7/64	0.11
	sodium phosphate	0.066	0.066	7/64	0.11
	magnesium chloride	0.2	0.2	7/64	0.11

^a Each drop was required to have one, and only one, buffer, precipitant, and salt. Within each class, the likelihood that any given ingredient was selected was weighted by the probability listed in the last column (i.e., the fraction of time that it was found in successful crystallizations). Bis-Tris pH range 5.5–7.5; MES pH range 5.6–6.0; Tris pH range 8.0–8.5; Na-HEPES pH range 7.0–7.5; and PCB pH range 4.0–9.0.

crystals into a solution of 80% well mix and 20% glycerol (w/v). Data were collected at NSLS beamline X29 on an ADSC Q315 CCD detector with 1.0 Å X-ray radiation. The data were processed using HKL2000 and the structure determined using CCP4 software.

RESULTS AND DISCUSSION

Over the past several years, we have determined the X-ray structure of HCVPr in complex with a large number of diverse ligands. While we were successful obtaining many co-crystal structures with custom screens based on past successes with this target, the failure rate of these screens was higher than our average derived from experience. Hence, we often had to return to broad screening to find hits. Even after broad screening, only about 40% of the compounds that were tried resulted in a co-crystal structure. This chemotype-dependent crystallization behavior is quite common but is slightly more challenging for HCVPr. We also noted that we obtained over 30 different crystal forms (i.e., unique combination of space group and unit cell parameters). No crystal form occurred more than three times. The compounds bind on a flat and open protein surface that we rationalized was impacting the crystal contacts formed. In fact, in most of the structures the compound was involved in crystal contacts. Thus, the strong chemotype dependency appeared to be driven by the modulation of the protein surface by the compounds which led to crystallization challenges and forced the protein to adopt a variety of packing choices when forming crystals.

One important chemotype completely failed to yield crystals even after resorting to broad screening. No crystal hits, or even poor leads, were obtained with 10 closely related compounds. Multiple additional approaches were attempted. The protein was formulated in various ways, co-solvents changed, and all manner of seeding employed. In addition, three different crystallization experts prepared complexes and independently attempted to

obtain crystals. No crystal hits or poor leads were obtained. Rather than completely abandon these compounds, an additional experiment was attempted. During the last several years of the HCVPr project, a well-annotated set of crystallization data had been accumulated within the ASPECT database. For all of the crystallization experiments, images of the drops had been recorded and these images were annotated by experts (Table 1, Figure 2). A SQL query of the data could be used to extract the most successful crystallization conditions, and that information could be used to create a new screen that might sample known favorable crystallization space in new ways.

The HCVPr crystallization data was mined for the year 2008, and all of the crystallization conditions associated with big crystals, small crystals, or “actionable precipitates” were extracted. The ingredients were binned into three categories: buffer, salt, precipitate. The number of observations of each ingredient were recorded and assembled into Tables 2–4. The ranked list shows the variables for each category most often associated with successful crystallization. Bis-Tris, MgCl₂, and PEG3350 were most highly correlated with crystals. The query was refined to exclude data from a routinely used custom screen, but similar results were obtained. This second query was the basis for the selection of ingredients for the new screen. The choice of buffers, salts, and precipitants was based on the top performers and limited to produce a reasonable sampling in 96 experiments (Table 5). These data were used to design a 96-well random screen using RockMaker 2.0. The screen was created on the Tecan using our standard robot-ready library of stock solutions and dispensed into a 96-well block. The crystallization experiment was prepared on the Innovadyne drop setter using an Innovadyne SD-2 tray and 0.5 + 0.5 μL drops of HCVPr in complex with the ligand at 1.4 mM. Crystals were obtained in only one drop after several days (Figure 3) and were large enough to cryopreserve. The data were collected at NSLS and

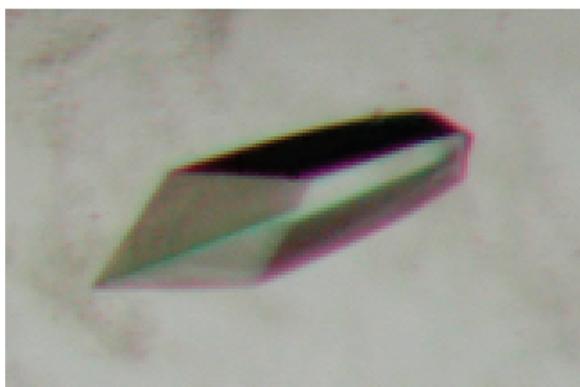


Figure 3. Crystals of HCVPr in complex with the chemotype that resisted co-crystallization were grown in a random screen derived from the ingredients most frequently associated with successful crystallization of the standard chemotypes. This crystal was grown in 100 mM PCB pH 7.33; 21.7% PEG3350; and 66.7 mM Na phosphate pH 8.8.

were 99.3% complete to 2.0 Å and 8-fold redundant with an $I/\sigma(I) > 3$ in the outermost shell. The space group was $P6_5$ with unit cell parameters $a = b = 91.8$ Å, $c = 36.0$ Å, $\alpha = \beta = 90^\circ$, and $\gamma = 120^\circ$. The structure was solved by molecular replacement and showed clear density for the ligand. The asymmetric unit contained one HCVPr–inhibitor complex. This was a unique crystal form for our collection of HCVPr structures.

A subsequent analysis of the crystallization conditions (PEG3350, PCB buffer, and phosphate) showed that this trio had not been tried previously. The binary combination of PEG3350 and phosphate was present in our commercial screens and had been associated with crystals seven times previously. The binary combinations of PEG3350/PCB buffer or phosphate/PCB buffer were not present in the commercial screening set and had not been previously tested in crystallization trials. These data confirmed our hope that new combinations of ingredients most commonly associated with the successful crystallization of complexes of HCVPr could generate crystallization conditions that might facilitate our efforts to mitigate the highly chemotype-dependent crystallization behavior of HCVPr.

CONCLUSION

Over the past several years, an infrastructure appropriate for high-throughput crystallography in the pharmaceutical setting has been implemented. The powerful combination of commercial robotics platforms along with our own custom-built imager and database infrastructure allows the rapid execution of crystallization experiments. Furthermore, our choice to annotate all of the crystallization drop images in our system adds a very rich layer of information to our database. We have exploited these data to design a new crystallization screen that enabled us to obtain a structure of a chemotype that had resisted co-crystallization. This experience has also encouraged us to enhance the infrastructure available for mining our extensive annotation data to routinely improve our success rate with future experiments.

ASSOCIATED CONTENT

S Supporting Information. Series of images at successive focal planes and the resulting extended depth-of-field image. Also, a series of time-lapse images from our system that show the

growth of a typical crystal and case where a crystal grows and then dissolves. The full table of the 96 conditions tested in the random screen described in this paper is included. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Author

*Dr. Eric T. Baldwin, Director, X-ray Crystallography, H3427B Bristol-Myers Squibb Company Route 206 & Provinceline Road Princeton, NJ 08540. Fax: 609-252-6012. Office: 609-252-4625. E-mail: eric.baldwin@bms.com.

ACKNOWLEDGMENT

We thank Jubilant BioSys for the diligent manual annotation of countless images of crystallization trials. Data for this study were measured at beamline X29 of the National Synchrotron Light Source. Financial support comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy, and from the National Center for Research Resources of the National Institutes of Health Grant Number P41RR012408. Bob Sweet and Howard Robinson provided assistance during data collection at the beamline. Valentina Goldfarb provided extensive protein purification expertise during the early phase of this project. We would like to thank Coleman Technologies and Edmund Optics for their help with the initial design of the automated digital microscope. We thank Steven Sheriff and Matt Pokross for reviewing the manuscript and offering useful suggestions.

ABBREVIATIONS

ASPECT=automated system for protein expression and crystallization technology
 SAR=structure–activity relationship
 HCVPr=hepatitis C virus nonstructural protein 3 protease
 SPG=succinic acid/sodium dihydrogen phosphate/glycine (molar ratio 2:7:7)
 MMT=DL-malic acid/MES/Tris base molar ratio 1:2:2
 PCB=sodium propionate/sodium cacodylate/BIS-TRIS propane (molar ratio 2:1:2)
 MIB=sodium malonate/imidazole/boric acid (molar ratio 2:3:3)
 MES=2-(*N*-morpholino)ethanesulfonic acid
 HEPES=4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
 DTT=dithiothreitol
 DMSO=dimethyl sulfoxide
 FTP=file transfer protocol
 VPN=virtual private network

REFERENCES

- (1) Adams, M.; Dailey, H.; DeLucas, L.; Prestegard, J.; Wang, B. J. *Acc. Chem. Res.* **2003**, *36*, 191–198.
- (2) Rupp, B.; Segelke, B.; Krupka, H.; Lekin, T.; Schafer, J.; Zemla, A.; Toppani, D.; Snell, G.; Earnest, T. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 1514–1518.
- (3) Graslund, S.; et al. *Nat. Methods* **2008**, *5*, 135–146.
- (4) Chayen, N.; Saridakis, E. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 921–927.
- (5) Chayen, N. *Trends Biotechnol.* **2002**, *20*, 98.
- (6) Abola, P., E.; Kuhn; Earnest, T.; Stevens, R. *Nat. Struct. Biol.* **2000**, *7*, 973–977.
- (7) Hosfield, D.; Palan, J.; Hilgers, M.; Scheibe, D.; McRee, D.; Stevens, R. J. *Struct. Biol.* **2003**, *142*, 207–217.

- (8) Lesley, S.; et al. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 1664–1669.
- (9) Stevens, R. *Nat. Biotechnol.* **2000**, *10*, 558–563.
- (10) Stevens, R. *Structure* **2007**, *15*, 1517–1519.
- (11) Bonanno, J.; et al. *J. Struct. Funct. Genomics* **2005**, *6*, 225–232.
- (12) McCarthy, A. *Chem. Biol.* **2005**, *12*, 407–408.
- (13) Mountain, V. *Chem. Biol.* **2003**, *10*, 95–98.
- (14) Ratner, M. *Nat. Biotechnol.* **2005**, *23*, 400.
- (15) Smalley, K. *Curr. Opin. Investig. Drugs* **2010**, *11*, 699–706.
- (16) Zartler, E.; Shapiro, M. *Curr. Opin. Chem. Biol.* **2005**, *9*, 366–370.
- (17) *Neurotech Insights* **2009**, Oct, 11–12.
- (18) Boggs, J. *BioWorld Today* **2009**, *20*, 1–2.
- (19) Saxty, G.; Woodhead, S.; Berdini, V.; Davies, T.; Verdonk, M.; Wyatt, P.; Boyle, R.; Barford, D.; Downham, R.; Garrett, M.; Carr, R. *J. Med. Chem.* **2007**, *50*, 2293–2296.
- (20) Warr, W. J. *Comput. Aided Mol. Des.* **2009**, *23*, 453–458.
- (21) Stevenson, J.; Umanoff, Z. www.formulatrix.com, 2009.
- (22) Murata, T. *Proc. IEEE* **1989**, *77*, 541–580.
- (23) Zhou, Y.; Murata, T.; Defanti, T. *IEEE Trans. Syst. Man. Cybern. B* **2000**, *30*, 737–756.
- (24) Russo, M.; Michalczyk, S.; Cahn, M.; Klei, H. *IEEE International Conference on Automation Science and Engineering*; 2008, CASE 2008.
- (25) Hadley, A. www.hadleyweb.pwp.blueyonder.co.uk, 2010.
- (26) Wittekind, M.; Weinheimer, S.; Zhang, Y.; Goldfarb, V. U.S. Patent 2001, US6333186 B1.
- (27) Wittekind, M.; Weinheimer, S.; Zhang, Y.; Goldfarb, V. U.S. Patent 2004, US6800456 B2.
- (28) Newman, J. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 610–612.