BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 20 2010, pages 2629–2630
doi:10.1093/bioinformatics/btq491

Structural bioinformatics

Advance Access publication August 24, 2010

Structural genomics of histone tail recognition

Minghua Wang[1,†], Man Wai Mok[2,†], Hong Harper[1], Wen Hwa Lee[1], Jinrong Min[2], Stefan Knapp[1], Udo Oppermann[1,3], Brian Marsden[1,4] and Matthieu Schapira[2,5,*]

[1]Structural Genomics Consortium, University of Oxford, Headington, Oxford OX37DQ, UK, [2]Structural Genomics Consortium, University of Toronto, Toronto, ON M5G1L7, Canada, [3]Oxford NIHR Biomedical Research Unit, Botnar Research Centre, [4]Nuffield Department of Clinical Medicine, Old Road Campus, University of Oxford, Headington, Oxford, OX3 7DQ, UK and [5]Department of Pharmacology and Toxicology, University of Toronto, ON M5S1A8, Canada

Associate Editor: Anna Tramontano

ABSTRACT

Summary: The structural genomics of histone tail recognition web server is an open access resource that presents within mini articles all publicly available experimental structures of histone tails in complex with human proteins. Each article is composed of interactive 3D slides that dissect the structural mechanism underlying the recognition of specific sequences and histone marks. A concise text html-linked to interactive graphics guides the reader through the main features of the interaction. This resource can be used to analyze and compare binding modes across multiple histone recognition modules, to evaluate the chemical tractability of binding sites involved in epigenetic signaling and design small molecule inhibitors.

Availability: http://www.thesgc.org/resources/histone_tails/

Contact: matthieu.schapira@utoronto.ca

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 7, 2010; revised on August 6, 2010; accepted on August 18, 2010

1 INTRODUCTION

Post-translational modifications of histone proteins regulate chromatin compaction, mediate epigenetic regulation of transcription and control cellular differentiation in health and disease (Grewal and Moazed, 2003). A variety of epigenetic marks are deposited, removed and sensed individually and in combinations by so-called 'writers, erasers and readers' of the 'histone code' (Jenuwein and Allis, 2001; Kouzarides, 2007; Strahl and Allis, 2000). Most post-translational modifications occur at lysine residues on the N-terminal unstructured portion of histones.

The structural mechanisms underlying specific recognition of histone tail sequences and post-translational modifications are central to the emerging concept of chromatin as a complex signaling platform where diverse epigenetic marks can make or break circuits in a synergistic or antagonistic manner (Fischle *et al.*, 2003a, b; Garske *et al.*, 2010; Latham and Dent, 2007). Additionally, the binding mode of individual histone tail side chains to catalytic domains or effector modules can be used as a guide to design competitive inhibitors (Culhane *et al.*, 2010; Liu *et al.*, 2009).

We have developed the structural genomics of histone tail recognition web server. This resource allows the user to browse within mini articles high-quality graphics and interactive 3D slides of all available crystal and nuclear magnetic resonance (NMR) structures of human proteins in complex with histone tail peptides. Each article highlights the main features of the structures, and the recognition mechanism at atomic resolution. The intuitive and easy-to-use interface has been designed to facilitate the analysis of structural features, even if the user does not have any previous experience in molecular graphics programs. It is hoped that this resource, updated on a monthly basis, will prove itself useful to decode the structural determinants of histone recognition and design chemical inhibitors targeting the writers, readers and erasers of histone marks.

2 RESULTS

The server contains, as of May 2010, structures of 47 histone–protein complexes organized along schematic representations of histone H3, H4 and H2A.x tails. The name of each gene is linked to its corresponding set of 3D slides, and color-coded based on the nature of the histone binding domain. Twelve structure classes are represented: histone methyltransferases (Cheng *et al.*, 2005; Copeland *et al.*, 2009), lysine demethylases (Marmorstein and Trievel, 2009), bromodomains (Mujtaba *et al.*, 2007; Taverna *et al.*, 2007), histone acetyltransferases (Berndsen and Denu, 2008; Marmorstein and Trievel, 2009), six readers of methylated lysine (MBTs, Tudor, PHDs, Chromo, PWWP and 'Other' domains) (Adams-Cioaba and Min, 2009; Taverna *et al.*, 2007) and the two readers of phosphorylation marks 14-3-3 and BRCT (Macdonald *et al.*, 2005; Stucki *et al.*, 2005).

All 3D articles follow the same template. The opening view shows a molecular surface of the protein with electrostatic potential coloring and bound histone. Successive bullet points in the mini-article, linked to interactive 3D slides, guide the reader through the main features of histone recognition as follows. (i) Overall topology of the structure, which highlights the tertiary arrangement of the molecule, and the relative orientation of multiple domains when applicable. This view is useful, for instance, to see how the histone H3 tail is sandwiched between the JMJ and PHD domains of PHF8 (Horton *et al.*, 2010). (ii) Hydrogen bonds between the histone

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the First two authors should be regarded as joint First Authors.

peptide and its recognition platform. This view highlights with atomic resolution interactions that typically act as determinant for specificity. Water molecules mediating interactions are also shown. Residue numbers are accessible by a single click, which can be useful to scientists wishing to follow-up with targeted mutagenesis. (iii) Molecular surface of the binding site with bound histone tail. This view reveals side chains that contribute the most to binding, and may be used as a guide to design selective inhibitors, as illustrated in the structure browser of the methyltransferase GLP. 3D slides clearly show how (a) two inhibitors occupy the histone binding site (b) both inhibitors mimic the interaction of Arg 8 of histone 3 (H3R8) with the receptor and (c) the most potent inhibitor also occupies the substrate lysine (H3K9) binding channel (Chang *et al.*, 2009; Liu *et al.*, 2009). (iv) Electrostatics of the entire structure, which provides a more realistic rendition of the overall complex.

Ad hoc bullet points are often complementing this basic template to present features reported in published articles. Finally, a color-coded amino acid sequence of the protein highlights residues hydrogen bonded to the histone tail.

3 IMPLEMENTATION

Structural analyses are carried out—and 3D graphics generated—with the software ICM (Molsoft LLC). The activeICM and iSee technology, recently developed for publishing 3D content, is used to illustrate the web version of the articles (Abagyan *et al.*, 2006; Raush *et al.*, 2009). On their first visit, users are asked to download a free, cross-platform, browser-independent plug-in, necessary to embed interactive graphics within a web browser. Articles are divided into two panes, one for text and one for graphics. HTML links in the text window trigger 3D slides in the graphics window, allowing rotation and zooming (Supplementary Fig. 1). For any structure, users can also choose to visualize histograms showing the number of hydrogen bonds engaged with side- and main chain of each histone residue.

The Protein Databank is scanned on a monthly basis to identify novel structures of human proteins in complex with histone tails. Updates are completed within 4 weeks, which ensures that the resource is not outdated by more than 2 months.

Funding: This work is supported by the Structural Genomics Consortium, a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Karolinska Institutet, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck & Co., Inc., the Novartis Research Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research and the Wellcome Trust. The study was also supported by the Oxford NIHR Biomedical Research Unit and the Nuffield Department of Clinical Medicine. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the article.

Conflict of Interest: none declared.

REFERENCES

Abagyan,R. *et al.* (2006) Disseminating structural genomics data to the public: from a data dump to an animated story. *Trends Biochem. Sci.*, **31**, 76–78.

Adams-Cioaba,M.A. and Min,J. (2009) Structure and function of histone methylation binding proteins. *Biochem. Cell Biol.*, **87**, 93–105.

Berndsen,C.E. and Denu,J.M. (2008) Catalysis and substrate selection by histone/protein lysine acetyltransferases. *Curr. Opin. Struct. Biol.*, **18**, 682–689.

Chang,Y. *et al.* (2009) Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. *Nat. Struct. Mol. Biol.*, **16**, 312–317.

Cheng,X. *et al.* (2005) Structural and sequence motifs of protein (histone) methylation enzymes. *Annu. Rev. Biophys. Biomol. Struct.*, **34**, 267–294.

Copeland,R.A. *et al.* (2009) Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug. Discov.*, **8**, 724–732.

Culhane,J.C. *et al.* (2010) Comparative analysis of small molecules and histone substrate analogues as LSD1 lysine demethylase inhibitors. *J. Am. Chem. Soc.*, **132**, 3164–3176.

Fischle,W. *et al.* (2003a) Binary switches and modification cassettes in histone biology and beyond. *Nature*, **425**, 475–479.

Fischle,W. *et al.* (2003b) Histone and chromatin cross-talk. *Curr. Opin. Cell. Biol.*, **15**, 172–183.

Garske,A.L. *et al.* (2010) Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nat. Chem. Biol.*, **6**, 283–290.

Grewal,S.I. and Moazed,D. (2003) Heterochromatin and epigenetic control of gene expression. *Science*, **301**, 798–802.

Horton,J.R. *et al.* (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.*, **17**, 38–43.

Jenuwein,T. and Allis,C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.

Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.

Latham,J.A. and Dent,S.Y. (2007) Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.*, **14**, 1017–1024.

Liu,F. *et al.* (2009) Discovery of a 2,4-diamino-7-aminoalkoxyquinazoline as a potent and selective inhibitor of histone lysine methyltransferase G9a. *J. Med. Chem.*, **52**, 7950–7953.

Macdonald,N. *et al.* (2005) Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. *Mol. Cell.*, **20**, 199–211.

Marmorstein,R. and Trievel,R.C. (2009) Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim. Biophys. Acta*, **1789**, 58–68.

Mujtaba,S. *et al.* (2007) Structure and acetyl-lysine recognition of the bromodomain. *Oncogene*, **26**, 5521–5527.

Raush,E. *et al.* (2009) A new method for publishing three-dimensional content. *PLoS ONE*, **4**, e7394.

Strahl,B.D. and Allis,C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41–45.

Stucki,M. *et al.* (2005) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell*, **123**, 1213–1226.

Taverna,S.D. *et al.* (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.*, **14**, 1025–1040.