Published in final edited form as:

Curr Proteomics. 2014 ; 11(3): 151–160. doi:10.2174/157016461103140922163444.

B-factor Analysis and Conformational Rearrangement of Aldose Reductase

Ganesaratnam K. Balendiran^{1,*}, J. Rajendran Pandian², Evin Drake¹, Anubhav Vinayak¹, Malkhey Verma³, and Duilio Cascio⁴

¹Department of Chemistry, WBSH 6017, Youngstown State University, One University Plaza, Youngstown, OH 44555

²Department of Business Administration, College of Business Administration, Gulf University for Science and Technology, Mubarak Al-Abdullah Area/West Mishref, Kuwait

³Manchester Institute of Biotechnology, 131 Princess Street, The University of Manchester, Manchester, M1 7DN, UK

⁴UCLA-DOE, 611 Charles E. Young Drive East, 220 Boyer Hall, Los Angeles, CA 90095, USA

Abstract

The NADPH-dependent reduction of glucose reaction that is catalyzed by Aldose Reductase (AR) follows a sequential ordered kinetic mechanism in which the co-factor NADPH binds to the enzyme prior to the aldehyde substrate. The kinetic/structural experiments have found a conformational change involving a hinge-like movement of a surface loop (residues 213-224) which is anticipated to take place upon the binding of the diphosphate moiety of NADPH. The reorientation of this loop, expected to permit the release of NADP⁺, represents the rate-limiting step of the catalytic mechanism. This study reveals: 1) The Translation/Libration/Screw (TLS) analysis of absolute B-factors of apo AR crystal structures indicates that the 212–224 loop might move as a rigid group. 2) Residues that make the flexible loop slide in the AR binary and ternary complexes. 3) The normalized B-factors separate this segment into three different clusters with fewer residues.

Keywords

Aldo-keto reductase; B-factor; clustering; crystal structure; statistical analysis; structural dynamics; TLS

©2014 Bentham Science Publishers

*Address correspondence to this author at the Department of Chemistry, WBSH 6017, Youngstown State University, One University Plaza, Youngstown, OH 44555, USA; Tel: 1-330-941-7103; Fax: 1-330-941-7103; pl_note@yahoo.com or gkbalendiran@ysu.edu.

Send Orders for Reprints to reprints@benthamscience.net

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

INTRODUCTION

The first enzyme of the polyol pathway is Aldose reductase (AR) which catalyses the reduction of glucose to sorbitol. Sorbitol dehydrogenase (SDH), the second enzyme of this pathway, has a lower flux than AR, causing an escalation of sorbitol within the cell [1, 2]. The sorbitol accumulation implicates AR in the pathogenesis of most diabetic complications.

In the forward reaction, AR follows a sequential ordered kinetic mechanism in which the cofactor, NADPH, binds prior to the aldehyde substrate and NADP^+ is subsequently discharged after the release of the alcohol product [3, 4]. A kinetically detectable isomerization was suggested to occur upon the binding of NADPH ($\text{E-NADPH} \rightarrow \text{E}^*\text{-NADPH}$) and during the release of the oxidized cofactor product NADP^+ [4–6]. The reorientation of this loop during the $\text{E}^*\text{-NADPH} \rightarrow \text{-NADP}^+$ step appears to be the rate-limiting step in the catalytic mechanism of the forward reaction [3, 4, 7–9]. As a result, under steady state conditions, most of the enzyme is expected to be present predominantly as an enzyme-nucleotide binary complex [7]. High resolution crystal structures are available in the apo, NADP^+ bound binary, and NADP^+ -inhibitor complex ternary forms for AR family members in several space groups. Comparison of these crystal structures reveals that there is a significant difference in the conformation attained by the loop region corresponding to the residues 212–224 in the AR family of proteins in which some of the residues have shifted over 5 Å in certain entries.

Atomic displacement [10], B –, Debye Waller or temperature factors which are evaluated for each atom during a crystal structure determination reflects the degree of thermal mobility and static disorder [11] of an atom in a crystal structure. Numerous attempts are being made to find the nonmathematical meaning of this crystallographically introduced B -factor over the years [12–34]. A B -factor for nonhydrogen atom j , B_j , is proportional to the mean square displacement of the atom from its equilibrium position: where $B_j = 8 \pi^2 \langle x_j^2 \rangle$. The term $\langle x_j^2 \rangle = \langle x_j^2 \rangle_{\text{TLS}} + \langle x_j^2 \rangle_{\text{V}} + \langle x_j^2 \rangle_{\text{CS}}$, where $\langle x_j^2 \rangle_{\text{TLS}}$ is from translation, rotation and screw component; $\langle x_j^2 \rangle_{\text{V}}$ is caused by vibrations faster than 0.1 ps and $\langle x_j^2 \rangle_{\text{CS}}$ is due to different conformational substates [35]. The term $\langle x_j^2 \rangle_{\text{TLS}}$ may come from motions of the whole molecule or from static displacements of the atoms/molecule in the crystal lattice [35]. Since the reported B -factors in several crystal structures are in different scales, the normalized version rather than the absolute (not normalized) form of them has been employed in the comparison [36–38]. Here, we attempt to explore the application of TLS and normalized B -factor analysis to determine the segments in the crystal structures of AR that are anticipated to undergo molecular motion/dynamics during catalysis. Also, we have extended the normalized B -factor based comparison across the AR family of proteins. Furthermore, the residues are clustered based on the normalized B -factors of their $\text{C}\alpha$ atoms across 20 related structures using statistical analysis.

EXPERIMENTAL SECTION

The B -factors used in our analysis originate from the crystal structures deposited in the Protein Data Bank (PDB) [39] for all the entries. A total of 20 coordinate sets of AR and its

analogs deposited in the PDB are utilized for this study (Table 1). The factors that are different about these structures are: use of distinct software for X-ray data collection, their structure determination/refinement, the crystal forms, resolution of the data used, experimental parameters, X-ray source, and type of complexes and sequence variation/modification. We verified all the PDB entries utilized in this study and only 3Q65 and 3Q67 used TLS group restraints in the crystallographic refinement and do not bias our results reported here.

TLS Analysis of AR

TLS method [40] extracts a linear combination of anisotropic translational (T), librational (L), and screw (S) motions from an analysis of the isotropic temperature factors (absolute form) of the atomic coordinates refined against X-ray diffraction data. The rigid-body motions of the peptide segments of human AR were analyzed by the TLS method using the coordinates for apo AR (1XGD, [41]) where, the 315 residue amino acid sequence was partitioned into 20 TLS groups containing consecutive residues in continuous chains using the web-based server, <http://skuld.bmsc.washington.edu/~tlsmd> [42, 43]. During this process, the calculated B-factors (absolute) per residue were compared with the experimentally observed values and their best-fit correlated values for the TLS groups were evaluated.

Sequence Based Analogs

A blast search using the amino acid sequence corresponding to 212-224 segment of the AR was performed. Each hit from the blast search was subjected to TLS analysis independently as described above for the AR apo enzyme.

Statistical Analysis of Normalized B-factors

Only the C α atoms of the entries shown in (Table 1) were included in the current study because Smith *et al.*, [38] have established that the profile of B-factors of C α is very similar to that obtained by averaging the B-factors over the backbone atoms and Micheletti *et al.*, [44] also demonstrated that the degrees of freedom of a side-chain are controlled by C α positions. In the case of hits containing atoms with partial occupancies, only the C α atoms with the highest occupancy were taken into consideration for the B-factor calculations. Since a few entries are missing B-factors for the first two or last four residues, these residues were omitted for all entries included in the study. The normalized B-factors were calculated according to the definition derived previously in literature [36, 45, 46]: $B\text{-normalized} = (B - \langle B \rangle) / \sigma(B)$, where B is the B-factor of each C α atom in the entry, $\langle B \rangle$ is the mean B-factor of all the C α atoms used in the entry and $\sigma(B)$ is the standard deviation.

To verify whether the mean values of peaks corresponding to different residue segments are distinct in B-factor, the probability (p) was estimated using the Student's t -test. The peaks were considered to be of similar height when p approached unity and were considered different when p approaches zero.

Statistical Analysis of B-factor and Clustering

The B-factor for each Ca atom in every entry is normalized as described above for the clustering studies. The Ward's [47] method is a hierarchical clustering technique which starts with one cluster of all the entries and moves one member which is different from the rest into the second cluster. This procedure is repeated until each member is in different cluster. At every step, the variance explained by the new cluster is tested in order to ensure that the difference between clusters is high and the variance between cluster members (intra-cluster difference) is low. The new cluster reported is based on the marginal increase in variance. Since the maximum explanatory power is when each member is a cluster by itself, the marginal explanatory power of the new cluster is used to identify the appropriate number of clusters.

In order to determine the changes in the location, the difference between the atomic coordinates of the same Ca atom was used. The root-mean-squared sum of differences (RMSD) is calculated as defined previously [48]. The mean squared deviation (MSD) = RMSD^2 ,

where $\text{MSD} = (\{\sum_{j=1 \text{ to } 20} \sum_{i=1 \text{ to } 310} [(x_{ji}-x_{j(i+1)})^2 + (y_{ji}-y_{j(i+1)})^2 + (z_{ji}-z_{j(i+1)})^2]\})/210$.

Here j is the number of entry, i is number of residues and coordinates are x_{ij} , y_{ij} and z_{ij} .

RESULTS

The B-factors can fluctuate significantly in proteins due to many reasons [10, 13, 49]. In particular, segments undergo movements on a much larger relative scale than the rest of the molecule in many proteins when B-factors are taken into consideration over the length of the polypeptide chain [36, 37].

Conformational Rearrangement and TLS Analysis During AR Catalysis

In this study, TLS formalism has allowed the visualization of the anisotropic sub-loop dynamics of AR and various complexes of the AR family, which are relevant for the understanding of the dynamics of cofactor binding process in the reaction pathway.

The TLS analysis of the apo AR (1XGD) shows that the average B-factors of the segments 212-224 and 22-29, which are the second highest and the highest among the 20 groups, respectively (Fig. 1). Kinetic studies [4–6] suggest that the former segment is predicted to undergo a conformational change during the binding of NADPH and release of NADP^+ while structures of binary complexes reveal 212-224 segment covers the mobile loop. Residues 210, 212, 214 and 216 in the segment 212-224 form polar interactions with NADP^+ . Similarly, residues 19-21 and 26 in the latter segment 22-29 also form molecular contacts with NADP^+ . Other residues that form molecular interactions with the cofactor are distributed over four additional TLS groups: 159-160 in 148-165; 183 and 209-210 in 181-211; 43 in 30-46; and 262-265, 268 and 271-272 in 242-273 segments.

The amino acid sequence corresponding to the segment 212-224 in 1XGD structure is LGSPDRPWAKPED. The blast search using this sequence resulted in the identification of

the following 23 PDB entries; 1Z3N, 2AGT, 1ZUA, 3Q67, 3LZ5, 3LQG, 3LD5, 3LBO, 2R24, 2PDX, 2PDQ, 2PDM, 2PDK, 2PDI, 2PDF, 2PDB, 2PD5, 1AH0, 1PWL, 2IS7, 1EKO, 2ACU and 1EF3. These entries are all wild-type or mutant human AR except one for AKR1B10 and two for pig AR (Table 1). Individually preformed TLS analysis of blast search hit entries 2PDQ, 2PD5, 2PDB and 2PDF show a highest mean B-factor of $35.4 \text{ \AA}^2$ for 299-304, $29.1 \text{ \AA}^2$ for 238-242, $24.7 \text{ \AA}^2$ for 234-239 and $29.5 \text{ \AA}^2$ for 119-126 segments that are different from the mobile loop (212-224) in their structures, respectively. Noticeably these entries (2PDQ, 2PD5, 2PDB, 2PDF) all have a number of discretely disordered residues and have employed anisotropic scaling in their structure refinement. Therefore, they were excluded from the B-factor analyses conducted here. Also, the TLS analysis could not be performed for PDB entries 1ABN and 1EF3 since the error on the standard deviation of temperature factors is less than 0.01 or greater than 60.0 for those residues. Evidently, these coordinates were refined with grouped B-factor constraints. The highest mean B-factor from TLS based regrouping that coincides with the 212-224 loop region is shown in (Table 1) for the rest of the hits. The TLS analysis identifies segments made of successive residues (Table 1) with a highest mean absolute B-factor that is based on the absolute B-factor which corresponds to each atom in the deposited crystal structures. Also, the segments revealed by TLS analysis are independent of space groups and lattice parameters (Table 1). Among isomorphous crystal forms, TLS analysis consistently identifies the similar range of residues as having high mean B factors, regardless of the ligand that is bound to AR. However, the group of residues (segment) associated with the mobile loop seem to shift in the complexes compared to that in the apo structure (212-224) (Table 1).

The TLS analysis of aldehyde reductase (1HQT) from the aldo-keto reductase family with conserved structure, function and activity to AR was also performed. Although this entry was not a hit in the blast search when the flexible loop sequence of wild-type AR (1XGD) was utilized, the TLS analysis of 1HQT resulted in the identification of the highest mean B-factor corresponding to 214-226 segment (Table 1). This segment in entry 1HQT does not share amino acid sequence similarity with that of 1XGD. In 1HQT, O γ of 216, N of 215, N of 213, N of 211 and O γ of 211 are located at $2.8 \text{ \AA}$, $3.1 \text{ \AA}$, $2.8 \text{ \AA}$, $2.9 \text{ \AA}$ and $2.6 \text{ \AA}$ from O4B, O2A, O2A, O5D and O1N atoms of NADP⁺, respectively in the binary complex structure. Interestingly it was reported that segment containing residues 215–225 in the 1HQT structure is very flexible, and the RMSD is $1.23 \text{ \AA}$ for the main-chain atoms between the wild-type and Y50F mutant pig aldehyde reductase [50]. The results predicted by the TLS analysis for 1HQT is consistent with the previous observation and it implies the ability of B-factor over the sequence similarity in the concept of conformation.

Normalized B-factor Analysis

The comparison of absolute and normalized B-factors for each PDB entry reveals that the pattern of the B-factor profile are not significantly different. This observation is preserved in all of the hits as well. Furthermore, the 212-224 segment has the largest mean normalized B-factor in 1XGD and is significantly different from all segments but 22-29 in this entry (Table 2 and Fig. 2).

Nine entries considered here are refined using the same software, CNS/SHELXL and these structures also pertain to isomorphous space group, $P2_1$. The entries, which used the same refinement software (CNS/SHELXL), reveal that the B-factor of the flexible loop region is noticeably high among them (Table 1). The flexible loop of sixteen entries for all the AR-NADP⁺-inhibitor ternary complexes corresponds to significantly high B-factor regardless of the crystal form or space group ($P2_12_12_1$ or $P2_1$) of their structures (Table 3). Furthermore, this feature is apparently present in the B-factor analysis of a homologous structure, AKR1B10 (1ZUA).

Clustering of Residues

The hierarchical cluster analysis of B-factors performed with the Ward method routine can be used to identify similar residues. The Ward method, while maximizing the variance across clusters, minimizes the variance intra-clusters (within clusters). That is, if the members of each cluster are similar, then the variance of their B-factors among these members should be the lowest and hence, the intra-cluster variance should be low. In a similar vein, the inter-cluster variance should be high if the members of each cluster differ from members of the other clusters in their B-factor. The number of clusters could vary from one to the number of residues with no missing value. The appropriate number of clusters is chosen by the marginal explanatory power of the last cluster which should be greater or equal to 5%. The analysis of B-factor across all residues resulted in four clusters which explained 86% of the total variance. The fifth cluster explained only 2.8% of the variance and hence the four cluster choice was viewed as the best solution. The membership of these four clusters is shown in (Table 4). Two of the four clusters have a few residues (one with four residues is from 221 through 224 and the other with six residues is from 125 through 130) whereas the other two clusters have a large number of residues as their members (Table 4).

Clustering of PDB Entries

In the current case, the residues are clustered on the basis of their B-factor which is the only measurement that varies. PDB entries are observations of the B-factor for each of the residue. It is statistically meaningless to cluster PDB entries (observations) based on only B-factor because the number of variables (residues) are more than the number of observations (PDB entries).

Analysis of Root Mean Squared Deviations

Analysis of variance (ANOVA) (Table 5) helps to identify how the dependent variable differs across different groups of observations. Evaluation of the variances of the B-factors and RMSD across the same clusters indicates that RMSD varied significantly across these clusters, confirming their correlation with B-factor. As previously established, the RMSDs reflect the dynamics of atoms [37, 51], and hence their values differ across clusters created by B-factors of the corresponding atoms. The B-factor based clustering model explains over 21.7% of the total variance and the difference across clusters were statistically significant with over 99% certainty.

DISCUSSION

This study identified 1) the 212-224 segment in the apo structure (1XGD) as the most flexible group by TLS analysis, 2) sliding of this segment by about ten residue segments are in different entries in TLS analysis and 3) loops with tighter residues clustered into different groups by statistical analysis of B-factor only approach. In general this segment overlaps with the flexible loop, which is independently predicted previously as loop 7 [5, 7], which is expected to undergo conformational change during cofactor binding and release.

TLS analysis of several entries (Table 1) displays a shifting of the boundaries of the flexible loop around 212-224 segment that is seen in the apo enzyme. Also, it is interesting to note that several PDB entries (3LZ5, 3LQG, 3LD5, 3LBO, 2PDM, 2PDK, 2PDI, 2AGT, 2PDX) analyzed here use the same refinement software (CNS/SHELXL), belong to an isomorphous crystal system and contain ternary complex structures with inhibitors which are very close chemical analogs of each other but the flexible loop shifts and slides significantly. However the method that uses clustering of the B-factor of the residues of all the entries results in a much tighter loop with residues 221-224.

The theory underlying TLS parameterization has been presented in detail by Schomaker and Trueblood [40, 52] and Howlin *et al.*, [53]. The comparison of TLS parameters with the amplitudes of normal modes describing the internal motion of a protein indicates that TLS parameters for the overall motion make the largest contribution to atomic displacements [54, 55]. Furthermore, TLS analysis has been shown to provide biologically meaningful information, for example in the detection of mobile domains which contribute to the induced fit in a protein. These regions are otherwise with restricted mobility, and comprise of critically relevant functional sites, like the active sites of enzymes [56, 57]. Interestingly, the TLS analysis of the R form of Aspartate carbamoyltransferase (ATCase) reveals the relationship between the atomic fluctuations and the temperature factors associated with its catalytic and zinc domains [58].

Analysis of B-factor using clustering methodology confirms that certain residues are significantly different from others. B-factors capture not only the static features of the molecular structures but also the dynamics of the internal movements of atoms within these molecules. This dynamics of internal movements are reflected in the RMSD of each atom in every residue (and across other residues in another observations/entries). Our analysis of the variance of the RMSDs confirms that these RMSDs are different across clusters created by B-factors.

The B-factor is one of the principal scale terms between the experimental and calculated structure factors in the crystal structure determination. This parameter has taken a more meaningful role as a measure of the conformational flexibility [33, 36, 37, 59–61], even though a fraction of the B-factor reflects the errors in the fitting of the structural model against the experimental data. Since the component of the B-factor that corresponds to the errors in the model is uniform at a given final R-value, it is conceivable that the resultant mean B-factor will contain significant flexibility information in each corresponding residue within the molecule.

CONCLUSION

Overall, AR and its analogs make a good experimental set of structures to explore the importance of B-factor based analysis because 1) they have been crystallized in more than one space group from various laboratories, 2) apo enzyme as well as different ligand bound crystal structures are available in PDB for this family of proteins and 3) fairly well characterized flexible loop phenomenon is independently demonstrated by other techniques. No correlation stood out between the experimental parameters, space group, software or X-ray source and the B-factor in clustering approach. On the other hand interestingly B-factor clustering analysis reveals a clear correlation between the B-factor and RMSD of their C α atoms corresponding to the residues in the flexible loop in these PDB entries. In summary, concurrence of the segment with a significant difference in the B-factor and the flexible loop that is identified to undergo conformational change during catalysis from kinetic studies is not a random coincidence at least in the collection of Aldo-keto reductase family protein structures that are considered in this study.

Major difference between TLS and cluster analysis is the finding that the variation in the number of residues and the sliding of the loop between 212-234 compares to only 221-224 in cluster 3, 215-220 in cluster 1 and 212-214 in cluster 4. In TLS type analysis segments with successive residues are used to determine the groups that undergo dynamics whereas in the B-factor clustering approach individual residue is evaluated in grouping without any connectivity constrain. Therefore the latter approach resolves the dynamics at a single residue level where the residues don't have to be organized consecutively in the molecule.

Fundamentally the information obtained from this analysis may become useful for the understanding of the conformational changes that the enzyme undergoes during the function. This finding may also become valuable in the design of most effective inhibitor.

Acknowledgments

FUNDING

This work is supported by funding DK085496 from the National Institute of Diabetes and Digestive and Kidney Diseases and Dean's Reassign time. We thank Dr. Michael Sa-way for helpful discussion and Dr. Ethan Merritt for explaining the aspects related to TLS analysis in multiple occasions.

ABBREVIATIONS

AR	Aldose reductase
AKR	Aldo-keto reductase
ATCase	Aspartate carbamoyltransferase
PDB	Protein Data Bank
SDH	Sorbitol dehydrogenase
TLS	Translational, librational and screw motion
STD	Standard deviation

RMSD Root Mean Squared deviations

References

1. Van Heyningen R. Formation of polyols by the lens of the rat with sugar cataract. *Nature*. 1959; 184:194–195.
2. Hers HG. The mechanism of the formation of seminal fructose and fetal fructose. *Biochim Biophys Acta*. 1960; 37:127–38. [PubMed: 14401391]
3. Grimshaw CE, Shahbaz M, Putney CG. Mechanistic basis for nonlinear kinetics of aldehyde reduction catalyzed by aldose reductase. *Biochemistry*. 1990; 29(42):9947–9955. [PubMed: 2125486]
4. Kubiseski TJ, Hyndman DJ, Morjana NA, Flynn TG. Studies on pig muscle aldose reductase. Kinetic mechanism and evidence for a slow conformational change upon coenzyme binding. *J Biol Chem*. 1992; 267(10):6510–6517. [PubMed: 1551865]
5. Borhani DW, Harter TM, Petrash JM. The crystal structure of the aldose reductase. NADPH binary complex. *J Biol Chem*. 1992; 267(34):24841–24847. [PubMed: 1447221]
6. Wilson DK, Bohren KM, Gabbay KH, Quiocho FA. An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications. *Science*. 1992; 257(5066):81–84. [PubMed: 1621098]
7. Grimshaw CE, Bohren KM, Lai CJ, Gabbay KH. Human aldose reductase: rate constants for a mechanism including inter-conversion of ternary complexes by recombinant wild-type enzyme. *Biochemistry*. 1995; 34(44):14356–14365. [PubMed: 7578039]
8. Grimshaw CE, Bohren KM, Lai CJ, Gabbay KH. Human aldose reductase: subtle effects revealed by rapid kinetic studies of the C298A mutant enzyme. *Biochemistry*. 1995; 34(44):14366–14373. [PubMed: 7578040]
9. Liu SQ, Bhatnagar A, Srivastava SK. Does sorbinil bind to the substrate binding site of aldose reductase? *Biochem Pharmacol*. 1992; 44(12):2427–2429. [PubMed: 1472112]
10. Trueblood KN, Brghi HB, Burzlaff H, Dunitz JD, Gramaccioli CM, Schulz HH, Shmueli U, Abrahams SC. Atomic displacement parameter nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature. *Acta Cryst*. 1996; 52:770–781.
11. Drenth, J. Principles of protein crystallography. Springer-Verlag; New York: 1994.
12. Carugo O, Argos P. Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic thermal factors. *Proteins*. 1998; 31(2):201–213. [PubMed: 9593193]
13. Stroud RM, Fauman EB. Significance of structural changes in proteins: expected errors in refined protein structures. *Protein Sci*. 1995; 4(11):2392–2404. [PubMed: 8563637]
14. Carugo O, Argos P. Correlation between side chain mobility and conformation in protein structures. *Protein Eng*. 1997; 10(7):777–787. [PubMed: 9342144]
15. Carugo O, Argos P. Protein-protein crystal-packing contacts. *Protein Sci*. 1997; 6(10):2261–2263. [PubMed: 9336849]
16. Carugo O. Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Eng*. 1999; 12(12):1021–1024. [PubMed: 10611392]
17. Vihinen M. Relationship of protein flexibility to thermostability. *Protein Eng*. 1987; 1(6):477–480. [PubMed: 3508295]
18. Altman RB, Hughes C, Zhao D, Jardetsky O. Compositional characteristics of relatively disordered regions in proteins. *Prot Pept Lett*. 1994; 1:120–127.
19. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S, Dunker AK. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomp*. 1998:437–448.
20. Khabiri M, Minofar B, Brezovsky J, Damborsky J, Ettrich R. Interaction of organic solvents with protein structures at protein-solvent interface. *J Mol Model*. 2013; 19(11):4701–4711. [PubMed: 22760789]
21. Cousido-Siah A, Petrova T, Hazemann I, Mitschler A, Ruiz FX, Howard E, Ginell S, Atmanene C, Van Dorsselaer A, Sanglier-Cianferani S, Joachimiak A, Podjarny A. Crystal packing modifies

- ligand binding affinity: The case of aldose reductase. *Proteins*. 2012; 80(11):2552–2561. [PubMed: 22752989]
22. Zhang JH, Lin Y, Sun YF, Ye YR, Zheng SP, Han SY. High-throughput screening of B factor saturation mutated *Rhizomucor miehei* lipase thermostability based on synthetic reaction. *Enzyme Microb Technol*. 50(6–7):325–330. [PubMed: 22500900]
 23. Fisher CK, Stultz CM. Protein structure along the order-disorder continuum. *J Am Chem Soc*. 133(26):10022–10025. [PubMed: 21650183]
 24. Dugdale ML, Vance ML, Wheatley RW, Driedger MR, Nibber A, Tran A, Huber RE. Importance of Arg-599 of beta-galactosidase (*Escherichia coli*) as an anchor for the open conformations of Phe-601 and the active-site loop. *Biochem Cell Biol*. 88(6):969–979. [PubMed: 21102659]
 25. DeLorbe JE, Clements JH, Teresk MG, Benfield AP, Plake HR, Millspaugh LE, Martin SF. Thermodynamic and structural effects of conformational constraints in protein-ligand interactions. Entropic paradox associated with ligand preorganization. *J Am Chem Soc*. 2009; 131(46):16758–16770. [PubMed: 19886660]
 26. Urzhumtseva L, Afonine PV, Adams PD, Urzhumtsev A. Crystallographic model quality at a glance. *Acta Crystallogr D Biol Crystallogr*. 2009; 65(Pt 3):297–300. [PubMed: 19237753]
 27. Rao MK, Chapman TR, Finke JM. Crystallographic B-factors highlight energetic frustration in aldolase folding. *J Phys Chem B*. 2008; 112(34):10417–10431. [PubMed: 18671365]
 28. Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. *Gene*. 2008; 422(1–2):7–13. [PubMed: 18577430]
 29. Halle B. Flexibility and packing in proteins. *Proc Natl Acad Sci USA*. 2002; 99(3):1274–1279. [PubMed: 11818549]
 30. Yuan Z, Bailey TL, Teasdale RD. Prediction of protein B-factor profiles. *Proteins*. 2005; 58(4):905–912. [PubMed: 15645415]
 31. Ruvinsky AM, Kirys T, Tuzikov AV, Vakser IA. Structure fluctuations and conformational changes in protein binding. *J Bioinform Comput Biol*. 10(2):1241002. [PubMed: 22809338]
 32. Glyakina AV, Lobanov MY, Galzitskaya OV. Structural-Functional Analysis of Biopolymers and Their Complexes A search for structural factors responsible for the stability of proteins from thermophilic organisms. *Mol Biol*. 2007; 41(4):616–621.
 33. Schlessinger A, Punta M, Yachdav G, Kajan L, Rost B. Improved disorder prediction by combination of orthogonal approaches. *PLoS One*. 2009; 4(2):e4433. [PubMed: 19209228]
 34. Yuan Z, Zhao J, Wang ZX. Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng*. 2003; 16(2):109–114. [PubMed: 12676979]
 35. Parak FG. Physical aspects of protein dynamics. *Rep Prog Phys*. 2003; 66:103–129.
 36. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins- a tool for the selection of peptide antigens. *Naturwissenschaften*. 1985; 72:212–213.
 37. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins*. 1994; 19(2):141–149. [PubMed: 8090708]
 38. Smith DK, Radivojac P, Obradovic Z, Dunker AK, Zhu G. Improved amino acid flexibility parameters. *Protein Sci*. 2003; 12(5):1060–1072. [PubMed: 12717028]
 39. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28(1):235–242. [PubMed: 10592235]
 40. Schomaker V, Trueblood KN. On the rigid-body motion of molecules in crystals. *Acta Crystallogr B*. 1968; 24:63–76.
 41. Bohren KM, Brownlee JM, Milne AC, Gabbay KH, Harrison DH. The structure of Apo R268A human aldose reductase: hinges and latches that control the kinetic mechanism. *Biochim Biophys Acta*. 2005; 1748(2):201–212.
 42. Painter J, Merritt EA. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Crystallogr D*. 2005; 61(Pt 4):465–471. [PubMed: 15809496]
 43. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D*. 2006; 62(Pt 4):439–450. [PubMed: 16552146]

44. Micheletti C, Carloni P, Maritan A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins*. 2004; 55(3):635–645. [PubMed: 15103627]
45. Parthasarathy S, Murthy MR. Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng*. 2000; 13(1):9–13. [PubMed: 10679524]
46. Siglioccolo A, Gerace R, Pascarella S. “Cold spots” in protein cold adaptation: Insights from normalized atomic displacement parameters (B'-factors). *Biophys Chem*. 2010; 153(1):104–114. [PubMed: 21075502]
47. Ward JH. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*. 1963; 58:236–244.
48. Cohen FE, Sternberg MJ. On the prediction of protein structure: The significance of the root-mean-square deviation. *J Mol Biol*. 1980; 138(2):321–333. [PubMed: 7411610]
49. Wampler JE. Distribution Analysis of the Variation of B-Factors of X-ray Crystal Structures: Temperature and Structural Variations in Lysozyme. *J Chem Inf Comp Sci*. 1997; 37:1171–1180.
50. Ye Q, Hyndman D, Green N, Li X, Korithoski B, Jia Z, Flynn TG. Crystal structure of an aldehyde reductase Y50F mutant-NADP complex and its implications for substrate binding. *Proteins*. 2001; 44(1):12–19. [PubMed: 11354001]
51. Kuzmanic A, Zagrovic B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys J*. 98(5):861–871. [PubMed: 20197040]
52. Schomaker V, Trueblood KN. Correlation of Internal Torsional Motion with Overall Molecular Motion in Crystals. *Acta Crystallogr B*. 1998; 54:507–514.
53. Howlin B, Moss DS, Harris GW. Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallogr A*. 1989; 45:851–861. [PubMed: 2619965]
54. Diamond R. On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr A*. 1990; 46(Pt 6):425–435. [PubMed: 1694442]
55. Kidera A, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. *J Mol Biol*. 1992; 225(2):457–475. [PubMed: 1593630]
56. Yousef MS, Clark SA, Pruett PK, Somasundaram T, Ellington WR, Chapman MS. Induced fit in guanidino kinases--comparison of substrate-free and transition state analog structures of arginine kinase. *Protein Sci*. 2003; 12(1):103–111. [PubMed: 12493833]
57. Chaudhry C, Horwich AL, Brunger AT, Adams PD. Exploring the structural dynamics of the E.coli chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. *J Mol Biol*. 2004; 342(1):229–245. [PubMed: 15313620]
58. Tanner JJ, Smith PE, Krause KL. Molecular dynamics simulations and rigid body (TLS) analysis of aspartate carbamoyl-transferase: evidence for an uncoupled R state. *Protein Sci*. 1993; 2(6): 927–935. [PubMed: 8318897]
59. Willis, BTM.; Pryor, AW. *Thermal Vibrations in Crystallography*. Cambridge University Press; London, New York: 1975.
60. Petsko GA, Ringe D. Fluctuations in protein structure from X-ray diffraction. *Annu Rev Biophys Bioeng*. 1984; 13:331–371. [PubMed: 6331286]
61. Chen K, Kurgan LA, Ruan J. Prediction of flexible/rigid regions from protein sequences using k-spaced amino acid pairs. *BMC Struct Biol*. 2007; 7:25. [PubMed: 17437643]
62. Ye Q, Hyndman D, Green NC, Li L, Jia Z, Flynn TG. The crystal structure of an aldehyde reductase Y50F mutant-NADP complex and its implications for substrate binding. *Chem Biol Interact*. 2001; 130–132(1–3):651–658.
63. Bohren KM, Grimshaw CE, Lai CJ, Harrison DH, Ringe D, Petsko GA, Gabbay KH. Tyrosine-48 is the proton donor and histidine-110 directs substrate stereochemical selectivity in the reduction reaction of human aldose reductase: enzyme kinetics and crystal structure of the Y48H mutant enzyme. *Biochemistry*. 1994; 33(8):2021–2032. [PubMed: 8117659]

64. Koch C, Heine A, Klebe G. Tracing the detail: how mutations affect binding modes and thermodynamic signatures of closely related aldose reductase inhibitors. *J Mol Biol.* 406(5):700–712. [PubMed: 21185307]
65. Petrova T, Steuber H, Hazemann I, Cousido-Siah A, Mitschler A, Chung R, Oka M, Klebe G, El-Kabbani O, Joachimiak A, Podjarny A. Factorizing selectivity determinants of inhibitor binding toward aldose and aldehyde reductases: structural and thermodynamic properties of the aldose reductase mutant Leu300Pro-fidarestat complex. *J Med Chem.* 2005; 48(18):5659–5665. [PubMed: 16134934]
66. Blakeley MP, Ruiz F, Cachau R, Hazemann I, Meilleur F, Mitschler A, Ginell S, Afonine P, Ventura ON, Cousido-Siah A, Haertlein M, Joachimiak A, Myles D, Podjarny A. Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase. *Proc Natl Acad Sci USA.* 2008; 105(6):1844–1848. [PubMed: 18250329]
67. Steuber H, Heine A, Podjarny A, Klebe G. Merging the binding sites of aldose and aldehyde reductase for detection of inhibitor selectivity-determining features. *J Mol Biol.* 2008; 379(5):991–1016. [PubMed: 18495158]
68. Brownlee JM, Carlson E, Milne AC, Pape E, Harrison DH. Structural and thermodynamic studies of simple aldose reductase-inhibitor complexes. *Bioorg Chem.* 2006; 34(6):424–444. [PubMed: 17083960]
69. Van Zandt MC, Jones ML, Gunn DE, Geraci LS, Jones JH, Sawicki DR, Sredy J, Jacot JL, Dicioccio AT, Petrova T, Mitschler A, Podjarny AD. Discovery of 3-[(4,5,7-trifluorobenzothiazol-2-yl)methyl]indole-N-acetic acid (lidorestat) and congeners as highly potent and selective inhibitors of aldose reductase for treatment of chronic diabetic complications. *J Med Chem.* 2005; 48(9):3141–3152. [PubMed: 15857120]
70. El-Kabbani O, Darmanin C, Schneider TR, Hazemann I, Ruiz F, Oka M, Joachimiak A, Schulze-Briesse C, Tomizaki T, Mitschler A, Podjarny A. Ultrahigh resolution drug design. II. Atomic resolution structures of human aldose reductase holoenzyme complexed with Fidarestat and Minalrestat: implications for the binding of cyclic imide inhibitors. *Proteins.* 2004; 55(4):805–813. [PubMed: 15146479]
71. Calderone V, Chevrier B, Van Zandt M, Lamour V, Howard E, Poterszman A, Barth P, Mitschler A, Lu J, Dvornik DM, Klebe G, Kraemer O, Moorman AR, Moras D, Podjarny A. The structure of human aldose reductase bound to the inhibitor IDD384. *Acta Crystallogr D.* 2000; 56(Pt 5):536–540. [PubMed: 10771421]
72. Urzhumtsev A, Tete-Favier F, Mitschler A, Barbanton J, Barth P, Urzhumtseva L, Biellmann JF, Podjarny A, Moras D. A ‘specificity’ pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. *Structure.* 1997; 5(5):601–612. [PubMed: 9195881]
73. Gallego O, Ruiz FX, Ardevol A, Dominguez M, Alvarez R, de Lera AR, Rovira C, Farres J, Fita I, Pares X. Structural basis for the high all-trans-retinaldehyde reductase activity of the tumor marker AKR1B10. *Proc Natl Acad Sci USA.* 2007; 104(52):20764–20769. [PubMed: 18087047]
74. Balendiran GK, Sawaya MR, Schwarz FP, Ponniah G, Cuckovich R, Verma M, Cascio D. The role of Cys298 in aldose reductase function. *J Biol Chem.* 2010; 286:6336–6344. [PubMed: 21084309]

Fig. 1.

Topology of human AR. The sequential organization of secondary structural elements in the three-dimensional architecture of AR is arranged from N-terminal to C-terminal. Arrows indicate the β strands and their directions; elongated cylinders represent α helices. Single, solid lines with arrows indicate the loop regions connecting the defined secondary elements. The secondary structural elements shown are experimentally determined as defined by their structural parameters. TLS segments are marked with the colored text and coded with the same color as 1-11 (■), 12-21 (■), 22-29 (■), 30-46 (■), 47-57 (■), 58-77 (■), 78-84 (■), 85-94 (■), 95-112 (■), 113-121 (■), 122-137 (■), 138-147 (■), 148-165 (■), 166-180 (■), 181-211 (■), 212-224 (■), 225-241 (■), 242-273 (■), 274-295 (■) and 296-315 (■).

Fig. 2.

The TLS analysis predicts that the motions AR may undergo during the conformational dynamics. These are 20 backbone traces (identities of the molecules are listed in Table 1) showing zero/small/large displacement along the maximum librational component of the TLS model (Dr. Ethan Merritt personal communication). The boundaries and colors of 20 groups used in the TLS analysis are 1-11 (■), 12-21 (■), 22-29 (■), 30-46 (■), 47-57 (■), 58-77 (■), 78-84 (■), 85-94 (■), 95-112 (■), 113-121 (■), 122-137 (■), 138-147 (■), 148-165 (■), 166-180 (■), 181-211 (■), 212-224 (■), 225-241 (■), 242-273 (■), 274-295 (■) and 296-315 (■).

Table 1

Entries of AR family in PDB that were used in the TLS analysis along with their results.

PDB Code	 value (Å ²)	TLS Predicted Flexible Loop	Cell Parameters (Å, °)	Resolution (Å)/R-value (%)	Temperature (K)	Molecular Assembly
1XGD	33.9	²¹² LGSPPDRPWAKPED ²²⁴	64.2, 64.2, 176.6, α=β=γ=90, P ₄ ₃ 2 ₁ 2	2.1/16.8	277	Apo R268A hAR
1EKO	68.4 ^a	²¹² LGSPPDRPWAKP ²²²	68.8, 68.8, 154.8, α=β=γ=90, P ₄ ₃ 2 ₁ 2	2.2/21.6	273	Pig AR NADP ⁺ , IDD384
2ACU	49.5	²²⁰ AKPEDPS ²²⁶	49.99, 67.1, 92.1, α=β=γ=90, P ₂ ₁ 2 ₁ 2 ₁	1.76/18.7	Null	Y48H hAR NADP ⁺ , Citrate
2IS7	29.6	²²⁰ AKPEDPSL ²²⁷	50.0, 67.2, 92.1, α=β=γ=90, P ₂ ₁ 2 ₁ 2 ₁	1.70/17.0	277	hAR NADP ⁺ , CAS:6575-24-2
3LZ5	28.9	²¹⁹ WAKPED ²²⁴	49.3, 66.8, 47.3, β=92.2, P ₂ ₁	0.95/11.8	100	T113V hAR NADP ⁺ , IDD594
3LQG	18.3	²¹⁶ DRPWAKP ²²²	49.1, 66.7, 47.1, β=92.4, P ₂ ₁	1.35/11.7	100	T113A hAR NADP ⁺ , IDD388
3LD5	25.2	²¹⁷ RPWAKPED ²²⁴	49.2, 66.8, 47.3, β=92.3, P ₂ ₁	1.27/12.5	100	T113S hAR NADP ⁺ , IDD594
3LBO	26.1 ^b	²¹⁸ PWAKPED ²²⁴	49.2, 66.7, 47.3, β=92.4, P ₂ ₁	1.10/11.4	100	T113C hAR NADP ⁺ , IDD594
2R24	33.9	²¹⁸ PWAKPEDPSLLED ²³⁰	50.1, 67.1, 47.9, β=92.4, P ₂ ₁	1.75/13.2	298	hAR NADP ⁺ , IDD594
2PDM	53.6	²²¹ KPEDP ²²⁵	49.5, 66.4, 47.4, β=92.2, P ₂ ₁	1.75/17.4	100	S302R hAR NADP ⁺ , Zopolrestat
2PDK	29.8 ^c	²²² EDPSLLEDPRIK ²³⁴	49.4, 66.6, 47.4, β=91.9, P ₂ ₁	1.55/15.9	100	L301M hAR NADP ⁺ , Sorbinil
2PDI	19.4 ^d	²²¹ KPEDPSLL ²²⁸	49.4, 66.9, 47.3, β=93, P ₂ ₁	1.55/15.5	100	L300A hAR NADP ⁺ , Zopolrestat
1Z3N	29.9	²¹⁸ PWAKPEDP ²²⁵	50.1, 67.3, 47.9, β=92.5, P ₂ ₁	1.04/9.8	100	hAR NADP ⁺ , Lidorestat
2AGT	26.2	²²⁰ AKPEDP ²²⁵	49.1, 66.7, 47.1, β=92.79, P ₂ ₁	1.00/10.6	100	L300P hAR NADP ⁺ , Fidarestat, Citrate
1PWL	24.0	²²¹ KPEDP ²²⁵	40.0, 47.1, 47.4, 75.7, 67.5, 76.8, P ₁	1.10/10.0	100	hAR NADP ⁺ , Minalrestat
1AH0	61.3	²²³ EDPSLLEDP ²³¹	67.9, 67.9, 152.8, α=β=γ=90, P ₄ ₃ 2 ₁ 2	2.30/19.2	273	Pig AR NADP ⁺ , Sorbinil
2PDX	23.8 ^e	²²¹ KPEDPSLLEDPR ²³³	49.3, 67.3, 47.8, β=91.4, P ₂ ₁	1.65/17.7	100	S302R-C303D hAR NADP ⁺ , Zopolrestat
1ZUA	21.0 ^f	²²² KPEDPSLLEDPK ²³⁴	89.1, 89.1, 78.4, β=120, P ₆ ₁	1.25/13.1	298	AKR1B10 NADP ⁺ , Tolrestat
1HQT	87.6	²¹⁴ GSSDRAWRDPNEP ²²⁶	53.3, 70.1, 92.7, α=β=γ=90, P ₂ ₁ 2 ₁ 2 ₁	2.20/21.2	100	Y50F pig Aldehyde Reductase NADP ⁺
3Q67	TLS restraints used in the refinement		83.7, 86.2, 104.4, α=β=γ=90, P ₂ ₁ 2 ₁ 2 ₁	1.55/17.1	100	C298S hAR NADP ⁺

- ^ahighest 83.6 Å² corresponds to 223-237 in IEKO.
- ^bhighest 41.6 Å² corresponds to 311-315 in 3LBO.
- ^chighest 34.5 Å² corresponds to 300-305 in 2PDK.
- ^dhighest 33.5 Å² corresponds to 312-315 in 2PDL.
- ^ehighest 25.5 Å² corresponds to 299-303 in 2PDX.
- ^fhighest 23.9 Å² corresponds to 235-243 in 1ZUA.
- IDD384: [2,6-Dimethyl-4-(2-O-tolyl-acetylamino)-benzenesulfonyl]-glycine
- CAS Number: 6575-24-2: (2,6-Dichlorophenyl)acetic acid
- IDD594: [2-(4-Bromo-2-fluoro-benzylthiocarbamoyl)-5-fluoro-phenoxy]-acetic acid
- IDD388: (2-([(4-Bromo-2-fluorobenzyl)amino]carbonyl)-5-chlorophenoxy)acetic acid
- IDD594: 2-(4-Bromo-2-fluoro-benzylthiocarbamoyl)-5-fluoro-phenoxy]-acetic acid
- Zopolrestat: 3,4-Dihydro-4-oxo-3-((5-trifluoromethyl-2-benzothiazolyl)methyl)-1-phthalazine acetic acid
- Sorbinil: (4S)-6-Fluoro-2,3-dihydro-spiro[4H-1-benzopyran-4,4'-imidazolidine]-2',5'-dione
- Lidorestat: {3-[(4,5,7-Trifluoro-1,3-benzothiazol-2-yl)methyl]-1H-indol-1-yl}acetic acid
- Fidarestat: (2S,4S)-2-Aminoformyl-6-fluoro-spiro[chroman-4,4'-imidazolidine]-2',5'-dione
- Minalrestat: 2[4-Bromo-2-fluorophenyl)methyl]-6-fluorospir[isoquinoline-4-(1H),3'-pyrrolidine]-1,2',3,5' (2H)-tetrone
- Tolrestat: 2-[[6-methoxy-5-(trifluoromethyl)naphthalene-1-carbothioyl]-1-methylamino]acetic acid
- 3Q67-C298S human AR, 2PDQ- C303D human AR, 2PDF-L300P human AR, 2PDB-F121P human AR, 2PDS-V47I human AR, 1IEF3-Human AR.

Table 2
Statistical analysis of normalized B-factor ($\text{\AA}^2$) of fragments (residue range) in 1XGD.

Fragment	212-224	1-11	12-21	22-29	30-46	47-57	58-77	78-84	85-94	95-112
Average	1.7629	0.8396	-0.3332	1.7770	-0.3849	-0.4240	0.5984	-0.5206	-0.3055	-0.7783
Std	1.121	1.4992	0.8891	1.1923	0.5876	0.4102	1.175	0.6001	0.4618	0.5063
Observations	13	11	10	8	17	11	20	7	10	18
t-test value		1.72	4.85	-0.03	6.80	6.12	2.83	4.98	5.47	8.53
Probability <		0.0986	0.0001	0.9784	0.0001	0.0001	0.0081	0.0001	0.0001	0.0001

Fragment	212-224	113-121	122-137	138-147	148-165	166-180	181-211	225-241	242-273	274-295	296-313
Average	1.7629	0.321	0.8147	-0.7180	-0.8728	-0.3404	-0.7087	0.5781	-0.3550	0.2815	0.2665
Std	1.121	0.5809	0.7701	0.4439	0.4529	0.5173	0.5490	0.5046	0.5858	0.5416	0.4137
Observations	13	9	16	10	18	15	31	17	32	22	18
t-test value		3.53	2.69	6.75	9.05	6.52	9.87	3.89	8.33	5.28	5.22
Probability <		0.0021	0.0120	0.0001	0.0001	0.0001	0.0001	0.0006	0.0001	0.0001	0.0001

Table 3

X-ray sources, data collection devices and structure determination/refinement software used in the generation of the entries.

Entry	Rotating Anode	Synchrotron	Detector Type	Software	References
1XGD	Rigaku RU200	N	Rigaku R-axis IIC area detector	XPLOR	[41]
1HQT	Rigaku RU200	N	MARRESEARCH image plate	CNS	[62]
2ACU	Eliot GX-6	N	Siemens X-100A multiwire proportional X-ray detector	XPLOR	[63]
3LZ5	N	Berlin/BESSY 14.2	MARMOSAIC CCD	CNS/SHELXL	[64]
3LD5	N	Berlin/BESSY	MARMOSAIC CCD	CNS/SHELXL	[64]
3LQG	N	Berlin/BESSY	MARMOSAIC CCD	CNS/SHELXL	[64]
2AGT	N	APS/19-ID	SBC-2 CCD	CNS/SHELXL	[65]
3LBO	N	Berlin/BESSY	MARMOSAIC CCD	CNS/SHELXL	[64]
2R24	N	ANL/APS/SBC 19ID	Not clear	PHENIX	[66]
2PDK	Rigaku RU300	N	Rigaku RAXIS IV image plate	CNS/SHELXL	[67]
2IS7	Rigaku RU200	N	Rigaku RAXIS IIC	REFMAC	[68]
1Z3N	N	DW32	MARRESEARCH image plate	X-PLOR/SHELXL	[69]
1PWL	N	APS/19-BM	Custom made CCD	X-PLOR/SHELXL	[70]
2PDM	Rigaku RU300	N	Rigaku RAXIS IV image plate	CNS/SHELXL	[67]
2PDX	Rigaku RU300	N	Rigaku RAXIS IV image plate	CNS/SHELXL	[67]
1EKO	N	DW32	Image plate	Not provided	[71]
1AH0	N	LURE/ DW32	MARRESEARCH image plate	XPLOR	[72]
2PDI	Rigaku RU300	N	Rigaku RAXIS IV image plate	CNS/SHELXL	[67]
1ZUA	N	ESRF/ID14-4	ADSC QUANTUM 4R CCD	REFMAC	[73]
3Q67	N	ALS/8.2.2	ADSC QUANTUM 315 CCD	REFMAC	[74]

Table 4

Membership of clusters (based on normalized B-factors).

Cluster	Membership Observations (Residues)
1	2, 226, 225, 277, 227, 228, 229, 276, 293, 294, 234, 235, 279, 238, 239, 241, 240, 278, 230, 233, 231, 232, 236, 237, 242, 243, 280, 284, 281, 283, 282, 285, 288, 286, 287, 290, 3, 124, 60, 67, 61, 62, 69, 71, 70, 66, 132, 133, 121, 120, 131, 23, 24, 134, 135, 25, 26, 68, 118, 119, 63, 64, 65, 215, 216, 217, 218, 219, 220, 4, 5, 10, 11, 6, 87, 169, 198, 174, 202, 203, 138, 139, 170, 163, 164, 84, 197, 200, 194, 311, 296, 297, 7, 8, 9, 167, 168, 172, 171, 201, 173, 85, 86, 136, 137, 299, 303, 304, 305, 307, 306, 308, 300, 302, 301
2	125, 126, 127, 128, 129, 130 (ends with 128)
3	221, 222, 223, 224 (ends with 223)
4	12, 147, 179, 55, 145, 155, 13, 205, 82, 83, 113, 162, 88, 91, 99, 115, 152, 92, 154, 96, 153, 103, 144, 114, 141, 180, 146, 148, 149, 30, 98, 101, 93, 94, 166, 89, 90, 97, 100, 142, 150, 151, 309, 310, 140, 143, 165, 204, 192, 195, 196, 199, 175, 177, 178, 176, 14, 15, 189, 186, 187, 190, 191, 211, 34, 188, 250, 248, 252, 273, 251, 253, 36, 254, 256, 37, 268, 38, 255, 72, 272, 245, 246, 247, 257, 31, 49, 52, 39, 40, 81, 56, 32, 33, 58, 212, 263, 264, 21, 22, 214, 27, 28, 102, 266, 265, 29, 59, 116, 117, 122, 123, 213, 298, 193, 267, 270, 271, 244, 275, 274, 292, 289, 291, 16, 181, 18, 42, 74, 159, 105, 206, 207, 208, 259, 41, 258, 156, 157, 209, 260, 112, 160, 185, 184, 17, 19, 161, 261, 210, 249, 20, 262, 46, 47, 73, 104, 48, 51, 50, 53, 95, 54, 57, 43, 75, 109, 79, 80, 182, 183, 44, 45, 77, 78, 110, 111, 76, 158, 106, 107, 108

Clusters based on 20 variables and 308 observations (No factor analysis)

Table 5

Analysis of variance of B-Factors.

Source	Df	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	56.34	18.78	28.00	< 0.0001
Error	303	203.21	0.67		
Corrected Total	306	259.55			
R ²	0.2171				
Mean RMSD	1.39				

Dependent Variable: RMSD