A conformational switch in response to Chi binding converts RecBCD from phage destruction to DNA repair

Kaiying Cheng[1], Martin Wilkinson[1], Yuriy Chaban[2] and Dale B. Wigley[1*]

[1]*Section of Structural Biology, Dept. Infectious Disease, Faculty of Medicine, Imperial College London, London SW7 2AZ, U.K.*

[2]*Electron Bio-Imaging Centre (eBIC), Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire, OX11 0DE, U.K.*

*Corresponding author – D.Wigley@imperial.ac.uk

RecBCD complex plays key roles in phage DNA degradation, CRISPR array acquisition (adaptation) and host DNA repair. The switch between these roles is regulated by a DNA sequence called Chi. CryoEM structures of the *Escherichia coli* RecBCD complex bound to a DNA fork containing a Chi sequence conformational changes in regions of the protein that contact Chi and reveals a tortuous path taken by the DNA with sequence specificity resulting from interactions with both the RecC subunit and the sequence itself. These structures provide molecular detail of how Chi is recognised and provides insight into the changes that occur in response to Chi binding that switch RecBCD from bacteriophage destruction and CRISPR spacer acquisition, to constructive host DNA repair.

INTRODUCTION

RecBCD complex, and its close relatives AddAB and AdnAB, plays a critical role in bacterial DNA double-strand break repair (1,2). These complexes load onto DNA ends resulting from a DNA break and are highly processive helicase/nucleases that digest the duplex until they encounter a special sequence called Chi (Crossover hotspot instigator) (3) at which they load RecA protein to initiate repair by homologous recombination (1,2). In *E.coli*, approximately 75% of Chi sequences are present in the leading strand with an orientation bias in the direction of DNA replication (4). This arrangement has been proposed to favour activity towards the replication origin (4).

However, the complexes are voracious DNA degrading machines that are also a major defence system against bacteriophages (that typically lack Chi sequences) and digest the viral DNA from breaks created by restriction enzyme cleavage (5,6). Indeed, the default mode of action of RecBCD and AddAB is DNA destruction, that is only modified to a benign DNA repair mode upon interaction with Chi. Furthermore, phage DNA fragments that are produced by both RecBCD and AddAB digestion are incorporated into CRISPR arrays to recognise and respond rapidly to subsequent phage infection, as well as providing a mechanism to avoid incorporation of "self" DNA into these arrays via Chi recognition (7,8). Although structural, genetic and biochemical studies of RecBCD have contributed significantly to our understanding of the role of RecBCD in the repair of DNA double-strand breaks (1), how Chi regulates the switch between bacteriophage digestion mode and DNA repair has remained enigmatic for almost half a century.

Using gel mobility shift assays, we probed specificity of RecBCD for unwound DNA forks containing a Chi sequence in the 3'-tail to determine the requirements of the complex for Chi

3

recognition. In this way we were able to identify a DNA fork with the Chi sequence spaced appropriately for recognition and response. Using this substrate, we were able to obtain high resolution structural information of the complex by cryoEM. The structure revealed a surprisingly compacted and twisted conformation of the Chi sequence, that binds and is recognised quite differently to that in an equivalent structure of the related AddAB enzyme bound to its Chi sequence (9). This different mode of binding explains a conundrum in that the final cleavage of the 3'-tail after Chi recognition occurs at a different distance from the Chi sequence in each enzyme. Furthermore, the structure reveals an on-enzyme equilibrium between Chi recognised and unrecognised states that explains a mechanism to prevent phage escape from these destructive enzymes.

RESULTS

DNA substrate optimisation

The recognition of Chi by *E. coli* RecBCD requires the sequence to enter the enzyme complex as a DNA duplex which is then unwound to present the Chi sequence (GCTGGTGG) in a single-stranded DNA context in the 3'-terminated strand (10). Previous studies on both RecBCD and the related AddAB system have shown that the RecBCD complex is more highly regulated than AddAB and that this control emanates from the RecD subunit (11-15). Indeed, the nuclease activity of the RecB subunit1 (14) is controlled by the RecD subunit (11,13) by a mechanism that involves binding of the 5'-tail to RecD to induce a conformational change in RecB, releasing an α-helix that blocks the nuclease active site (15,16). Interactions of RecD with the 5'-tail also alter the conformation of domains of the RecC subunit, the subunit proposed to interact with Chi (15,17-22), indicating that a 5'-tail of at least 15 bases would be required to activate the enzyme. On this basis, we then conducted experiments to determine the

specificity requirements for the 3'-tail and design a synthetic DNA fork that would mimic an encounter with Chi on a physiological substrate (Figure 1a).

Cryo-EM structure of the Chi-recognition complex

On the basis of the experimental data, we chose a DNA fork substrate with 25 base pair duplex, a 15 base 5'-tail and a 20 base 3'-tail comprising eight T bases, followed by the eight base Chi sequence (GCTGGTGG), and terminated with four T residues (Figure 1a). CryoEM analysis resulted in three main structures – a 3.7Å structure of the Chi-recognised complex, a 3.9Å structure with a DNA fork bound but in which Chi has not been recognised, and an intermediate structure at 4.1Å (Figures 1b & 1c, Table 1, Supplementary Figures 1 – 3, Supplementary Movie 1).

The structure in which Chi is not recognised is indistinguishable from our previous structure with a long 5'-tail (16). Density corresponding to the bound DNA fork is visible for the duplex region, the 5'-tail, and for the 3'-tail across the RecB subunit, although it is disordered beyond the sixth base so the Chi sequence is not visible (Figure 2a). Consequently, this structure provides a high resolution internal control to delineate which conformational changes are specific to the Chi response. As further confirmation, the structure of a complex with an equivalent substrate that lacks a Chi sequence (Chi-minus) and does not induce gel mobility shifts (Figure 1a), was determined at 3.8Å resolution and shows no density in the Chi-binding site (Table 1, Supplementary Figures 2 & 3). Similarly, structures of complexes with forks in which the 3'-tail contained either six (Chi-minus 2, 3.6Å resolution) or ten (Chi-plus 2, 3.8Å resolution) bases between the junction and the Chi site also revealed no density in the Chi site (Table 2, Supplementary Figures 3 – 5). Consequently, we can be confident that the changes we observe in the Chi-bound complexes are a result of specific interactions with that sequence

and that a spacer of eight residues is required between the fork junction and the start of the Chi site.

In the Chi-recognised complex, density for the bound DNA is visible for all twenty bases of the 3'-tail, including the Chi sequence (Figure 2b, Supplementary Figures 3 & 4, Supplementary Movie 1). Furthermore, the overall structure of the protein component showed Chi-dependent domain shifts that likely explain the altered gel mobility (Figures 1b & 1c, Supplementary Movie 2). Conformational changes in the RecB nuclease and RecC domains were observed as well as smaller changes in specific regions of the RecC subunit, many of which make direct interactions with the Chi sequence (Figures 2 - 4). The conformational changes in RecC that are induced by the 5'-tail binding to RecD (15) appear to facilitate Chi recognition and bring several residues into proximity of the Chi-binding site prior to Chi binding, switching the subunit into a Chi-recognition mode. Binding to Chi induces a rotation of the 1A and 1B domains of RecC. The 1A domain of RecC forms a significant part of the interface with the RecB nuclease domain and the largest Chi-dependent movement is of the nuclease domain that rotates towards the Chi-binding site as part of a rigid-body movement with domain 1A of RecC (Supplementary Movie 2). The active site residues of the nuclease domain are now located close to the terminal T residue of the 3'-tail (Figure 5). The final cleavage of the 3'-tail is located 4-6 bases after the Chi sequence (23), suggesting the structure is that after the final cleavage event following Chi recognition.

Details of the interactions between RecBCD and the bound Chi sequence (Figures 2 - 4, Supplementary Movies 1 & 3) reveal multiple side chain interactions with residues in RecC, many of which have been identified from mutational studies (19,20,22). Interactions with each of the eight bases of the Chi sequence provide specificity while interactions with the

phosphodiester backbone (mainly arginine side chains) help to stabilise the twisted conformation of the DNA, allowing the first and fourth residues (both guanines) of the Chi sequence to stack upon each other which provides additional specificity. It has been determined that mutation of any one of the eight bases in Chi is sufficient to reduce affinity to a level that is not functional *in vivo* (20).

In addition to the particles we assigned to the Chi-recognised structure, a further group were isolated in a class that also showed good density for Chi (Supplementary Figures 1 & 3). The majority of the structure was the same as the Chi-recognised complex, but the nuclease domain was in a slightly altered location, intermediate between the positions of the recognised and unrecognised complexes, that was slightly misaligned for interaction with the 3'-terminus. The density for the nuclease was also less well ordered (Supplementary Figure 1) suggesting the domain is more mobile. This class appears to have recognised Chi but is not optimally aligned for cleavage so we suggest this may be an intermediate in which Chi has been recognised but the nuclease is not aligned for cleavage and during active translocation may not have time to cleave before the Chi sequence is pushed out. We, therefore, refer to this as the Chi-intermediate structure.

Comparison with AddAB/Chi recognition complex

RecBCD shares many aspects of its activity with AddAB (1,2) and the structure of *B.subtilis* AddAB bound to its Chi sequence (AGCGG) has been determined (9). However, comparison of the structures reveals that interactions with their respective Chi sequences are very different, even though the structures and conformations of the protein components, and their interactions with the DNA fork, are all very similar (Figure 4). In particular, the conformation of the Chi sequence itself is completely different despite the two sequences sharing some apparent

similarity (9) (Figure 4). However, the different interaction modes for Chi with RecC and AddB observed in the structures resolves a conundrum that the final cleavage event is 4-6 bases after Chi for RecBCD (23) yet is a single base for AddAB (24). In each case, the bound ssDNA is presented to the nuclease active site in a manner that is consistent with these differing cleavage sites (Figures 4 & 5). Consistent with our structural data, residues beyond the canonical eight base Chi site may also contribute to specificity, particularly in the 4-7 bases on the 3' side of the Chi sequence (21,22), albeit with reduced specificity. Furthermore, a RecBCD complex that carries a mutation in the RecC protein (17) recognises an extended eleven base Chi sequence (Chi* = GCTGGTGCTCG) that includes this region (25).

Mechanism of RecBCD stalling

Single-molecule studies have shown that, upon encountering Chi, RecBCD pauses for several seconds before recommencing translocation at (usually) a reduced speed (26,27). It is also known that an active RecB motor is required for Chi recognition and response (28). The RecB motor precedes the Chi-binding site in RecC so reactivation would push DNA towards the bound Chi sequence. However, the twisted and compacted Chi sequence would act as a physical block to prevent translocation and the DNA would become compressed within the space at the RecB/RecC interface. The conformational changes that provide additional interactions between RecC and the Chi sequence would tighten this grip even further and prevent translocation across and through the RecC subunit. As described previously (19,20), once translocation by the RecB motor recommences, the only exit for this DNA would be through a protein gate between the RecB and RecC subunits that is operated by an ion pair "latch". Such a mechanism has similarities to the "DNA scrunching" mechanism observed for RNA polymerase during initiation of transcription that forces the RNA product out of the enzyme complex (29). Blocking the 3'-channel would also explain why nuclease digestion of

the 3'-strand is attenuated after Chi binding (30). Although the structure explains why RecBCD stalls at Chi sites (26,27) it does not provide details of the subsequent steps that precede full conversion to repair mode such as reactivation of translocation or how the RecD subunit becomes inactivated (27).

A mechanism to prevent phage evasion

It is known that RecBCD only cleaves at a Chi sequence in 25-40% of encounters (30,31). There are two explanations for this counter intuitive behaviour. The first is that the rate of translocation by RecBCD is so rapid that the Chi sequence passes over the RecC binding site before the complex has a chance to recognise the sequence. A second alternative is that there is an intrinsic equilibrium between bound and unbound states that is reached at every step of translocation. For this to be the case, at a translocation speed of around 1000 bases per second, this equilibrium would need to be established within 1 ms. For the particles with bound DNA substrate, around 40% had the Chi sequence bound specifically and were in the Chi-recognised state with the nuclease correctly positioned (Supplementary Figure 1), suggesting that the latter explanation is correct. Although an additional 20% have also bound Chi (Chi-intermediate), the nuclease domain does not appear to be correctly positioned for cleavage. The remaining 40% of molecules have failed to recognise Chi. Since single-molecule studies of RecBCD indicate that translocation speeds vary considerably (26,27), this would ensure that Chi is recognised and cleaved with similar frequency no matter what the translocation speed is on the 3'-strand (RecB motor). Furthermore, bacteriophages that infect *E. coli*, such as lambda, typically lack Chi sequences or have them at very low frequency. Consequently, the reduced recognition frequency is an adaptation mechanism to ensure that even bacteriophages which acquire a Chi sequence will likely fail to escape degradation by RecBCD. By contrast, the elevated frequency of Chi sequences in the *E. coli* genome (4) ensures DNA repair is initiated

within a few kilobases of the break site. Thus, these contradictory requirements for Chi recognition can be achieved simultaneously; allowing RecBCD to make the switch between being friend to the host, yet foe to bacteriophages.

DISCUSSION

The twisted conformation of the Chi sequence when bound to RecBCD was unanticipated in light of the related AddAB/Chi structure reported previously (9). In AddAB, the Chi sequence is extended with each base sitting in a pocket on the surface of the AddB subunit, making side chain contacts with proteins side chains that explain the specificity for the sequence. By contrast, RecC contacts Chi in a quite different manner with fewer base-specific contacts. Instead, the specificity appears to come from an overall interaction with a specific shape of the bound ssDNA that is stabilised, at least in part, by contacts such as the intra-sequence contacts between the first and fourth guanine residues in the sequence. Nonetheless, many side chains have been shown to be essential for Chi recognition (19,20,22) so these are, presumably, assisting in the shape recognition that leads to sequence specificity.

RecBCD has been validated as a target for antibiotic development (32). The differing modes of Chi interaction adopted by AddAB and RecBCD have important implications for drug discovery. Recognition of Chi by AddAB provides opportunities for the design of specific drugs that might interact at any or each of the base specificity pockets to prevent Chi binding. Such interactions could be different in organisms with different Chi sequences creating the potential for narrow spectrum targeting of groups of organisms with similar Chi sequences. By contrast, blocking of Chi recognition in RecBCD may require a different approach to find compounds that interact with a part of the pocket in a different manner. Nevertheless, the

10

twisted conformation of the Chi sequence provides multiple surfaces and pockets that might provide specificity for drug design.

In *E.coli*, 75% of Chi sequence are located in the leading strand and oriented with a bias to favour activity towards the replication origin (4). Such an arrangement has been proposed to promote DNA repair processes during replication (4,33). Given this property, one would imagine that determination of Chi sequences would be straightforward in different organisms. However, that has proved not to be the case and analysis of bacterial genomic data has had limited success in identifying Chi sequences (34). Indeed, even the notion that Chi sequences have evolved to facilitate repair has been questioned. Instead, it was shown that G-rich sequences are naturally more prevalent and the orientation of G-rich sequences other than Chi also showed an apparent directional bias in *E.coli* but less so in other organisms (such as *H.influenzae*) that do not show an intrinsic sequence bias (35). This suggests that rather than Chi sequences evolving as a regulator of Chi, an alternative is that RecBCD adopted a naturally occurring sequence with appropriate characteristics as a regulator, possibly initially with a higher redundancy but later to be more specific as a co-evolution with the abundance of sequence of Chi itself. Such a proposal might explain why Chi sites have proved to be so hard to identify from bioinformatics analysis alone because under these circumstances there need be no conservation of sequences between organisms. To date, Chi sequences that bind to RecBCD have only been identified in two other organisms (*H.influenzae* and *P.synringae*) and these are both very similar to the *E.coli* sequence (*E.coli* – GCTGGTGG, *H.influenzae* – G(**G/C**)TGGTGG and *P.syringae* – GCTGG**C**G**C**).

The structures presented here help us to explain the long-standing observation of Chi-regulation and its effects on phage biology. They help explain the switch from destructive to

repair mode but several questions remain unanswered. For example, the structures show how the enzyme is stopped in its tracks during degradation but they do not explain how the complex changes in order to recommence translocation after pausing at the Chi site. Several events appear to take place after Chi recognition including inactivation of the RecD motor (27) and loading of RecA protein onto the 3'-tail. Hopefully, future structures will shed light on these processes.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

KC conducted biochemical experiments and prepared samples. KC, MW, & YC collected cryoEM data. KC and MW processed data. KC, MW & DBW analysed data. DBW wrote the manuscript with input from KC & MW.

REFERENCES

1) Dillingham, M.S. & Kowalczykowski, S.C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72**, 642-71 (2008).

2) Wigley, D.B. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat. Rev. Microbiol.* **11**, 9-13 (2013).

3) Lam, S.T., Stahl, M.M., McMilin, K.D. & Stahl, F.W. Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* **77**, 425-33 (1974).

4) Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y. The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-62 (1997).

5) Simmon, V.F. & Lederberg, S. Degradation of bacteriophage lambda deoxyribonucleic acid after restriction by Escherichia coli K-12. *J. Bacteriol.* **112**, 161-9 (1972).

6) Dharmalingam, K. & Goldberg, E.B. Mechanism localisation and control of restriction cleavage of phage T4 and lambda chromosomes in vivo. *Nature* **260**, 406-10 (1976).

7) Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U.

& Sorek, R. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505-510 (2015).

8) Modell, J.W., Jiang, W. & Marraffini, L.A. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101-104 (2017).

9) Krajewski, W.W., Fu, X., Wilkinson, M., Cronin, N.B., Dillingham, M.S. & Wigley, D.B. Structural basis for translocation by AddAB helicase/nuclease and its arrest at Chi sites. *Nature* **508**, 416-9 (2014).

10) Bianco, P.R. & Kowalczykowski, S.C. The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGGTGG-3'. *Proc. Natl. Acad. Sci. U S A.* **94**, 6706-11 (1997).

11) Lieberman, R.P. & Oishi, M. The recBC deoxyribonuclease of Escherichia coli: isolation and characterization of the subunit proteins and reconstitution of the enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4816-20 (1974).

12) Amundsen, S.K., Taylor, A.F., Chaudhury, A.M. & Smith, G.R. RecD: the gene for an essential third subunit of exonuclease V. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5558-62 (1986).

13) Chaudhury, A.M. & Smith, G.R. A new class of Escherichia coli recBC mutants: implications for the role of RecBC enzyme in homologous recombination. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7850-7854 (1984).

14) Yu, M., Souaya, J. & Julin, D.A. The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 981-6 (1998).

15) Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C. & Wigley, D.B. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**,187-93 (2004).

16) Wilkinson, M., Chaban, Y. & Wigley, D.B. Mechanism for nuclease regulation by RecBCD. *Elife* **5**, pii: e18227 (2016).

17) Schultz, D.W., Taylor, A.F. & Smith G.R. *Escherichia coli* RecBC pseudorevertants lacking Chi recombinational hotspot activity. *J. Bacteriol.* **155**, 664-80 (1983).

18) Ponticelli, A.S., Schultz, D.W., Taylor, A.F. & Smith, G.R. Chi-dependent DNA strand cleavage by RecBCD enzyme. *Cell.* **41**, 145-51 (1985).

19) Handa, N., Yang, L., Dillingham, M.S., Kobayashi, I., Wigley, D.B. & Kowalczykowski, S.C. Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, χ, by RecBCD enzyme. *Proc. Natl. Acad. Sci. U S A.* **109**, 8901-6 (2012).

20) Yang, L., Handa, N., Liu, B., Dillingham, M.S., Wigley, D.B. & Kowalczykowski, S.C. Alteration of χ recognition by RecBCD reveals a regulated molecular latch and suggests a channel-bypass mechanism for biological control. *Proc. Natl. Acad. Sci. U S A.* **109**, 8907-12 (2012).

21) Taylor, A.F., Amundsen, S.K. & Smith, G.R. Unexpected DNA context-dependence identifies a new determinant of Chi recombination hotspots. *Nucleic Acids. Res.* **44**, 8216-28 (2016).

22) Amundsen, S.K., Sharp, J.W. & Smith, G.R. RecBCD Enzyme "Chi Recognition" Mutants Recognize Chi Recombination Hotspots in the Right DNA Context. *Genetics* **204**, 139-52 (2016).

23) Taylor, A.F., Schultz, D.W., Ponticelli, A.S. & Smith, G.R. RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation-dependence of the cutting. *Cell* **41**, 153-63 (1985).

24) Chedin, F., Ehrlich, S.D. & Kowalczykowski, S.C. The *Bacillus subtilis* AddAB helicase/nuclease is regulated by its cognate Chi sequence *in vitro*. *J. Mol. Biol.* **298**, 7-20 (2000).

25) Handa, N., Ohashi, S., Kusano, K. & Kobayashi, I. χ^*, a χ-related 11-mer sequence partially active in an *E. coli recC** strain. *Genes Cells* **2**, 525-36 (1997).

26) Spies, M., Bianco, P.R., Dillingham, M.S., Handa, N., Baskin, R.J. & Kowalczykowski, S.C. A Molecular Throttle: The Recombination Hotspot χ Controls DNA Translocation by the RecBCD Helicase. *Cell* **114**, 647-654 (2003).

27) Spies, M., Amitani, I., Baskin, R.J. & Kowalczykowski, S.C. RecBCD enzyme switches lead motor subunits in response to chi recognition. *Cell* **131**, 694-705 (2007).

28) Spies, M., Dillingham, M.S. & Kowalczykowski, S.C. Translocation by the RecB motor is an absolute requirement for χ-recognition and RecA protein loading by RecBCD enzyme. *J. Biol. Chem.* **280**, 37078-87 (2005).

29) Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S & Ehbright, R.H. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1139-43 (2006).

30) Dixon, D.A. & Kowalczykowski, S.C. The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. *Cell* **73**, 87-96 (1993).

31) Taylor, A.F. & Smith, G.R. RecBCD enzyme is altered upon cutting DNA at a chi recombination hotspot. *Proc. Natl. Acad. Sci. U S A.* **89**, 5226-30 (1992).

32) Wilkinson, M., Troman, L., Wan Nur Ismah, W.A., Chaban, Y., Avison, M.B., Dillingham, M.S. & Wigley, D.B. Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. *Elife* **5**, pii: e22963 (2016).

33) El Karoui, M., Biaudet, V., Schbath, S. & Gruss, A. Characteristics of Chi distribution on different bacterial genomes. *Res. Microbiol.* **150**, 579-87 (1999).

34) Uno, R., Nakayama, Y., Arakawa, K. & Tomita, M. The orientation bias of Chi sequences is a general tendency of G-rich oligomers. *Gene* **259**, 207-15 (2000).

35) Halpern, D., Chiapello, H., Schbath, S., Robin, S., Hennequet-Antier, C., Gruss, A. & El Karoui, M. Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genet.* **3**, 1614-21 (2007).

36) Saikrishnan, K., Griffiths, S.P., Cook, N., Court, R. & Wigley, D.B. DNA binding to RecD: role of the 1B domain in SF1B helicase activity. *EMBO J.* **27**, 2222-9 (2008).

37) Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., & Agard, D.A. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-2 (2017).

38) Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1-12 (2016).

39) Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., & Scheres, S.H. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, pii:e42166 (2018).

40) Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E. C., & Ferrin, T.E. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* **25**, 1605-1612 (2004).

41) Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-55 (1997).

42) Burnley, T., Palmer, C.M., & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol.* **73**, 469-477 (2017).

43) Emsley, P., Lohkamp, B., Scott, W.G., & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).

44) Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., & Adams, P.D. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531-544 (2018).

45) Kucukelbir, A., Sigworth, F.J,. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods.* 11, 63-5 (2014).

FIGURE LEGENDS

Fig. 1: Structural changes associated with Chi binding.

a, Native gel mobility shift assays show an altered shift of the RecBCD-DNA complex band when the substrate contained eight bases between the fork junction and the Chi sequence (Y = 8), suggesting a Chi-specific conformational change. The star indicates position of the Cy5 fluorescent label. b, Overview of the conformational changes between the Chi-recognised and the Chi-unrecognised states. The two structures were superimposed on the RecB motor domains. The unchanged domains of both structures are shown in the same colours: RecB in wheat, RecC in light blue and RecD in pale green. The areas that undergo large conformational changes in the Chi-recognised state are coloured darker: RecB nuclease domain in brown and RecC helicase-like domains in blue. The Chi sequence is coloured red and the four bases after Chi are yellow, with the rest of the DNA in grey. c, Expanded and rotated view to emphasise the conformational changes in RecC and RecB nuclease domain around Chi (red). Arrows indicate direction of movements.

Fig. 2: Details of the Chi-binding site interactions.

a, Density corresponding to the bound DNA fork substrate in the Chi-unrecognised state with the protein complex shown as a faint cartoon. Strong density is observed for the duplex and 5'-tail regions. Density is also observed for the first six bases of the 3'-tail, but is disordered beyond that. **b**, Density corresponding to the bound DNA fork substrate in the Chi-recognised state. The density is similar to that observed in the Chi-unrecognised state except that additional density is observed for the entire 20 bases of the 3'-tail, including the Chi sequence. **Inset**, Expanded view of the density (contoured at 2.5 σ) for the eight Chi bases

with the Chi bases individually coloured as in panel (**c**) demonstrating the highly twisted conformation of the DNA. **c**, Details of the interactions with each base of the Chi sequence. Hydrogen bonds are shown as dotted lines.

Figure 3: Details of the full Chi-binding site.

A cross-eye stereo of the Chi-binding site and interacting residues is shown. Interacting residues are shown as sticks and coloured according to previously published point mutation studies (19,20). Residues which failed to recognize Chi are in violet, those that showed altered/relaxed Chi recognition specificity are in green, and those that showed little or no influence after mutation are grey. Additional residues which failed to recognize Chi after mutation (22) are coloured in yellow.

Fig. 4: Comparison of RecBCD and AddAB Chi-bound complexes.

a, Transparent surfaces of RecBCD (above) and AddAB (below) coloured by subunit with the DNA fork backbones overlaid as a cartoon in grey with their respective Chi sequences depicted in red and the bases between Chi and the nuclease active site in yellow. **b**, Expanded cartoon view of the Chi-binding region in each structure with the RecC domains labelled to highlight the different Chi-binding surfaces utilised by the two enzymes. **c**, Cartoon representation detailing the contacts between the proteins and DNA.

Fig. 5: The nuclease active site in the Chi-bound RecBCD and AddAB complexes.

The nuclease domains of RecBCD and AddAB are shown as cartoons. Key residues participating in DNA binding and catalysis are shown as sticks. Hydrogen bonds are shown as dotted lines. The DNA forks used for both enzymes represent substrates after the final cleavage, with four bases after Chi for RecBCD but just a single base for AddAB (in each case depicted in yellow). The difference in the final cleavage site relates to the distance between the end of the Chi-binding site and the nuclease even though the relative position of the nuclease, relative to RecC or AddB, are similar. The normal cleavage of DNA after Chi in RecBCD would result in a terminal 3'-phosphate but this was not added to the substrate synthesised for the structure determination.

Table 1 Cryo-EM data collection, refinement and validation statistics

	Chi-recognised (EMDB-10214) (PDB 6SJB)	Chi-intermediate (EMDB-10215) (PDB 6SJE)	Chi-unrecognised (EMDB-10216) (PDB 6SJF)	Chi-minus (EMDB-10217) (PDB 6SJG)
Data collection and processing				
Magnification	47,755	47,755	47,755	47,710
Voltage (kV)	300	300	300	300
Detector	Gatan K2	Gatan K2	Gatan K2	Gatan K2
Electron exposure (e–/Å^2)	45	45	45	45
Dose rate (e–/pixel/s)	5	5	5	7
Pixel size (Å)	1.047	1.047	1.047	1.048
Defocus range (μm)	-1.0 to -3.0	-1.0 to -3.0	-1.0 to -3.0	-1.0 to -2.5
Initial particle images (no.)	373,476	373,476	373,476	99,545
Final particle images (no.)	74,496	44,320	74,273	62,812
Symmetry imposed	C1	C1	C1	C1
Map resolution (Å)	3.7	4.1	3.9	3.8
FSC threshold	0.143	0.143	0.143	0.143
Refinement				
Initial model used (PDB code)	5LD2	6SJB	5LD2	6SJF
Model resolution (Å)	3.7	4.1	3.9	3.8
FSC threshold	0.143	0.143	0.143	0.143
Refinement program	Phenix 1.12-2829	Phenix 1.12-2829	Phenix 1.12-2829	Phenix 1.12-2829
Model to map CC (masked)	0.80	0.77	0.78	0.73
Map sharpening B factor (Å^2)	-50	-50	-50	-50
Model composition				
Non-hydrogen atoms	24,115	24,115	23,826	23,826
Protein residues	2,847	2,847	2,847	2,847
DNA residues	68	68	54	54
B factors (Å^2)				
Protein	105	118	106	109
DNA	155	172	199	241
R.m.s. deviations				
Bond lengths (Å)	0.003	0.003	0.002	0.002
Bond angles (°)	0.56	0.58	0.52	0.54
Validation				
Clashscore	4.4	4.3	4.4	4.3
Poor rotamers (%)	2.6	3.1	3.0	2.9
Ramachandran plot				
Favored (%)	97.7	97.9	98.0	97.8
Allowed (%)	2.3	2.1	2.0	2.2
Disallowed (%)	0.0	0.0	0.0	0.0

FSC = Fourier shell correlation, R.m.s = root means squared, CC = correlation coefficient as calculated by phenix.real_space_refine

Table 2 Cryo-EM data collection, refinement and validation statistics

	Chi-minus 2 (EMDB-10369) (PDB-6T2U)	Chi-plus 2 (EMDB-10370) (PDB-6T2V)
Data collection and processing		
Magnification	75,000	75,000
Voltage (kV)	300	300
Detector	Falcon3 (Integrating)	Falcon3 (Integrating)
Electron exposure (e–/Å^2)	75.9	75.9
Dose rate (e–/pixel/s)	5	5
Pixel size (Å)	1.085	1.085
Defocus range (μm)	-1.2 to -2.7	-1.2 to -2.7
Initial particle images (no.)	754,150	795,810
Final particle images (no.)	379,790	380,912
Symmetry imposed	C1	C1
Map resolution (Å)	3.6	3.8
FSC threshold	0.143	0.143
Refinement		
Initial model used (PDB code)	5LD2	5LD2
Model resolution (Å)	3.6	3.8
FSC threshold	0.143	0.143
Refinement program	Phenix 1.12-2829	Phenix 1.12-2829
Model to map CC (masked)	0.71	0.73
Map sharpening B factor (Å^2)	-100	-100
Model composition		
Non-hydrogen atoms	23,826	23,826
Protein residues	2,847	2,847
DNA residues	54	54
B factors (Å^2)		
Protein	94	135
DNA	147	208
R.m.s. deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.46	0.43
Validation		
Clashscore	4.5	4.6
Poor rotamers (%)	3.0	3.0
Ramachandran plot		
Favored (%)	98.0	98.0
Allowed (%)	2.0	2.0
Disallowed (%)	0.0	0.0

FSC = Fourier shell correlation, R.m.s = root means squared, CC = correlation coefficient as calculated by phenix.real_space_refine

METHODS

Purification of RecBCD

To prevent digestion of the single-stranded tails of the DNA substrate, a nuclease-deficient *E. coli* RecBCD mutant (D1080A in RecB) was used in this work. The complex was produced from three plasmids, pETduet-His6-TEVsite-recBD1080A, pRSFduet-recC and pCDFduet-recD in a ΔrecBD *E. coli* strain as described previously (14). His-tagged RecBCD complex was expressed and purified as described previously (36) with the His-tag removed by TEV protease. The purified protein was flash-frozen with 15% glycerol and stored at -80 °C until use.

Preparation of forked DNA substrates

The forked DNA substrates used during this work were made by annealing equal molar quantities of two respective ssDNA oligonucleotides to generate forks with 25 bp duplex, 15 base single-stranded 5'-tail and a 20 base single-stranded 3'-tail. The same upper strand, responsible for generating the 5'-tail, was used in all cases and had the sequence: 5' TTTTTTTTTTTTTTTgagcgactgcactacaacagaacca 3' (lower case represents the bases that form the duplex region). The variable lower strand, responsible for presenting Chi (GCTGGTGG), contained the sequence: 5' tggttctgttgtagtgcagtcgctc(Y)GCTGGTGG(Z) 3'. Y and Z represent the variable number of T bases used to alter the position of the Chi sequence whilst keeping the total number of unpaired bases to 20 (see Figure 1). In the Chi-

minus negative control substrate, the Chi sequence was replaced by eight T bases to give a 3'-tail of twenty unpaired T bases.

Native gel mobility assays

For native gel mobility assays, all DNA substrates contained the upper strand labelled with fluorophore (Cy5) at the 5'-terminus and for the lower strand Y values of 1-10 with corresponding Z values of 11-2 were used. Typically, 500 nM RecB^{D1080A}CD was incubated with 500 nM DNA substrate in a 5µl binding solution containing 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.5 mM TCEP and 5% glycerol, which was incubated on ice for 10 min. Samples were separated on 5% native polyacrylamide gels in 1 × TB buffer. Gels were scanned in fluorescence mode (Cy5) on a BIO-RAD ChemiDocTM MP imaging system.

Cryo-electron microscopy grid preparation and data collection

Prior to preparing grids, RecBCD was thawed and desalted into 20 mM Tris-HCl (pH 8.0), 50mM NaCl, 0.5 mM TCEP using Sephadex G25 spintrap columns (GE Healthcare). The protein was mixed with a 1.5 fold excess of DNA substrate for 10 min at room temperature. Final concentrations were 1 µM RecBCD and 1.5 µM DNA substrate. A RecBCD-Chi complex was prepared using a Chi-containing DNA substrate with a Y spacing of 8 and a Z spacing of 4 respectively (see Figure 1). Separately, for the Chi minus dataset, a complex was prepared using the negative control substrate lacking Chi. Quantifoil R2/2 µm holey carbon film grids (300 mesh) were treated by plasma cleaning for 30 s before being covered with graphene oxide sheets. In order to improve grid preparation reproducibility and hydrophilicity of the graphene oxide, the following method was used: 5µl of graphene oxide

solution (Aldrich 763705, 2mg/ml) were mixed with 5µl of 1.5% w/v solution of nonionic detergent n-Dodecyl-β-D-Maltoside. The mixture was diluted 100 times with water and 5ul was applied to the carbon side of the grid. A sharp edge of a piece of filter paper was applied to the centre of the opposite surface of the grid to pull the graphene oxide solution through the grid. Grids were used within 30 minutes of graphene oxide application. Sample (4 µL) was evenly applied to the graphene oxide-coated side of the grid, followed by a 5 s wait time, 1 s blot time, and freezing in liquid ethane using a Vitrobot Mark IV (FEI). The Vitrobot chamber was maintained at close to 100% humidity at 4°C.

The dataset with the Chi substrate was collected using a Titan Krios microscope operated at 300 KV at eBIC, Diamond, UK. Zero loss energy images were collected automatically using EPU (FEI) on a Gatan K2-Summit detector in counting mode with a pixel size of 1.047 Å. A total of 3,721 images were collected with a nominal defocus range of −1.3 to −2.5 µm in 0.3 µm increments. Each image consisted of a movie stack of 40 frames with a total dose of 45 e-/Å2 over 10 s corresponding to a dose rate of 5 e-/pixel/s. The dataset with the Chi-minus control substrate was collected using a similar collection strategy again with a Gatan K2-Summit detector and Titan Krios microscope at eBIC, Diamond, UK. The pixel size was 1.048 Å and a total of 788 images were collected with a similar defocus range to the above. A total dose of 45 e-/Å2 was split into 40 frames over 7 s, corresponding to a dose rate of 7 e-/pixel/s.

The datasets with the Chi-minus 2 and Chi-plus 2 were collected using a Titan Krios microscope operated at 300 KV at eBIC, Diamond, UK. Zero loss energy images were collected automatically using EPU (FEI) on a Falcon3 detector in integrating mode with a pixel size of 1.085 Å. A total of 4,013 images (Chi-minus 2 substrate) and 3,613 images

(Chi-plus 2 substrate) were collected with a nominal defocus range of −1.2 to −2.7 μm in 0.3 μm increments. Each image consisted of a movie stack of 39 frames with a total dose of 76 e⁻/Å² over 1 s corresponding to a dose rate of 89 e⁻/pixel/s.

Data processing – The Chi-containing dataset

Movie stacks were aligned and summed using Motioncor2 (37). Template-free particle picking was done in Gautomatch, using a circular diameter of 150 Å and CTF parameters were estimated for each micrograph using Gctf (38). A total of 1,053,451 picked particles were extracted 2x binned from 3,721 images for two rounds of 2D classification in RELION3 (39) in which 373,476 real RecBCD particles were kept and picking artefacts/noise discarded. A consensus refinement was run on all the particles followed by unmasked 3D classification without alignment. From this, 204,633 particles with complete density for the complex were selected and sub-stoichiometric classes discarded (no DNA density, no RecB nuclease domain density or weaker RecD density). These 204,633 particles were re-extracted unbinned, 3D refined and Bayesian polished prior to 3D classification without alignment using a mask of RecC around the prospective Chi-binding site. This separated 127,268 particles with ordered density for the complete ssDNA 3' tail, including the Chi sequence, from 77,365 particles with no ssDNA density beyond the RecB helicase domains.

Each class was further refined and classified without alignment, this time using a mask around both RecD and the RecB nuclease. The classifications removed 8,452 and 3,092 particles from the Chi and no Chi classes respectively, which had weaker density for the RecD helicase domains. This left 74,263 homogenous particles from the no Chi class, that were refined to give a 3.9 Å resolution (0.143 FSC cutoff, RELION3) map representing a Chi-unrecognised complex. This structure was very similar a previous RecBCD structure

28

with a forked DNA substrate containing long ssDNA tails (PDB: 5LD2) (16). From the Chi class, two significant conformations were identified. The major state, with 74,496 particles, refined to a resolution of 3.7 Å (0.143 FSC cutoff, RELION3) and represented a Chi-recognised complex with ordered Chi density and significant local conformational changes involving RecC and the RecB nuclease. The second Chi state contained 44,320 particles, refined to 4.1 Å (0.143 FSC cutoff, RELION3) and similarly showed ordered Chi density and the associated local conformational changes in RecC. However, this third state did not show such a substantial movement in the RecB nuclease domain, therefore may represent an intermediate between the Chi-unrecognised and Chi-recognised states, which we have called the Chi-intermediate complex.

Data processing – The Chi-minus control dataset

The Chi-minus movie stacks were initially processed similarly to the Chi dataset. A total of 271,210 particles were extracted 2x binned from 788 images for two rounds of 2D classification in RELION3, resulting in 99,545 RecBCD particles. A consensus 3D refinement was run followed by unmasked 3D classification without alignment from which a homogeneous subset of 87,229 particles with complete density for the complex was selected. The particles were re-extracted unbinned, refined and Bayesian polished prior to 3D classification without alignment using the mask of RecC around the Chi binding site. This time there was only one class, with no DNA density observed in the Chi-binding channel. 3D classification with a mask of RecD plus the RecB nuclease domain separated a class of sub-stoichiometric particles (24,417 particles). The remaining major class (62,812 particles) was refined to produce a 3.8 Å resolution map (0.143 FSC cutoff, RELION3). This represents the

Chi-minus complex, which resembles the Chi-unrecognised complex and similarly does not contain ordered 3' ssDNA beyond the RecB helicase.

Data processing – The Chi-minus 2 dataset

The movie stacks were initially processed similarly to the Chi-containing dataset. A total of 1,694,205 particles were extracted 2x binned from 4,013 images for two rounds of 2D classification in RELION3, resulting in 754,150 RecBCD particles. A consensus 3D refinement was run followed by unmasked 3D classification without alignment from which a homogeneous subset of 687,825 particles with complete density for the complex was selected. The particles were re-extracted unbinned, refined and Bayesian polished prior to 3D classification without alignment using the mask of RecC around the Chi binding site. Only one class, with no DNA density observed in the Chi-binding channel was obtained. 3D classification with a mask of RecD plus the RecB nuclease domain separated a class of sub-stoichiometric particles (308,035 particles). The remaining major class (379,790 particles) was refined to produce a 3.6 Å resolution map (0.143 FSC cutoff, RELION3). This structure resembles the Chi-unrecognised complex and similarly does not contain ordered 3' ssDNA beyond the RecB helicase.
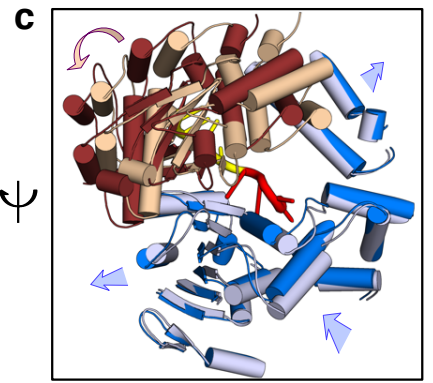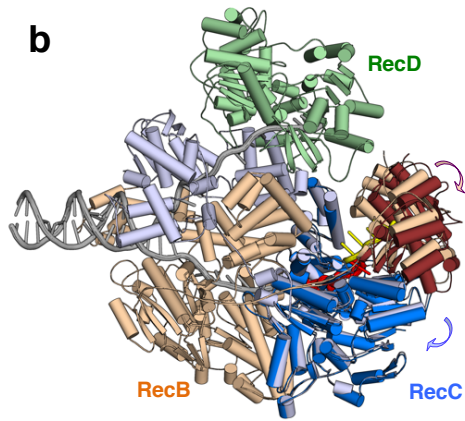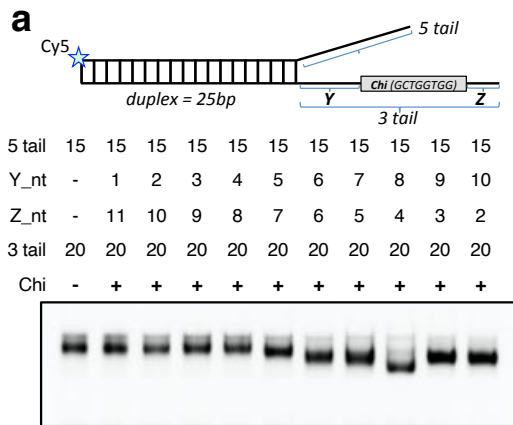
Data processing – The Chi-plus 2 dataset

The movie stacks were initially processed similarly to the Chi-containing dataset. A total of 1,809,768 particles were extracted 2x binned from 3,613 images for two rounds of 2D classification in RELION3, resulting in 795,810 RecBCD particles. A consensus 3D refinement was run followed by unmasked 3D classification without alignment from which a homogeneous subset of 776, 381 particles with complete density for the complex was

selected. The particles were re-extracted unbinned, refined and Bayesian polished prior to 3D classification without alignment using the mask of RecC around the Chi binding site. Only one class, with no DNA density observed in the Chi-binding channel was obtained. 3D classification with a mask of RecD plus the RecB nuclease domain separated a class of sub-stoichiometric particles (395,469 particles). The remaining class (380,912 particles) was refined to produce a 3.8 Å resolution map (0.143 FSC cutoff, RELION3). This structure resembles the Chi-unrecognised complex and similarly does not contain ordered 3' ssDNA beyond the RecB helicase.

Model building and refinement

The structure of RecBCD in a complex with a DNA fork (PDB: 5LD2) (16) was used as a starting model for global docking in Chimera (40) for the Chi-recognised, Chi-unrecognised, Chi-minus 2, and Chi-plus 2 complexes. Once finalised, the Chi-recognised model was then used as a starting template for the Chi-intermediate complex. The Chi-unrecognised model was used a starting template for the Chi minus complex. Each of the six models was edited initially by jelly-body refinement with Refmac (41) in CCP-EM (42) followed by cycles of manual rebuilding in Coot (43) and real-space refinement with PHENIX (44). A 3.6 Å resolution (0.143 FSC cutoff, RELION3) map from a masked focused refinement of the Chi region for the 127,268 Chi-bound particles was used to help build the novel Chi bases and surrounding protein contacts. The six models were finally checked for consistency with one another, apart from where obvious density differences occurred in the maps. A final run of PHENIX real-space refinement with group B-factor refinement was run to generate the final models and model statistics (Tables 1 & 2).

DATA AVAILIBILITY

The cryo-electron density maps have been deposited at the Electron Microscopy Databank and the coordinates of the final refined models have been deposited at the Protein Data Bank with the codes listed in Tables 1 & 2.

Figure 1

a

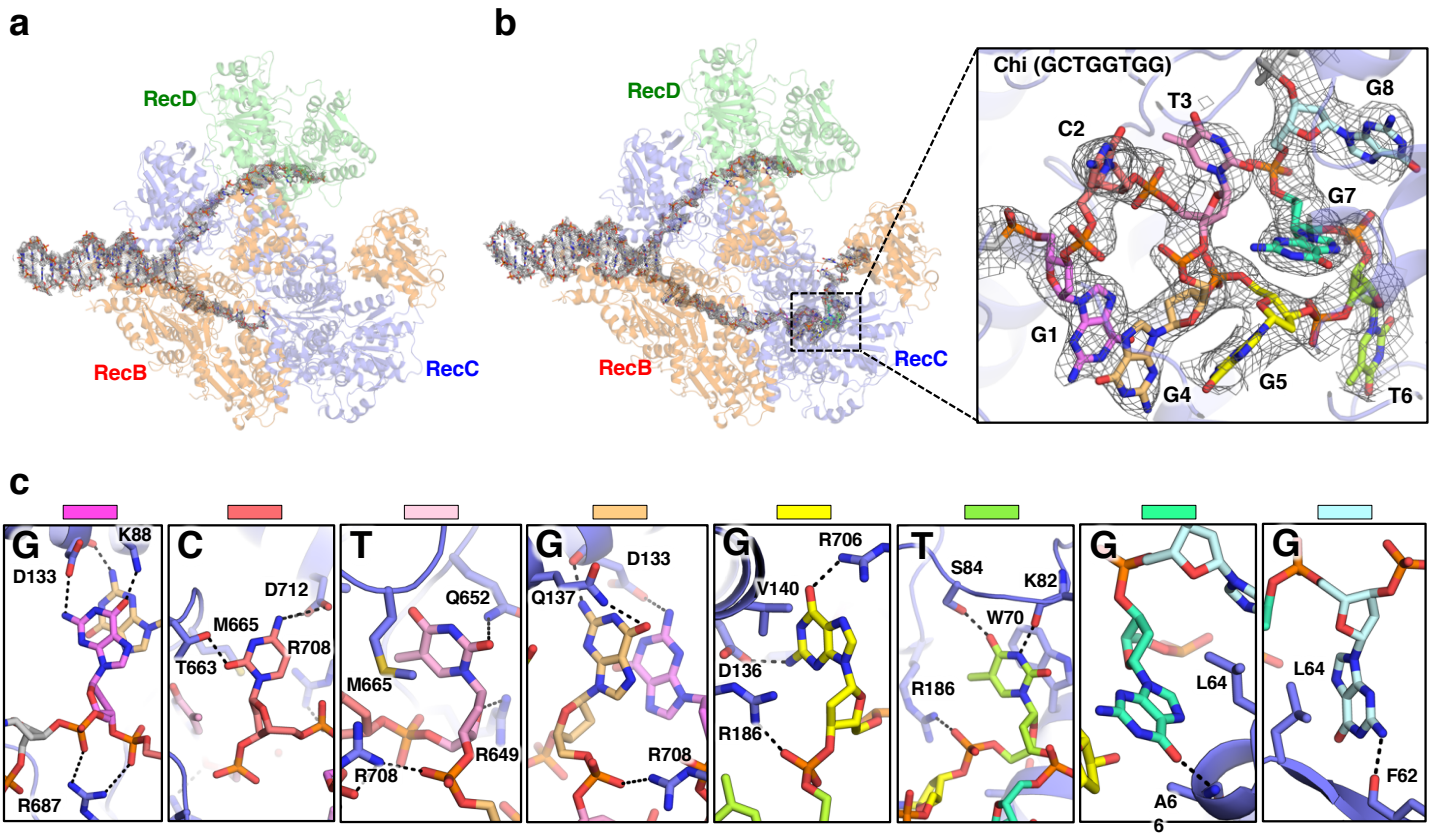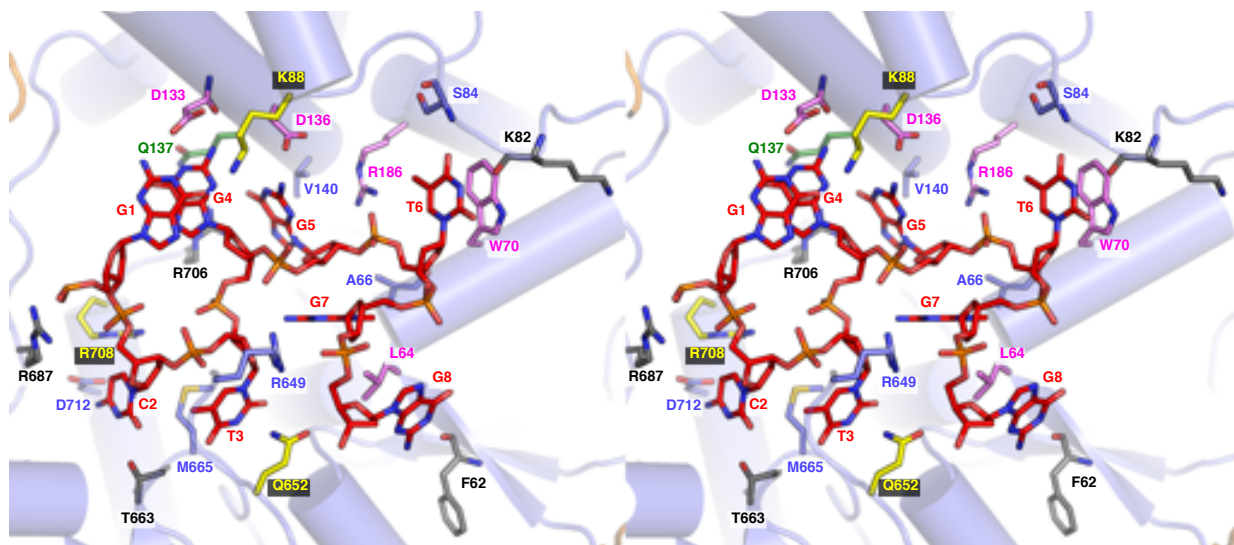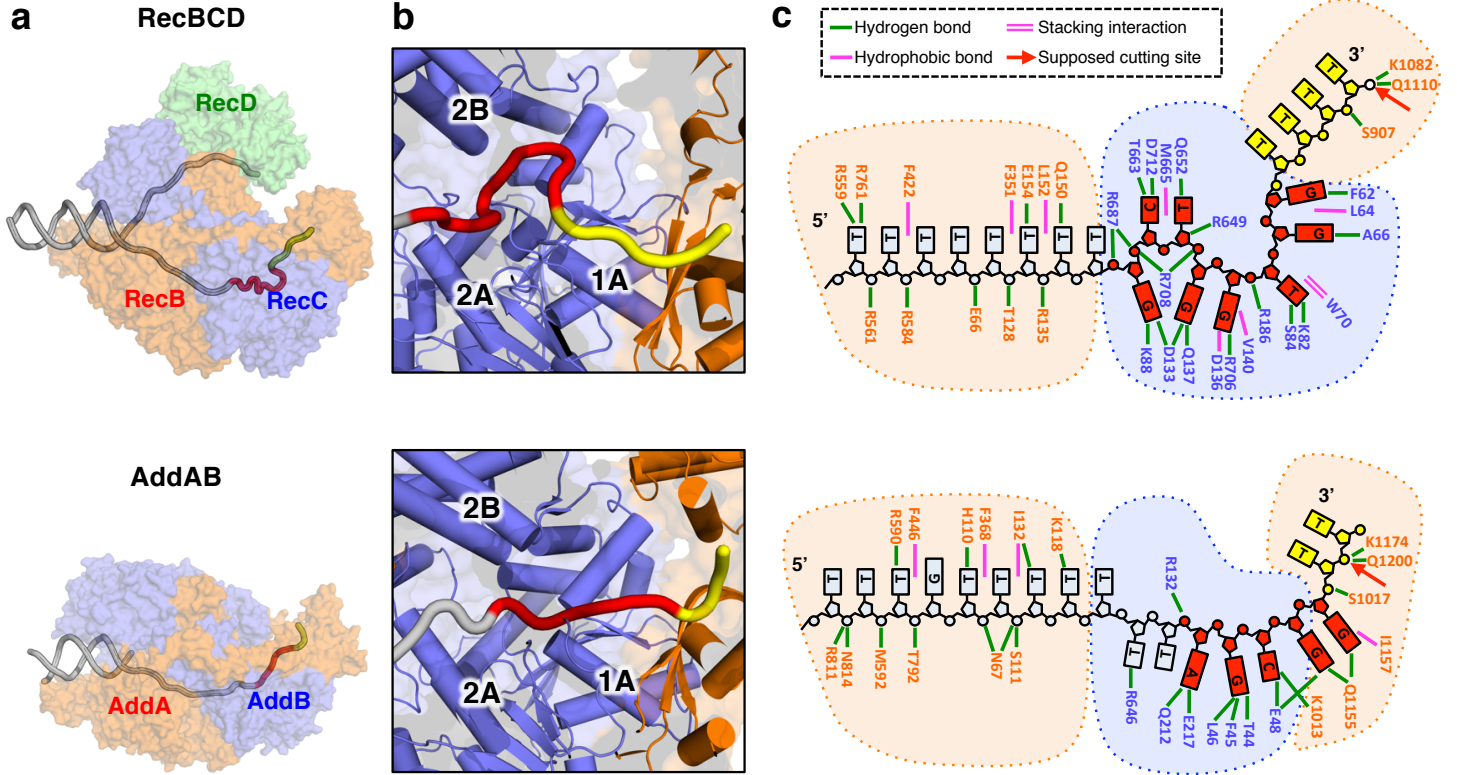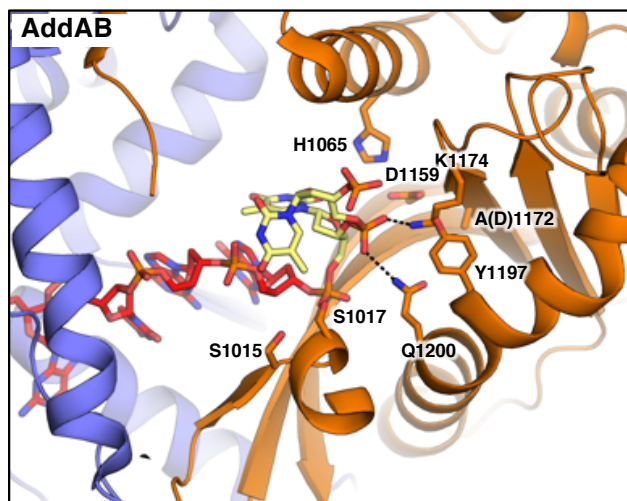Cy5

duplex = 25bp

5 tail

Chi (GCTGGTGG)

Y   Z

3 tail

5 tail	15	15	15	15	15	15	15	15	15	15	15
Y_nt	-	1	2	3	4	5	6	7	8	9	10
Z_nt	-	11	10	9	8	7	6	5	4	3	2
3 tail	20	20	20	20	20	20	20	20	20	20	20
Chi	-	+	+	+	+	+	+	+	+	+	+

b

RecD

RecB   RecC

c

Figure 2

Figure 3

Figure 4

Figure 5

Supplementary Figure 1

a 3,721 micrographs

autopick

1,053,451 particles

2D classification

373,476 particles

3D classification

54.8 %

204,633 particles, full complex

4.4 %

16,432 particles, no DNA bound

15.1 %

56,505 particles, no RecBnuc density, no Chi bound density

25.8 %

95,906 particles, no RecD motor domains density, no Chi bound density

Refinement, polishing and 3D classification (Mask around Chi)

62.2 %

127,268 particles

Chi

No Chi

37.8 %

77,365 particles

Refinement and 3D classification (Mask of RecBnuc+RecD)

58.5 %

74,496 particles

34.8 %

44,320 particles

6.6 %

8,452 particles

96.0 %

74,273 particles

4.0 %

3,092 particles

3.0 Å
4.0 Å
5.0 Å
6.0 Å
≥7.0 Å

Chi-recognised complex (3.7 Å)

Chi-intermediate complex (4.1 Å)

Chi-unrecognised complex (3.9 Å)

b

c

Chi-recognised (3.7 Å)
Chi-intermediate (4.1 Å)
Chi-unrecognised (3.9 Å)

0.143
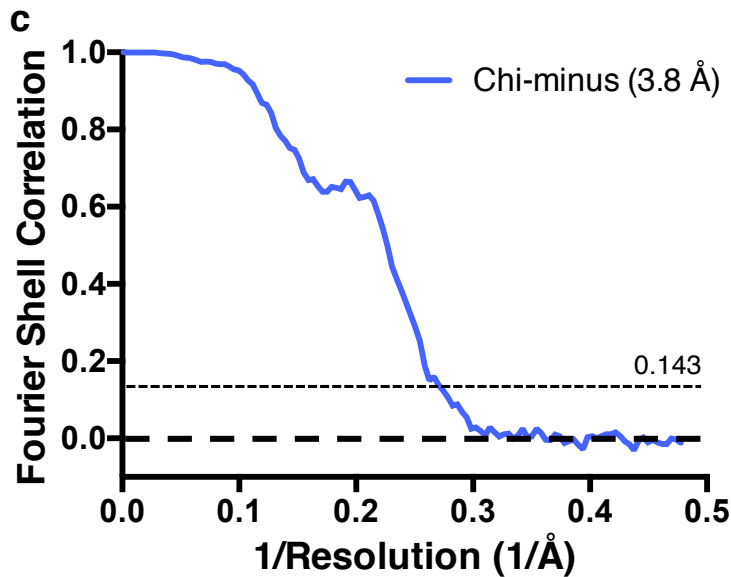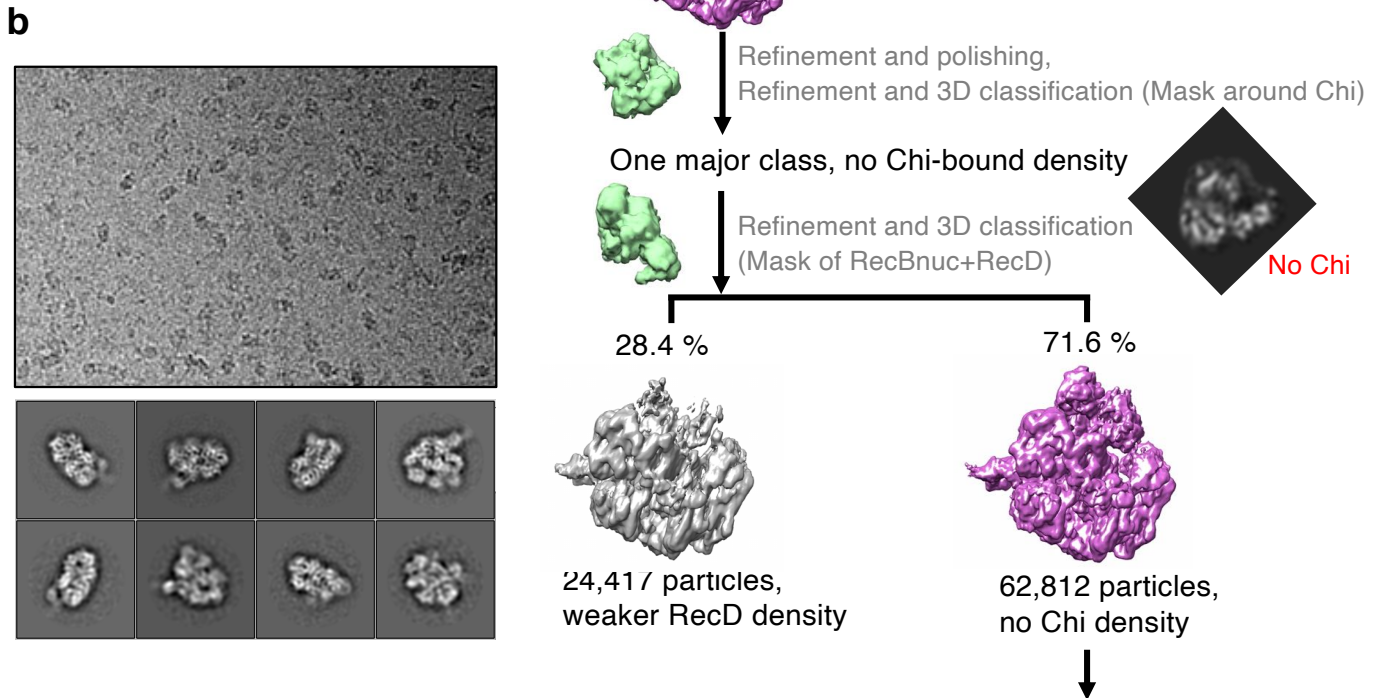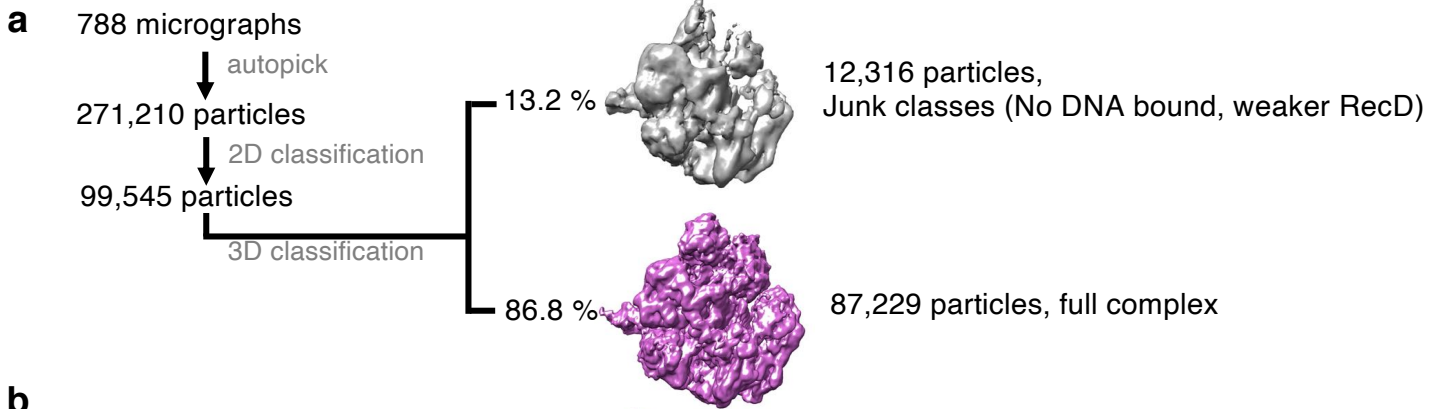
Supplementary Figure 1: CryoEM processing flow chart (Chi-containing substrate).
a, Scheme overview of the cryoEM processing steps for RecBCD with the Chi-containing substrate. Masks used during 3D classifications are shown in green. To display the separation of particles with ordered and disordered Chi sequences, 2D central slices are displayed of each respective masked 3D model. Final maps are coloured by local resolution as estimated by ResMap40. The scale bar shows the colour scale with resolution in Å. **b,** Upper, an example of a typical micrograph showing particle distribution. Lower, a selection of 2D class averages representing different particle views. **c,** 'Gold standard' FSC curve showing estimated global resolution of the three structures.
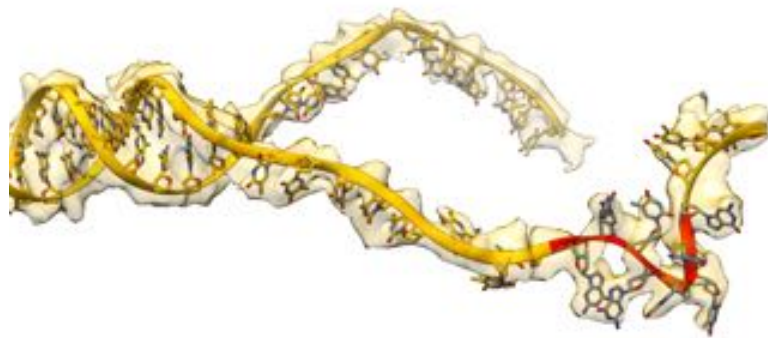
Supplementary Figure 2

a

788 micrographs
↓ autopick
271,210 particles
↓ 2D classification
99,545 particles
3D classification

13.2 % → 12,316 particles,
Junk classes (No DNA bound, weaker RecD)

86.8 % → 87,229 particles, full complex

↓ Refinement and polishing,
Refinement and 3D classification (Mask around Chi)

One major class, no Chi-bound density

No Chi

↓ Refinement and 3D classification
(Mask of RecBnuc+RecD)

28.4 %

24,417 particles,
weaker RecD density

71.6 %

62,812 particles,
no Chi density

b

c

Chi-minus (3.8 Å)

0.143

Chi-minus complex (3.8 Å)

3.0 Å
4.0 Å
5.0 Å
6.0 Å
≥7.0 Å

Supplementary Figure 2: **CryoEM processing flow chart (Chi-minus substrate).**
**a,** Scheme overview of the cryoEM processing steps for the RecBCD with the Chi-minus (20 base T residue 3'-tail) control substrate. Masks used during 3D classifications are shown in green. Classification with a mask around Chi showed one class with no Chi density as indicated in the displayed 2D central slice of the masked 3D model. The final map is coloured by local resolution as estimated by ResMap40. The scale bar shows the colour scale with resolution in Å. **b,** Upper, an example of a typical micrograph showing particle distribution. Lower, a selection of 2D class averages representing different particle views. **c,** 'Gold standard' FSC curve showing estimated global resolution of the structure.

Supplementary Figure 3

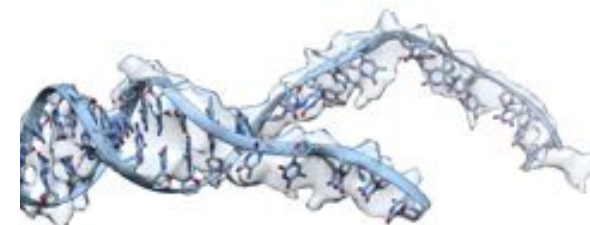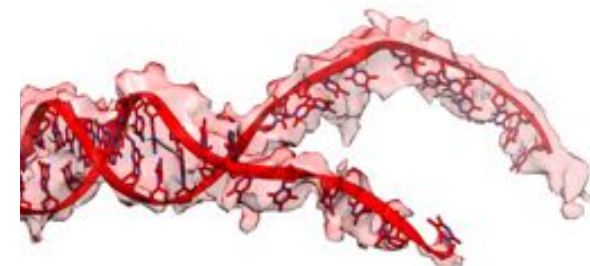**Chi-recognised complex**

Chi-intermediate complex
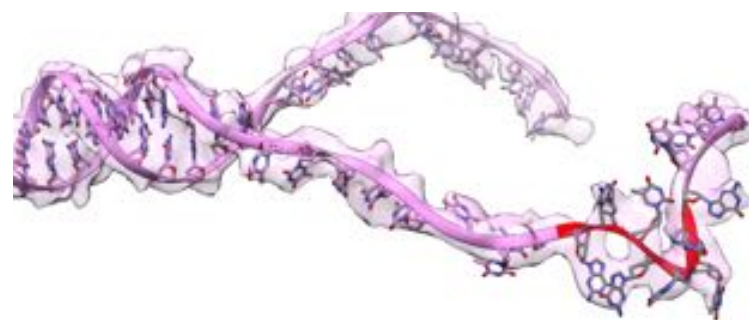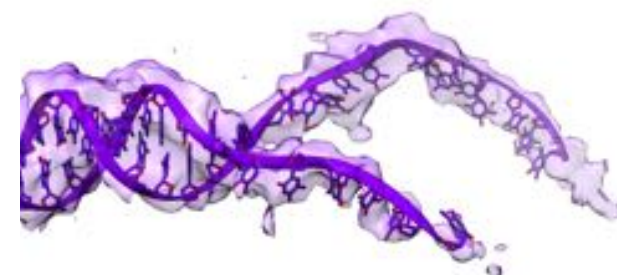

Chi-unrecognised complex
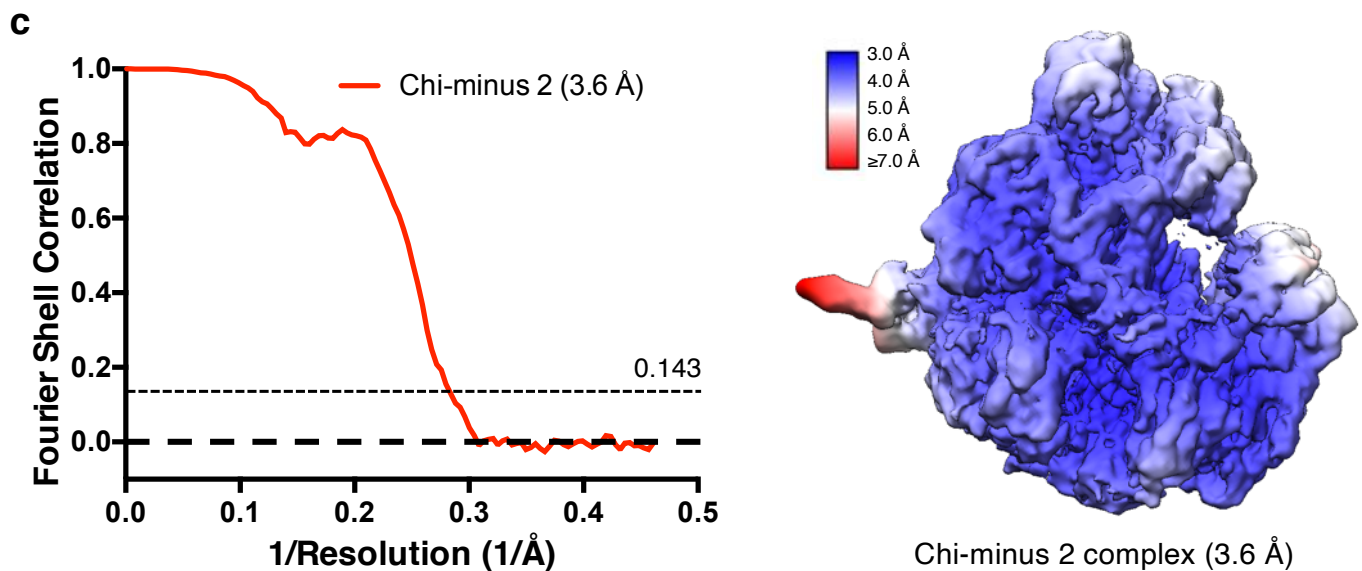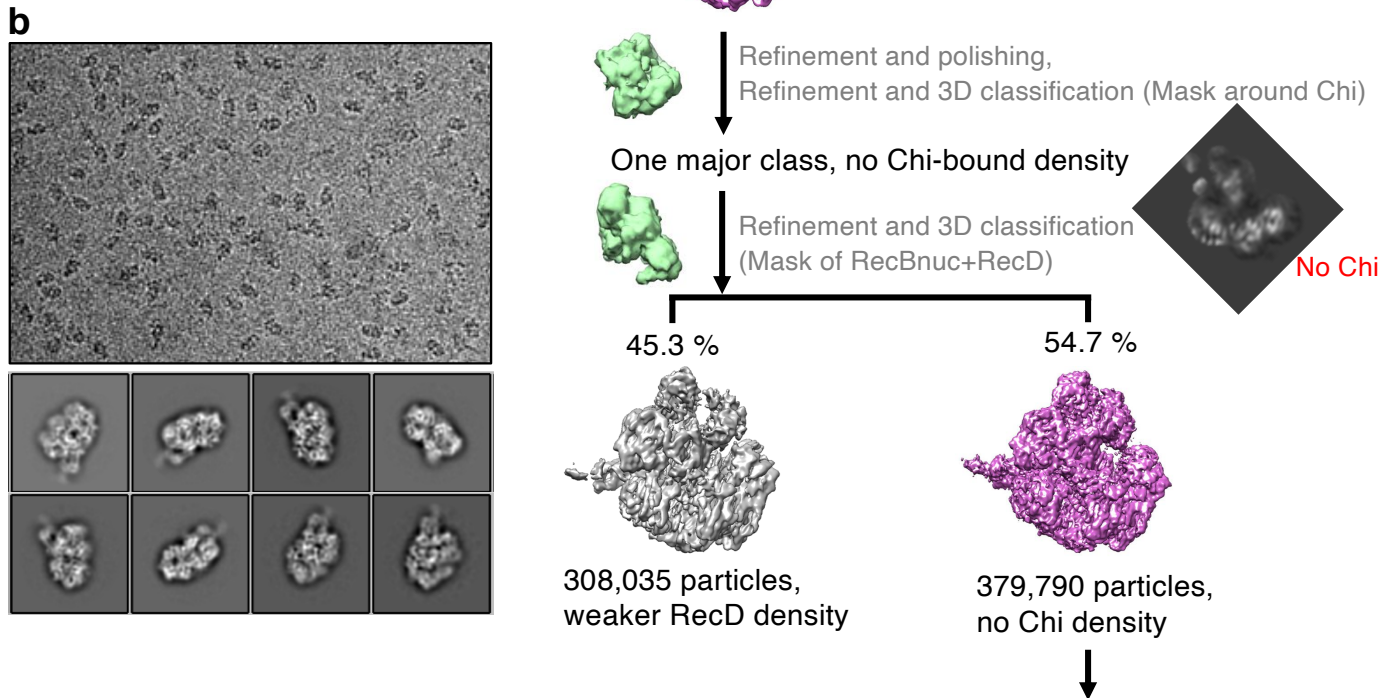

Chi-minus complex

Chi-minus 2 complex

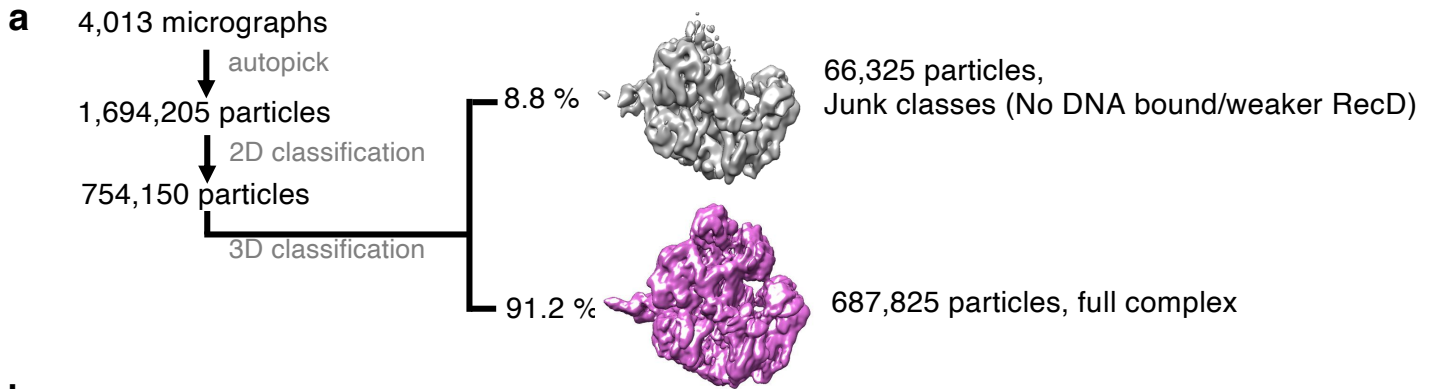**Chi-plus 2 complex**

Supplementary Figure 3: DNA density of the six structures.
The DNA density of the Chi-recognised, Chi-intermediate, Chi-unrecognised, Chi-minus, Chi-minus 2 and Chi-plus 2 complexes are shown as transparent surfaces with the built models in cartoon and sticks. The Chi sequence, where present, is coloured in red. All the maps are contoured to same level in Chimera (0.008).

Supplementary Figure 4

a

4,013 micrographs

↓ autopick

1,694,205 particles

↓ 2D classification

754,150 particles

3D classification

8.8 % 66,325 particles, Junk classes (No DNA bound/weaker RecD)

91.2 % 687,825 particles, full complex

↓ Refinement and polishing, Refinement and 3D classification (Mask around Chi)

One major class, no Chi-bound density

No Chi

↓ Refinement and 3D classification (Mask of RecBnuc+RecD)

45.3 % 308,035 particles, weaker RecD density

54.7 % 379,790 particles, no Chi density

↓

b

c

Chi-minus 2 (3.6 Å)

0.143

3.0 Å
4.0 Å
5.0 Å
6.0 Å
≥7.0 Å

Chi-minus 2 complex (3.6 Å)
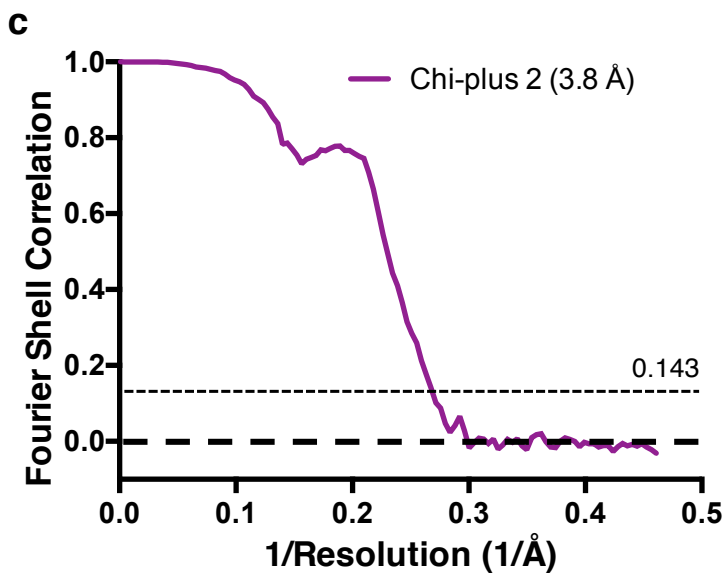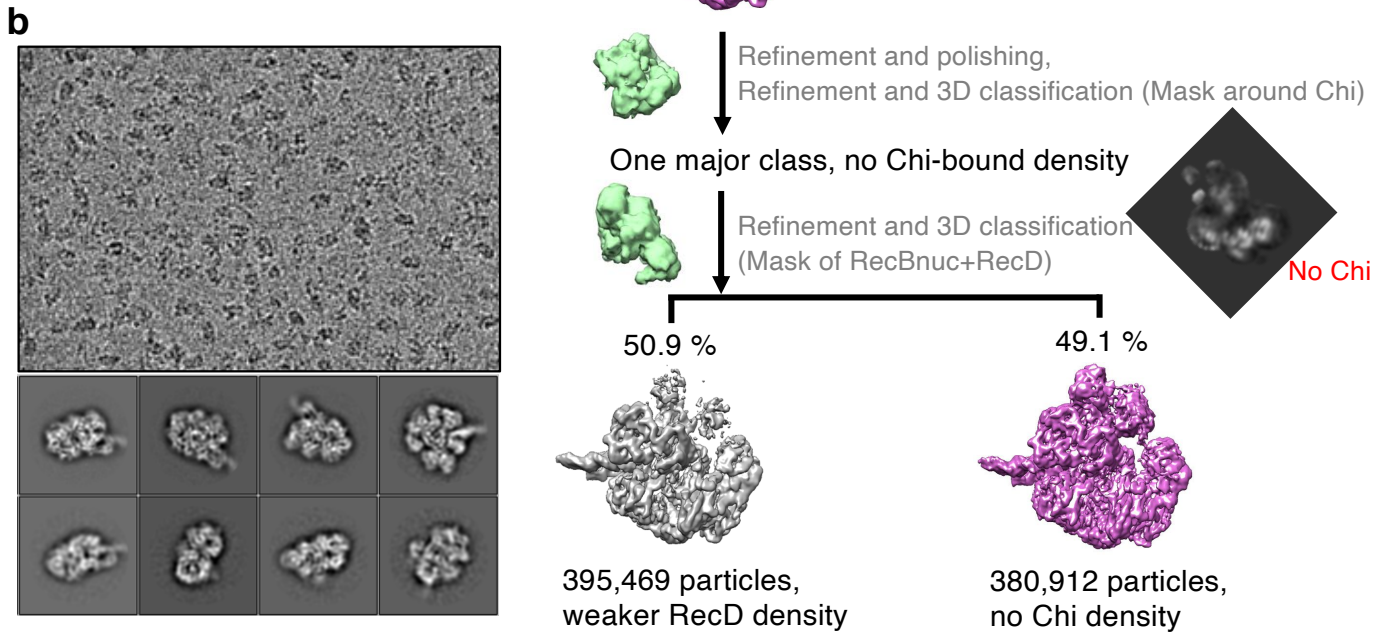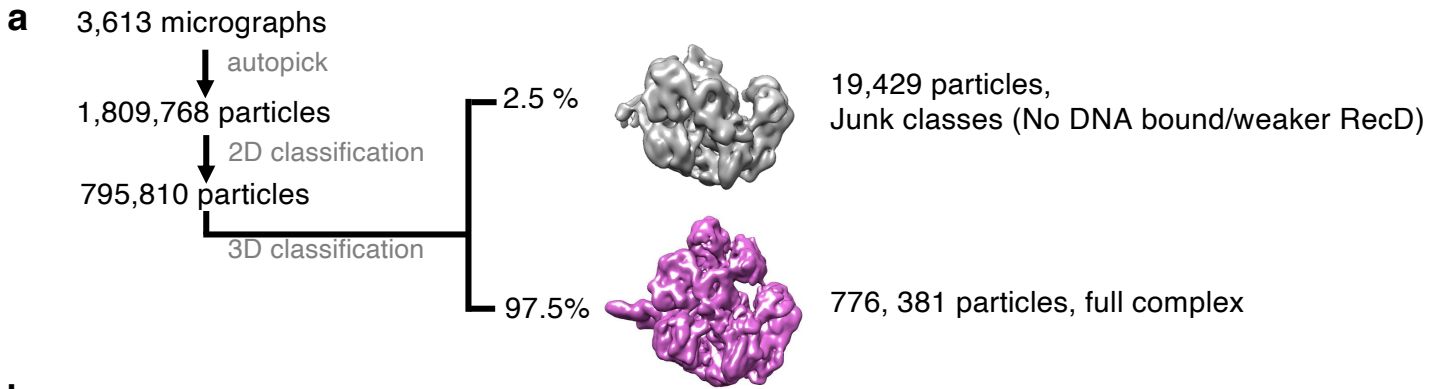
Supplementary Figure 4: CryoEM processing flow chart (minus 2 substrate).
a, Scheme overview of the cryoEM processing steps for the RecBCD with the Chi-minus 2 substrate (a 3'-tail contained Chi with 6 base T before Chi and 4 base T after Chi, the rest are the same as Chi-containing or Chi-minus substrates). Masks used during 3D classifications are shown in green. Classification with a mask around Chi showed one class with no Chi density as indicated in the displayed 2D central slice of the masked 3D model. The final map is coloured by local resolution as estimated by ResMap40. The scale bar shows the colour scale with resolution in Å. **b,** Upper, an example of a typical micrograph showing particle distribution. Lower, a selection of 2D class averages representing different particle views. **c,** 'Gold standard' FSC curve showing estimated global resolution of the structure.

Supplementary Figure 5

a

3,613 micrographs

↓ autopick

1,809,768 particles

↓ 2D classification

795,810 particles

3D classification

2.5 %

 19,429 particles,
Junk classes (No DNA bound/weaker RecD)

97.5%

 776, 381 particles, full complex

↓ Refinement and polishing,
Refinement and 3D classification (Mask around Chi)

One major class, no Chi-bound density

↓ Refinement and 3D classification
(Mask of RecBnuc+RecD)

No Chi

50.9 %

 395,469 particles,
weaker RecD density

49.1 %

 380,912 particles,
no Chi density

b

c

Chi-plus 2 (3.8 Å)

0.143

3.0 Å
4.0 Å
5.0 Å
6.0 Å
≥7.0 Å

Chi-plus 2 complex (3.8 Å)

Supplementary Figure 5: CryoEM processing flow chart (plus 2 substrate).
a, Scheme overview of the cryoEM processing steps for the RecBCD with the Chi-plus 2 substrate (a 3'-tail contained Chi with 10 base T before Chi and 4 base T after Chi, the rest are the same as Chi-containing or Chi-minus substrates). Masks used during 3D classifications are shown in green. Classification with a mask around Chi showed one class with no Chi density as indicated in the displayed 2D central slice of the masked 3D model. The final map is coloured by local resolution as estimated by ResMap40. The scale bar shows the colour scale with resolution in Å. **b,** Upper, an example of a typical micrograph showing particle distribution. Lower, a selection of 2D class averages representing different particle views. **c,** 'Gold standard' FSC curve showing estimated global resolution of the structure.
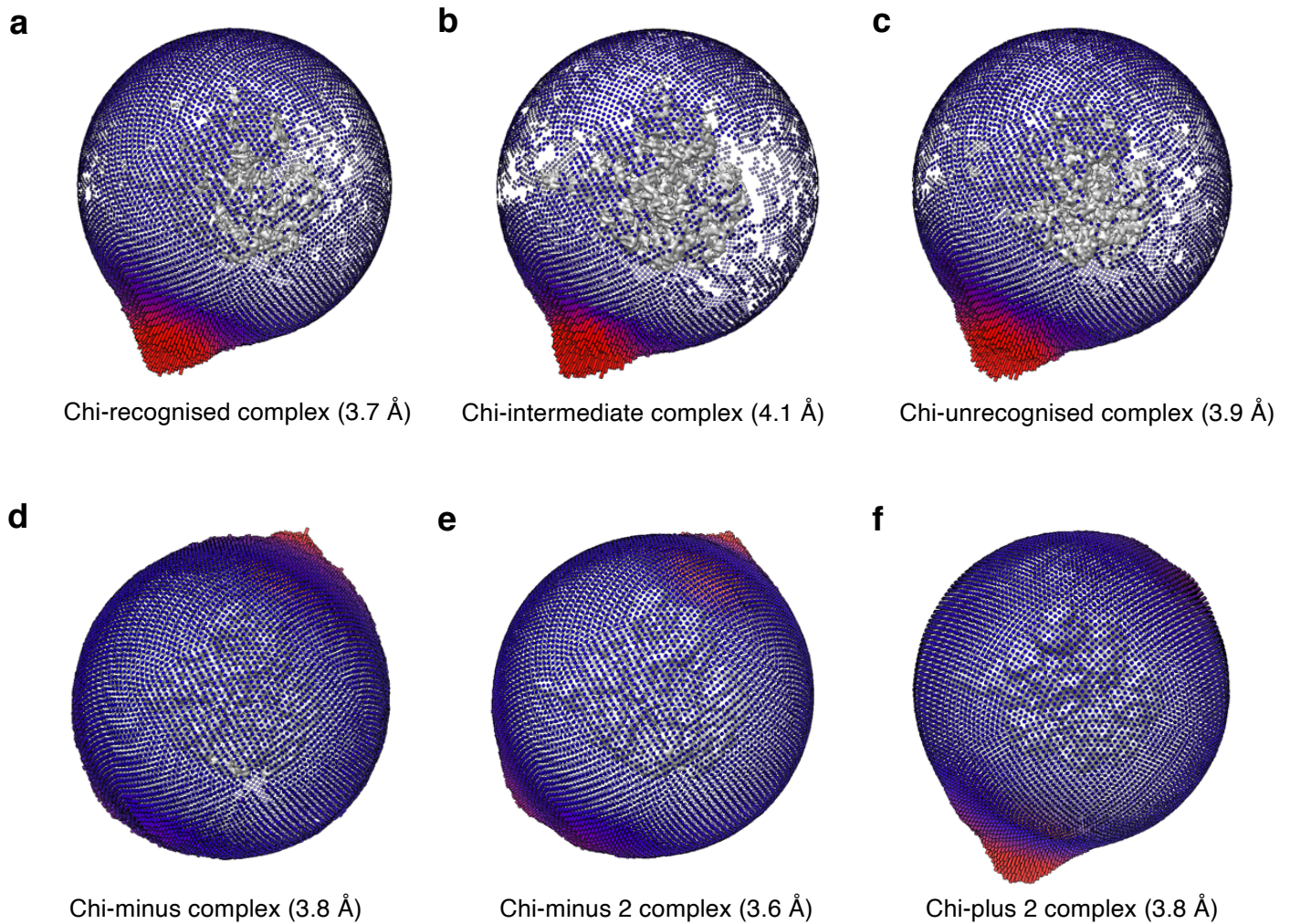
Supplementary Figure 6

a

Chi-recognised complex (3.7 Å)

b

Chi-intermediate complex (4.1 Å)

c

Chi-unrecognised complex (3.9 Å)

d

Chi-minus complex (3.8 Å)

e

Chi-minus 2 complex (3.6 Å)

f

Chi-plus 2 complex (3.8 Å)

Supplementary Figure 6: Euler angle distribution of the cryo-EM reconstructions .

Euler angle distribution of the Chi-recognised complex (a), Chi-intermediate complex (b), Chi-unrecognised complex (c), Chi-minus complex (d), Chi-minus 2 complex (e) and Chi-plus 2 complex (f).