Common structural features of the *luxF* protein and the subunits of bacterial luciferase: Evidence for a $(\beta\alpha)_8$ fold in luciferase

STANLEY A. MOORE AND MICHAEL N.G. JAMES

MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

(RECEIVED March 3, 1994; ACCEPTED August 30, 1994)

Abstract

The amino acid sequence identity and potential structural similarity between the subunits of bacterial luciferase and the recently determined structure of the *luxF* molecule are examined. The unique β/α barrel fold found in *luxF* appears to be conserved in part in the luciferase subunits. From secondary structural predictions of both luciferase subunits, and from structural comparisons between the protein product of the *luxF* gene, NFP, and glycolate oxidase, we propose that it is feasible for both luciferase subunits to adopt a $(\beta\alpha)_8$ barrel fold with at least 2 excursions from the $(\beta\alpha)_8$ topology. Amino acids conserved between NFP and the luciferase subunits cluster together in 3 distinct “pockets” of NFP, which are located at hydrophobic interfaces between the β -strands and α -helices. Several tight turns joining the C-termini of β -strands and the N-termini of α -helices are found as key components of these conserved regions. Helix start and end points are easily demarcated in the luciferase subunit protein sequences; the N-cap residues are the most strongly conserved structural features. A partial model of the luciferase β subunit from *Photobacterium leiognathi* has been built based on our crystallographically determined structure of *luxF* at 1.6 Å resolution.

Keywords: bacterial bioluminescence; homology; luciferase; *luxF*; subunits

Bacterial bioluminescence is a fascinating example of light emission from a living organism. All light-emitting bacteria have a *lux* operon that is responsible for the luminous phenotype (Meighen, 1991, 1993). Most of these bacteria are marine in origin and many are symbionts in fish. A terrestrial bioluminescent bacterium, *Xenorhabdus luminescens*, has a symbiotic relationship with a parasitic nematode and its host moth. How these often highly selective symbioses arose, and what advantages they convey to both symbionts are intriguing biological questions. Perhaps more interesting still is how such luminous pathways were developed by bacteria, as several of the *lux* proteins seem to have been borrowed from other metabolic pathways (e.g., from riboflavin biosynthesis) (O’Kane et al., 1991; Lee & Meighen, 1992; O’Kane & Prasher, 1992; Richter et al., 1992).

The luminescence reaction,

(1)

is catalyzed by bacterial luciferase, a heterodimeric flavin monooxygenase that uses myristic aldehyde, FMNH_2 , and an oxygen molecule to produce light and the oxidized products: myristic acid, FMN, and water (Baldwin & Ziegler, 1992). Bacterial luciferase guards its secrets closely; very little detailed chemistry is known about what takes place at the active site, with the exception of a 4 α -peroxy-4 α ,5-dihydroflavin formed during the initial stages of the reaction (Lee et al., 1991). Structurally, little is known about luciferase, although the α and β subunits can be easily aligned; they share roughly 30% pairwise amino acid identity (Meighen, 1991). This level of sequence identity suggests that the subunits likely share a very similar overall fold, although some structural details will certainly be distinct for the 2 subunits. In fact the α subunit contains a 27-amino acid insertion not present in β . About 10% of the amino acids in our alignment of presently available sequences of α and β subunits (Moore et al., 1993) are completely conserved. Hence the subunits are clearly evolutionarily related and likely descended from a common ancestor (O’Kane & Prasher, 1992).

When all luciferase α subunit sequences are aligned, 48% of the amino acids in the 6 sequences are identical (Meighen, 1991). This compares to 31% of identical residues in a β subunit alignment. Hence, the luciferase α subunit sequences are more strongly conserved, supporting the proposal that the active site of the

Reprint requests to: Michael N.G. James, MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.

enzyme is mainly on the α subunit. In general, genetic and enzymological evidence indicates a strong participation of the α subunit in catalysis (Baldwin & Ziegler, 1992).

We have recently determined the 3-dimensional structure of NFP (the protein product of the *luxF* gene) from *Photobacterium leiognathi* (Moore et al., 1993), another component of the *lux* operon that sequence analysis clearly shows is related to the luciferase subunits (O'Kane & Prasher, 1992). The *luxF* gene has only been found in the genus *Photobacterium*. Pairwise sequence alignments show the predicted protein sequence of the *luxF* gene to be 30% identical to the luciferase β subunit sequences and about 16% to α (Soly et al., 1988; Moore et al., 1993). Many of the strongly conserved residues in a luciferase subunit alignment are also strictly conserved in alignments with the 2 available NFP sequences.

Here we use the refined 1.6-Å NFP structure (to be published elsewhere) and amino acid sequence data to analyze conserved structural features in this family of proteins. A partial model of the luciferase β subunit from *P. leiognathi* has been built to illustrate the structural similarity. We have paid particular attention to structural motifs that are apparently conserved in all luciferase subunit protein sequences. A domain of approximately 100 amino acids of both luciferase subunits is not found in NFP. Secondary structural predictions of this domain and comparison of NFP with glycolate oxidase suggest that this region of the molecule likely completes a $(\beta\alpha)_8$ barrel structure. Only a modified 7-stranded barrel is found in NFP.

Like luciferase, NFP is a dimer, but it is a homodimer. In the NFP structure, each monomer of 228 amino acids folds into a 7-stranded β/α barrel. Six of the 7 major β -strands are parallel and form a central elliptical barrel (Moore et al., 1993). Five of these β -strands are part of β/α units. However, the strands are arranged in such a way that there is a large gap between strands β_2 and β_3 . This "side" of the barrel lacks surface helices and makes up the majority of the dimer interface in the NFP molecule. Several water molecules are found in place of the missing eighth strand that would be required to make a "closed" β -barrel topology typical of triose phosphate isomerase (TIM), glycolate oxidase, and other enzymes (Farber & Petsko, 1990).

Hence the structure of NFP is analogous to the structure of a TIM barrel with a "bite" taken out of the side of it. In NFP, there are 5 $\beta\alpha$ units and 2 additional β -strands. In one of these $\beta\alpha$ units, a large surface excursion containing 2 α -helices and a short β -strand between strand β_6 and helix α_6 of NFP interrupts the topology. This structural feature also appears to be present in both luciferase subunits. Another surface excursion connects strands β_2 and β_3 . It is part of the dimer interface in NFP and contains a short β -hairpin. There does not appear to be an equivalent feature in the polypeptide fold in luciferase.

Results and discussion

P. leiognathi luciferase β subunit model

Our partial model of the luciferase β subunit from *P. leiognathi* includes 201 amino acids and mimics the core of the overall structure found in *luxF* (Fig. 1; Kinemage 1). The model does not include a 100-amino acid segment that is absent in NFP. Also, part of the dimer interface in NFP, including strands β_{2a} , β_{2b} , and β_3 was omitted as there is little homology with the luciferase β subunit in this region. The predicted model displays good overall stereochemical characteristics. All residues are found within allowed regions of the Ramachandran plot. Good hydrogen bonding geometry is observed in secondary structural units. An analysis of the final model using PROCHECK (Laskowski et al., 1993) puts all stereochemical criteria within acceptable windows for a protein structure experimentally determined at 2.0 Å resolution. Generally hydrophobic residues face the interior of the protein and hydrophilic residues face solvent-exposed regions.

In any comparison of related protein sequences the regions of greatest sequence similarity generally contain conserved structural motifs. In the case of NFP where the 3-dimensional structure is known and a relatively straightforward alignment with both luciferase subunits is available (Meighen, 1991; Moore et al., 1993), structurally conserved features in luciferase can be inferred. In fact, a pair of homologous proteins with 30% sequence identity would be expected to have an RMS difference

Fig. 1. Superposition of the partial luciferase β subunit model onto the structure of NFP. The NFP backbone is represented as a thick line with every 20th residue labeled. The β subunit model is represented as a solid thin line. The region corresponding to amino acids 52–70 of NFP is not included in the model.

(RMSD) in main-chain atom positions of ~ 1.5 Å (Chothia & Lesk, 1986). The actual RMSD between the main-chain atoms of NFP and the β subunit model is 1.4 Å for 190 structurally equivalent amino acids.

Evidence for a $(\beta\alpha)_8$ structure

Three bodies of evidence suggested to us that the $(\beta\alpha)$ barrel fold in NFP most likely results in a $(\beta\alpha)_8$ structure in the luciferase subunits. Firstly, NFP is most similar in structure to $(\beta\alpha)_8$ proteins such as glycolate oxidase, but it is also a homologue of the luciferase subunits. Secondly the packing of amino acid side chains in the β -barrel core of NFP is strikingly similar to that observed in $(\beta\alpha)_8$ barrels. Thirdly, the luciferase subunits contain a 100-amino acid insertion relative to NFP that contains enough secondary structural elements to complete a $(\beta\alpha)_8$ barrel from the NFP $\beta\alpha$ fold, and this insertion is in register with the remaining $\beta\alpha$ components of the NFP structure.

In an exhaustive search of the Brookhaven Protein Data Bank (Bernstein et al., 1977) for folds similar to NFP, glycolate oxidase was found to have the structure that most closely matched that of NFP (L. Holm & C. Sander, pers. comm.). This result was not surprising, as NFP looks more like an incomplete $(\beta\alpha)_8$ barrel than any other known $(\beta\alpha)$ fold. Glycolate oxidase is an FMN-dependent enzyme that contains a $(\beta\alpha)_8$ barrel motif. Although glycolate oxidase and NFP both bind an FMN moiety, they do not contain a common FMN binding site.

Starting with the residue equivalences for glycolate oxidase as supplied by Liisa Holm and Chris Sander, we superposed NFP and glycolate oxidase. This crude superposition was then scrutinized for topologically equivalent regions and optimized. Our results are given in Table 1. Superposing the entire set of equivalent regions gave an RMSD of 2.4 Å for 347 main-chain atoms. Superposing only C^α atoms gave an RMSD of 2.5 Å for 90 equivalent residues (Fig. 2; Kinemage 1). Although these RMSD values may appear high for arguing similar structures, RMSDs of main-chain atoms in comparisons of other TIM barrel enzymes are similar (Lebioda & Stec, 1988; Lesk et al., 1989; Wilmanns et al., 1991). In fact, 3 class A β -lactamases of known structure have identical folding topologies and active sites, yet

Table 1. Structurally equivalent polypeptide segments in NFP and glycolate oxidase^a

NFP	Glycolate oxidase	Secondary structure
N5-F10	P72-P77	β_1
N27-D36	G88-S97	α_1
D38-S40	A98-G100	α_1
D43-A46	T101-L104	β_2
Q71-L75	P226-K230	β_5
S78-K87	T234-Q243	α_5
M90-K95	A247-S252	β_6
S99-Q111	T265-Q277	α_6
V124-L130	I280-G286	β_7
C177-L188	R290-G301	α_7
D193-C197	G304-G308	β_8
F198-M201	V311-S314	
E205-S209	A321-K325	α_8

^a Structurally equivalent residues in NFP and glycolate oxidase are listed along with the secondary structural units of a canonical $(\beta\alpha)_8$ barrel with which they are equated. The aligned regions were taken from the superposition of the 2 molecules (Fig. 2).

superpose with RMSDs of only 2.0–2.4 Å (Strynadka et al., 1992). We stress that our superposition does not suggest an evolutionary relationship between glycolate oxidase and NFP, but that the 2 proteins have a similar overall fold (Fig. 2; Kinemage 1). It seems that the $(\beta\alpha)_8$ fold is fairly flexible, allowing reasonably large derivations in atomic coordinates, yet maintaining general structural features of the β -strands and α -helices. The NFP barrel is more elliptical than that of glycolate oxidase, largely due to distortions of the barrel near the missing β -strand at the dimer interface in NFP. Strand β_2 of NFP especially makes a poor superposition onto the second strand of the glycolate oxidase barrel (Fig. 2; Kinemage 1).

In addition to the similarity of the overall polypeptide fold, the packing of aliphatic residues in the interior of the β -barrel of NFP is similar to the packing of side chains in glycolate oxidase and other $(\beta\alpha)_8$ barrel enzymes (Lesk et al., 1989) (Table 2;

Fig. 2. Structural superposition of NFP with glycolate oxidase. The equivalent C^α 's of NFP and glycolate oxidase are superposed to illustrate the similarity of the central β -barrel and some of the α -helices. The structurally equivalent regions for the superposition are given in Table 1. The RMSD in main-chain atoms is 2.4 Å for 90 residues. NFP is drawn in thick lines and glycolate oxidase in thin lines. Every 20th residue of NFP is labeled. Breaks in the chains of NFP and glycolate oxidase are at amino acids 135 and 170 and 159 and 211, respectively. These regions of the structure do not form part of the $(\beta\alpha)_8$ barrel and have been omitted for clarity.

Table 2. Barrel core residues in NFP and glycolate oxidase^a

Strand number (glycolate oxidase)	Glycolate oxidase	NFP
Top layer		
1	A76	F9
1	P77	
3 ^b	Q127	
5	K230	L75
7	D285	M129
Middle layer		
2	T104	A46
4	A153	V66
6	I250	L93
8	F2306	L195
Bottom layer		
1	M74	G7
3 ^b	F125	L68
5	L228	H73
7	F283	R127
2	I102	V44

^a Equivalent amino acid residues making up the central core of the β -barrels in NFP and glycolate oxidase are listed. Residues making up the 3 layers of the β -barrel hydrophobic core in NFP obey the packing framework outlined by Lesk et al. (1989).

^b There is no strand equivalent to the third strand of a $(\beta\alpha)_8$ barrel in NFP.

Fig. 3). In this family of proteins, hydrophobic residues pack in 3 layers with 4 residues in each layer. We had initially analyzed the packing of hydrophobic amino acid side chains in the central β -barrel of NFP without any prior knowledge of the structural alignment between glycolate oxidase and NFP. However structurally equivalent side chains from packing arguments are found to be equivalent to pairs of amino acids in the NFP-glycolate oxidase structural superposition (see Table 2 and Fig. 3). Indeed, the structural similarity of the barrel cores goes further, as either the even-numbered or the odd-numbered β -strands can contribute to the top and bottom layers of a $(\beta\alpha)_8$ barrel (Lesk et al., 1989). For both NFP and glycolate oxidase, the top and bottom layers of the barrel core are made up from the odd-numbered strands of the barrel. TIM barrels that have even-numbered strands making up the top and bottom layers of the barrel core superpose much more poorly with NFP (results not shown). Hence, not only is NFP similar to $(\beta\alpha)_8$ barrels, it is much more similar to a particular subfamily of $(\beta\alpha)_8$ proteins. Because both luciferase subunits are clearly related in primary sequence to NFP, they likely bear this similarity in overall structure and in the packing of amino acid side chains in the β -barrel core.

With NFP, strand β_3 of the central 7-stranded barrel is antiparallel and takes the place of strand 4 in glycolate oxidase (Fig. 4). A large gap at the NFP dimer interface exists where strand 3 in glycolate oxidase is found. In NFP the middle layer of side chains is made up from even-numbered strands if we number the NFP strands by taking account of the missing canonical third strand in glycolate oxidase (see Fig. 4). As a consequence of strand β_3 in NFP being antiparallel, it actually contributes a residue to the top and the middle layers, hence

playing the role of the canonical strands 3 and 4 of a $(\beta\alpha)_8$ barrel (Lesk et al., 1989). The side chain of Val 66 on β_3 contributes to the middle layer of the barrel core (canonical strand 4), whereas the side chain of Leu 68 on β_3 makes up part of the bottom layer (behaving as canonical strand 3). As odd-numbered strands contribute to the top and bottom layers of the barrel core, and as there is no equivalent to a third strand in the NFP structure, there are only 3 amino acid side chains in the top layer of the NFP barrel. In NFP the bottom layer, which is closer to the N-terminal opening of the barrel, is made up of residues Gly 7, His 73, and Arg 127 (Fig. 3). Obviously Gly 7 has no side chain, but the lack of a C^β atom leaves room for the side chain of Arg 127. Both Val 44 and Leu 68 also contribute to the bottom layer in part making up for the vacancy left by Gly 7 and the missing third canonical strand. The middle layer is made up of Ala 46, Val 66, Leu 93, and Leu 195. On the top layer at the C-terminal opening of the barrel are side chains from Phe 9, Leu 75, and Met 129.

Our superposition of NFP and glycolate oxidase suggests that NFP probably evolved from a progenitor containing a TIM barrel fold (Figs. 2, 3; Kinemage 1). To complete a $(\beta\alpha)_8$ fold, luciferase would need to have at least 3 additional α -helices and 2 β -strands relative to the NFP structure. This would require that the luciferase subunits have an insertion relative to NFP somewhere following strand β_2 of NFP between amino acids 50 and 70. The first of these helices would likely follow strand β_2 of NFP and replace the β_{2a} β_{2b} hairpin at the NFP dimer interface. The first of the strands would be found at the dimer interface of NFP. The second required α -helix would again be found at the NFP dimer interface. Following the second helix, the second required strand would replace strand β_3 of NFP and run parallel to the other strands of the barrel. The third required helix would again be in the region of the NFP dimer interface and its C-terminus should meet up with the beginning of strand β_4 in NFP.

Is it possible for the luciferase subunits to adopt a fold similar to that observed in NFP and also contain a $(\beta\alpha)_8$ fold? We believe so. Secondary structure predictions for both luciferase subunits in regions of primary sequence homology to NFP were generally in good agreement with the experimentally determined structure of NFP (results not shown). This corresponds to 2 regions of NFP: namely the N-terminal stretch including β_1 , α_1 , and β_2 , and the segment running from strand β_4 to the C-terminus of the NFP molecule. Poor secondary structural agreement was observed between amino acids 56 and 72 of NFP and the corresponding segments of the luciferase subunits (Fig. 5). This region is predicted to contain an α -helix in both luciferase subunits but contains a β -hairpin and an antiparallel β -strand in NFP. This new helix, which has no counterpart in NFP, has been labeled α_{4L} in Figure 5. It also lies in the region of the NFP dimer interface. We feel that the structural similarity between NFP and the luciferase subunits will be poor in this portion of the NFP molecule. Directly following helix α_{4L} in the luciferase subunit amino acid sequences is an insertion of 100 amino acids relative to NFP. Figure 5 shows the secondary structure predictions in and around this 100-amino acid region for both luciferase subunits, as calculated by the PhD algorithm (Rost et al., 1993).

The insertion must contain 2 β -strands and 2 α -helices to complete a $(\beta\alpha)_8$ fold. Analogous regions in other TIM barrel enzymes are generally 50–70 amino acid residues in length. Hence

Fig. 3. Residues making up the β -barrel core in NFP and glycolate oxidase. **A:** Stereo view of the β -barrel core in NFP. Side chains are thick lines and the main-chain β -strands are thin. The N-terminal opening of the barrel is at the bottom of the figure. The NFP barrel core residues are listed in Table 2. **B:** The barrel core residues are overlaid for NFP (thick) and glycolate oxidase (thin) as per Table 2.

the 100-amino acid region could easily accommodate the secondary structural features required to complete a canonical $(\beta\alpha)_8$ barrel. The insertion starts at residue 64 of the α subunit and ends at residue 167. The numbering is based on the *Vibrio harveyi* α subunit amino acid sequence. Residues 72–78 of luciferase are predicted to form a β -strand (β_{3L} in Fig. 5). This strand would be equivalent to strand β_3 of glycolate oxidase (Fig. 4)

Fig. 4. The relationship between the β -strands in a canonical $(\beta\alpha)_8$ barrel and the 7-stranded NFP barrel is illustrated. The alignment of the barrel β -strands is based on an analysis of the packing of amino acids in the NFP β -barrel core (Table 2) and on the structural superposition of NFP with glycolate oxidase (Table 1).

and has no counterpart in NFP. As in most other $(\beta\alpha)_8$ barrels, strand β_{3L} would likely run parallel to the rest of the β -strands in the luciferase subunits. Residues 83–97 of luciferase likely adopt a helical structure and are labeled as helix α_{3L} in Figure 5. The other β -strand (β_{4L}) that is required to complete a parallel 8-stranded barrel in luciferase is predicted to occur at residues 101–105 in the luciferase sequence. This strand would be equivalent to strand 4 of the glycolate oxidase $(\beta\alpha)_8$ barrel (Fig. 4) and again would run parallel to the rest of the strands making up the luciferase β -barrel. Residues 110–120 are predicted to adopt different secondary structures in the luciferase subunits and are likely not part of the $(\beta\alpha)_8$ topology. The last helix required to complete the $(\beta\alpha)_8$ barrel is predicted to lie between residues 124 and 136 of both luciferase subunits. This helix is labeled α_{4L} in Figure 5. A region of approximately 30 residues follows before the 100-amino acid insertion ends and the homology (and structural similarity) with NFP resumes. This 30-residue region is not predicted to have any significant secondary structural features, except for 2 short regions of β structure. There are only a few strongly conserved amino acids in this 30-amino acid region (see Fig. 5). Hence it likely forms a mostly extended structure that would lie at the N-terminal opening of the predicted $(\beta\alpha)_8$ barrel in luciferase and is likely not important for luciferase function.

One consequence of luciferase folding as a $(\beta\alpha)_8$ barrel, is that the dimer interface as observed in NFP cannot exist. Both

to form a complete $(\beta\alpha)_8$ barrel in luciferase is structurally incompatible with the bulk of the NFP dimer interface. For whatever reason, when NFP diverged from the luciferase subunit genes, it gave up its progenitor's dimeric structure and found a totally new way of dimerizing. But this does point to the 100-

amino acid insertion in the luciferase subunits as forming the dimerization interface. Firstly this region displays the highest sequence identity in a luciferase α/β subunit alignment. This suggests the interface could be 2-fold symmetrical. It is also known that the β subunit of luciferase can form a homodimer in the absence of the α subunit (Friedland & Hastings, 1967), again arguing for a symmetrical dimeric structure.

The active site of luciferase

Many $(\beta\alpha)_8$ barrel enzymes, including glycolate oxidase, utilize substrates that contain a phosphate group. Indeed both glycolate oxidase and luciferase bind FMN. All members of this subfamily contain a structurally conserved phosphate binding motif. Part of this motif is on a short α -helix, just before the eighth helix of the $(\beta\alpha)_8$ barrel. In fact there is 1 turn of the 3_{10} -helix at the C-terminus of strand β_7 in NFP including residues 198–201 that overlaps reasonably well with residues 311–314 of the phosphate binding helix in glycolate oxidase. More interesting still is that the 1 invariant feature of these phosphate binding motifs is a glycine residue preceding the first amino acid of the helix. In NFP the equivalent residue is Cys 197. Cys 197 aligns with Gly 326 of the *V. harveyi* luciferase α subunit. This amino acid is glycine in all other α subunit sequences, and mutation of Gly 326 is known to affect the function of the enzyme (Johnston, 1993). A reasonable hypothesis is that the α subunit of luciferase binds FMN and that at least the mode of binding of the ribityl phosphate is similar to the binding of phosphate in glycolate oxidase and many other $(\beta\alpha)_8$ enzymes (Wilmanns et al., 1991). If one assumes further that the overall orientation of FMN binding is similar in the luciferase α subunit and glycolate oxidase, it would put the pyrimidine moiety of the isoalloxazine ring in close proximity with 2 histidine residues ($\alpha 44$ and $\alpha 45$) that are critical for catalysis in *V. harveyi* (Xin et al., 1991).

All $(\beta\alpha)_8$ barrel enzymes have active sites constructed from residues in loops at the C-terminal ends of the β -strands. Hence, as we believe both luciferase subunits fold as $(\beta/\alpha)_8$ barrels, the active site of luciferase will likely be found at the C-terminal opening of the β -barrel of the luciferase α subunit. Residues on the luciferase α subunit in the 100-amino acid insertion that are known to affect catalysis are predicted to be found at the

C-terminal region of β -strands. These residues include Cys 106 and Asp 113 in *V. harveyi*. Interestingly Cys 106 lies at the C-terminus of the predicted strand β_{4L} and would be, by analogy with other $(\beta\alpha)_8$ enzymes, at the active site of luciferase. Cys 106 is known to affect the enzyme's function when alkylated. The 2 catalytically important histidine residues on the luciferase α subunit (Xin et al., 1991) map to the C-terminus of strand β_2 in NFP (Moore et al., 1993). Our proposed phosphate binding motif on the α subunit of luciferase would also be found at the C-terminal opening of the β -barrel.

Conservation of NFP helices in luciferase

α -Helices are often capped at the N-terminus by residues whose amino acid side chains accept a hydrogen bond from the $i + 2$ or $i + 3$ main-chain amide nitrogen atoms in the first turn of the helix (Baker & Hubbard, 1984; Richardson & Richardson, 1988). Helix-capping residues include serine, threonine, aspartic acid, and asparagine. A perusal of the NFP, luciferase subunit alignment (Moore et al., 1993) suggests that many of these N-terminal helix-capping residues are conserved. In addition, residues that terminate helices also are moderately well conserved in the luciferase subunits. These results are summarized in Table 3.

Helix α_1 in *P. leiognathi* NFP begins with a proline residue at position 21 (throughout this section, the *P. leiognathi* NFP protein sequence numbering will be used). A glutamate is found at the N-cap position and adopts the correct main-chain conformation to accept a hydrogen bond from the $i + 3$ amide nitrogen ($\phi = -77$, $\psi = 151$). However, in the NFP structure the glutamate side chain makes a salt bridge with histidine 179 from a symmetry-related molecule and does not make the capping hydrogen bond. In both luciferase subunit sequences, there is a strong preference for a good N-cap residue at position 20 (Ser, Thr, Asp). Position 21 in the luciferase subunits is always a polar residue but it is not well conserved. The last residue in the helix, Asp 36, is always aspartate in the β subunits and always glutamate in the α subunits. The structural rationale for this conserved carboxylate is not evident. A plausible explanation, though, is that a negatively charged Asp or Glu could interact unfavorably with the helix dipole, thereby destabilizing the he-

Table 3. Helix start and end points in NFP and luciferase

NFP helices	N-cap, NFP			Luciferase, start points		C-cap, NFP			Luciferase, end points	
	Residue	ϕ	ψ	Residues in β	Residues in α	Residue	ϕ	ψ	Residues in β	Residues in α
Helix α_1	E20	-77	151	T, S, V	T, S, D	D38	-100	-12		
Helix α_2	S78	-172	162	S	S	G89	66	25	G, A	G
Helix α_3	S99	-67	159	T, S, N	T, N, G	N118	59	50	N, G, D	G
Helix α_4	D137	-99	103	D	D, N	Q157	52	47	Y, H	N
Helix α_5						N169	-121	-158	H, N	N
Helix α_6	N173	-81	167	T, S, N	T	N189	57	24	G	G
Helix α_7	D203	-82	112	N, D	S, T					

^a Helix start and end points along with capping residues are listed for the 7 α -helices in NFP. For each NFP helix, the main-chain torsion angles are given for the listed residues. Amino acid residues found at these positions for the luciferase subunits are also tabulated. The sequence alignment used is the same for Figures 6, 7, and 8. All numbers correspond to the NFP amino acid sequence. Because there are 11 sequences for the luciferase subunits, more than 1 amino acid is found for a particular position in NFP.

lix. The C-cap at Asp 38 has no counterpart in the luciferase sequences as there is a gap at this position of the alignment.

In helix α_2 of NFP, although Pro 79 is the first residue in a helical conformation, it is not conserved in the luciferase sequences, and Ala is the only residue found in that position for the α subunit. Ser 78 makes a good helix-capping hydrogen bond in the NFP structure and lies in the far upper left-hand region of the Ramachandran map ($\phi = -172$, $\psi = 162$). A serine residue is strictly conserved in all luciferase sequences. At the C-terminus of the helix, Tyr 88 lies just outside the accepted helical region ($\phi = -106$, $\psi = 9.5$) and does not make an α -helical hydrogen bond. In fact, the amide nitrogen atoms of both Tyr 88 and Gly 89 make reasonable $i - 3$ hydrogen bonds. Gly 89 is in a left-handed conformation and is a glycine residue in 10/13 of the *luxA*, B, F sequences. Tyr 88 is less well conserved, but it is always a large bulky group in the luciferase A and B subunits. Most of the residues that precede the C-cap in NFP helices lie just outside the window defining α -helical residues. The majority of C-cap residues for the NFP helices are in a left-handed conformation (see Table 3) and their amide groups generally make hydrogen bonds to the carbonyl group at position $i - 3$. The majority of these C-cap residues are Gly, Asn, or Asp, residues that are often found in a left-handed helical conformation.

Ser 99 is the N-cap residue in helix α_3 ; this feature appears to be conserved in the luciferase sequences as only Ser, Thr, or Asn are found at this position. Like many other helix-capping amino acids, Ser 99 is in a β conformation ($\phi = -67$, $\psi = 159$). Another conserved residue near the N-terminus of helix 3 is Arg 103 (always Arg or Lys). The main-chain carbonyl groups of the turn preceding helix α_3 are arranged so as to accept hydrogen bonds from the arginine guanidinium group. Residues in this turn are also strongly conserved. Hence the side chain of Arg 103 appears not only to stabilize the start of helix α_3 in NFP, but also to anchor it in position. At the C-terminus of helix α_3 , there is a strong preference for histidine or tyrosine at position 117 in luciferase. In NFP the (χ_1 , χ_2) values are (-51 , -45) for the phenylalanine residue at this position. This is similar to the side-chain conformation of Tyr 88 at the C-terminus of helix α_2 ($\chi_1 = -52$, $\chi_2 = -79$). Like Tyr 88, Phe 117 is just outside the helical range of torsion angles ($\phi = -86.3$, and $\psi = -3.1$) and has a χ_1 value near -60 . In contrast, aromatic groups in helical regions usually have χ_1 values near 180° . It has been noted (Dunbrack & Karplus, 1993) that aromatic side chains just outside the helical range have a χ_1 value near -60° . We believe this could be a helix termination signal. Immediately following Phe 117 residue 118 is in a left-handed helical conformation. It is Asp, Asn, or Gly in all NFP and luciferase sequences.

In helix α_4 , Asp 137 makes a helix-capping hydrogen bond, albeit an unusual one. Instead of making the typical $i + 3$ hydrogen bond, it makes an $i + 2$ hydrogen bond (actually a type III Asx turn) (Rees et al., 1983) to the amide nitrogen of Thr 139. Just previous to this Asx turn there is another strongly conserved type III Asx turn starting at Asn 135 ($\phi = -60$, $\psi = -27$). Residue 135 is Asn or Asp in all luciferase sequences. Like most other capping residues, Asp 137 is in a β conformation ($\phi = -99$, $\psi = 103$), and in 12/13 luciferase subunit sequences this residue is Asp or Asn. Pro 138, which is the first residue in a helical conformation ($\phi = -53$, $\chi = -42$), is not conserved in the luciferase subunits. The sequence of the loop preceding he-

lix α_4 is reasonably well conserved especially in the β subunit of luciferase. Another apparently conserved structural feature in helix α_4 is a pronounced bend centered at position 150. Conserved aromatic residues at positions 148 and 152 anchor each helical segment relative to the rest of the protein by making hydrophobic interactions. Partially conserved Gly 17 (7/13 sequences), at the C-terminus of strand β_1 in NFP, also makes close contacts with this helix bend region. The carbonyl oxygen of Ile 147 and the C^α of Gly 17 are only 3.3 Å apart. In addition, the amide nitrogen of Gly 17 makes a favorable (3.1 Å) hydrogen bond with the carbonyl oxygen of Tyr 148 in the NFP structure. The second segment of helix α_4 is broken by Gln 157, which is in a left-handed conformation and makes an $i - 3$ hydrogen bond to the carbonyl oxygen of Ser 154. The helix end point is rather well conserved in the luciferase α subunit sequences, but not in the β subunits.

Helix α_5 is short and begins at Ile 162, this position has an isoleucine in 8 of 13 luciferase and *luxF* sequences. However, the cap position at residue 161 is not conserved. Asn 169 breaks the helix and is in a β conformation ($\phi = -121$, $\psi = -158$). It makes a good C-cap hydrogen bond from its amide nitrogen to the carbonyl oxygen of Ile 166. The $N^{\delta 2}$ nitrogen of Asn 169 makes an additional C-capping interaction by donating a hydrogen bond to the carbonyl oxygen of Ile 165. Asn 169 is strongly conserved as it is asparagine in 11/13 sequences. In the 2 exceptions, this residue is histidine and the imidazole side chain could easily make an analogous capping hydrogen bond with its $N^{\delta 1}$ atom.

Asn 173 is the N-cap residue for helix α_6 and it is in a β conformation ($\phi = -81$, $\psi = 167$). The $O^{\delta 1}$ side-chain atom of Asn 173 makes an $i + 3$ capping hydrogen bond. This residue is always asparagine, threonine, or serine but predominantly threonine (9/13 examples). In the luciferase α subunit sequences, the capping residue is always threonine and the predicted first helical residue at position 174 is invariantly proline. Proline often occupies the first position in α -helices (Richardson & Richardson, 1988). Helix α_6 ends at Asn 189, which is in a 3_{10} conformation. This residue is glycine in all luciferase sequences.

The last helix, helix α_7 , is capped by Asp 203 in NFP. The side-chain carboxyl oxygen makes an $i + 3$ hydrogen bond to the main-chain nitrogen of Asn 206. Asp 203 is in a β conformation ($\phi = -82$, $\psi = 112$) and is a conserved capping residue (Asp, Asn, Ser, Thr) in all luciferase and NFP sequences. As there is variability in the C-termini of the luciferase subunit and NFP sequences, the end of helix α_7 could not be accurately predicted.

For each helix in NFP, amino acid sequence similarities argue for analogous helices in the luciferase subunits. In addition, the start and end points of the helices appear to be well demarcated in the luciferase subunit amino acid sequences. These observations also support our hypothesis that there is extensive structural similarity between the luciferase subunits and NFP.

Conserved clusters involving surface loops

Another interesting feature of the NFP and luciferase subunit comparison is that clusters of conserved residues are near to, or are part of surface loops on NFP. This suggests a strong tendency to maintain key structural features in this family of proteins. Conserved hydrophobic residues are not found in the barrel core, but between the barrel strands and the surface he-

lices. There are 3 major regions in NFP where conserved residues in the luciferase subunits pack closely together. The residue numbering used below follows the *P. leiognathi* NFP amino acid sequence.

The first conserved cluster is located in the vicinity of the N-terminal β strand region including residues Phe 9–Phe 13, the surface loops Glu 49–Lys 54 and Phe 94–Arg 103, and residue Tyr 152 on helix α_4 (Fig. 6). Both conserved surface loops are at the C-terminal openings of the β -strands in NFPs central 7-stranded β -barrel. In addition, at least 2 residues in the loop corresponding to residues Glu 49–Lys 54 in the luciferase α subunit (His 50 and His 51) are crucial for bioluminescence light output (Xin et al., 1991). Alignments for conserved amino acid segments in these regions for NFP and the luciferase subunits are also presented in Figure 6. Both Phe 9 and Phe 10, are conserved in the luciferase β subunits. Phe 9 makes good van der Waals interactions with invariant Trp 96 and with a hydrophobic residue at position 131. Leu 11 is a moderately well-conserved hydrophobic residue (Leu or Phe) and packs against Phe 13, Trp 96, and Tyr 152. Phe 13 is conserved in the β subunits and is a conserved tyrosine in the luciferase α subunits. Its side chain interacts with the side chains of conserved aromatic residues His 51 and Tyr 152. Glu 49, His 50, and His 51 are in a 3_{10} -helical conformation on a surface loop and are well conserved in all luciferase sequences. The carbonyl oxygen of Glu 49 makes a 3_{10} hydrogen bond to the amide nitrogen of Ile 52.

A second conserved region includes helices α_2 and α_3 and strands β_5 and β_6 (see Fig. 7). The critical feature here appears to be helix α_2 , as it packs against the other structural units. It is found at the dimer interface in NFP. Loops preceding and fol-

lowing helix α_2 are well conserved, suggesting an important structural role for this portion of the molecule in NFP and also in the luciferase subunits. In NFP, residues on helix α_2 make several important interactions with one of the bound flavin cofactors. What function this part of the polypeptide would fulfill in the luciferase subunits is not clear. Conserved Ala 76 makes a close contact, via its methyl group, with the side chain of Phe 94. Residue 94 is always Phe or Tyr in the luciferase β subunits, and a Leu in the α subunits. Ser 78 is invariant in all luciferase sequences and is at the N-cap position of helix α_2 . In NFP Ser 78 makes a strong hydrogen bond to the ribityl phosphate of one of the bound myristylated flavins (Moore et al., 1993). Thr 81 is conserved in the luciferase α subunits and is a valine in the β subunits. This residue also forms a hydrogen bond to the flavin ribityl phosphate in NFP. Position 82 is always occupied by a β -branched residue in the luciferase sequence. Ala 86 in NFP is invariant in both luciferase subunits, and its side chain packs against hydrophobic residues on the adjacent helix α_3 , including the invariant Tyr 110. Evidently the abundance of Val, Thr, and Ala residues on this small helix allow it to pack closely to the neighboring helix α_3 . Met 90 and Pro 91 are well-conserved residues and form a loop leading from helix α_2 into strand β_5 . The carbonyl oxygen of Met 90 makes a strong (2.7 Å) hydrogen bond to the hydroxyl group of invariant Tyr 110 whereas the amide NH of Leu 92, a conserved hydrophobe, makes a hydrogen bond to the unprotonated $N^{\delta 1}$ side-chain nitrogen of conserved His 126. In NFP, His 126 makes close contacts with Tyr 110 and Val 124. Val 124 is either Val or Ile in all luciferase sequences. Invariant Ala 114 on helix α_3 is in close proximity with Ala 86 and also makes good van der Waals in-

Fig. 6. Stereo view of the cluster of conserved loops in NFP near Trp 96. A section of the NFP-luciferase multiple sequence alignment corresponding to the conserved regions is also shown. Residues marked by an * are well conserved in both luciferase subunits and NFP. Residues marked with a \$ are only conserved in NFP and the luciferase β subunit sequences. Conserved residues are also highlighted in bold. Sequences are: P1 F, luxF from *P. leiognathi*; P1 α and P1 β , luciferase α and β subunits, respectively, from *P. leiognathi* (Illarionov et al., 1990); Pp F, luxF from *Photobacterium phosphoreum* (Soly et al., 1988); Pp α and Pp β , α and β subunit sequences from *P. phosphoreum* (Meighen, 1991); Vf α and Vf β , the α and β subunit sequences from *Vibrio fischeri* (Baldwin et al., 1989); Vh α , the α subunit sequence from *V. harveyi* (Cohn et al., 1985); Vh β , the β subunit sequence from *V. harveyi* (Johnston et al., 1986); X1 α and X1 β , the α and β subunit sequences from *X. luminescens* (Johnston et al., 1990; Szittner & Meighen, 1990); Ka α , the α subunit sequence for *Kryptophanaron alfredi* (Haygood, 1990). The amino acid sequence references also correspond to the sequences in Figures 5, 7, and 8.

Fig. 7. Stereo view of the conserved region including helices α_2 and α_3 and strands β_5 and β_6 of NFP. A multiple sequence alignment for the area shown is provided.

teractions with the side chain of Ile 121. Ile 121 is either Ile or Val in all but 1 of the luciferase sequences and is the central residue of a conserved type I Asx turn ($\phi = -117$, $\chi = 20$). The previous residue, Asp 120, makes the $i + 2$ Asx hydrogen bond and is Asp in all but 2 luciferase and NFP protein sequences, where it is Asn or Ser. This constitutes a rich set of interactions involving helices α_2 and α_3 and also the loops that lead from these helices into the first residues of the ensuing β strands, β_5 and β_6 . These conserved interactions argue strongly for a similar $\alpha\beta\alpha\beta$ structural motif in the luciferase subunits.

A third region that exhibits remarkable conservation of sequence and therefore of tertiary structure among NFP and the luciferase subunits includes the C-terminus of strand β_6 , helix α_4 , and the loop between strand β_7 and helix α_7 (Fig. 8). Also included is strand β_{6a} of NFP. Met 201, which is conserved in all luciferase β subunit sequences makes hydrophobic interactions with Phe 198. Phe 198 is conserved in all luciferase and NFP sequences and makes van der Waals interactions with Val 132, a well-conserved residue in the luciferase β subunit se-

quences. Phe 198 and Met 201 appear to act as a hydrophobic template, against which the hydrophobic face of the C-terminal α -helix packs. Both asparagines 133 and 135 on strand β_6 are strictly conserved in the β subunit sequences, and conservative amino acid substitutions are present in the α subunit sequences at these positions. As mentioned previously Asn 135 makes a conserved type III Asx turn. In NFP the side-chain carbonyl oxygen of Asn 133 makes a hydrogen bond to the O^γ oxygen of Ser 200 in a conserved loop between strand β_7 and helix α_7 . Ser 200 is invariant in the luciferase β subunit sequences. The $N^{\delta 1}$ nitrogen of Asn 133 also makes a hydrogen bond to the $O^{\epsilon 2}$ carboxylate oxygen of Glu 144. Glu 144 is located on the adjoining helix α_4 and is conserved in all luciferase β subunit sequences. Hence this is likely a conserved hydrogen bonding interaction in the β subunits of bacterial luciferase. In addition, conserved Asn 133, Asn 135, and Glu 144 form hydrogen bonds to a well-ordered water molecule at the turn between strand β_6 and helix α_4 , in the NFP structure. This feature could be conserved in the luciferase β subunit. Ala 141 is invariant in all lu-

Fig. 8. Stereo view of the conserved residues in NFP near Phe 198. Amino acid residues in this region are much more strongly conserved in the luciferase β subunit sequences.

	25			30								200																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
--	----	--	--	----	--	--	--	--	--	--	--	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

ciferase subunit sequences; it makes a close contact with the carbonyl oxygen of Asn 133. Invariant Gly 172 is on strand β_{6A} , directly opposite to Ala 141. On the loop between strand β_7 and helix α_7 , the side-chain carboxylate of invariant Glu 199 makes a hydrogen bond to the O^γ oxygen atom of Thr 31 (Fig. 8). This interaction appears to be conserved in all β subunit sequences except for *Xenorhabdus luminescens* where the corresponding residue is Ile 31. However Thr 31 is on helix α_1 in NFP and the analogous $i-4$ residue is arginine in the *X. luminescens* β subunit sequence and also in all α subunit sequences. This arginine could easily make a salt bridge with Glu 199.

Conclusions

We have presented structural and amino acid sequence information which suggest that both subunits of bacterial luciferase fold as $(\beta\alpha)_8$ barrels. The packing of the central β -barrel core of NFP, a homologue of the luciferase subunits, is found to be similar to the β -barrel core in glycolate oxidase. The structure of NFP is similar to glycolate oxidase in several respects, including the β -barrel and several α -helices. For 90 structurally equivalent residues, the RMSD in C^α atoms is 2.4 Å between NFP and glycolate oxidase. A region of 100 amino acids not found in NFP, but present in both luciferase subunits, is predicted to

have enough secondary structural elements (2 α -helices and 2 β -strands) to complete the $(\beta\alpha)_8$ barrel proposed for the fold of the luciferase subunits. The third required helix comes before the insertion and replaces a region of β structure in NFP. Amino acid sequences for the luciferase subunits are found to be generally consistent with the protein fold of NFP, especially at positions that demarcate the start and end points of α -helices. Several tertiary interactions involving loops at the C-terminal opening of the NFP β -barrel, also appear to be well conserved in luciferase. We present evidence that suggests the luciferase α subunit and glycolate oxidase bind the phosphate moiety of FMN in a similar manner at the N-terminus of a short α -helix between strand β_8 and helix α_8 of a $(\beta\alpha)_8$ barrel.

Materials and methods

A database search looking for protein folds similar to the NFP fold was carried out using an algorithm that aligns distance matrices between different proteins (Holm & Sander, 1993). Glycolate oxidase (Lindqvist, 1989; Lindqvist & Brändén, 1989) was found to have a fold that most closely resembled that of NFP (L. Holm & C. Sander, pers. comm.). Atomic coordinates for glycolate oxidase were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977). The initial alignment and superposition were optimized within O (Jones et al., 1991) and the BIOMOL program package. The new PhD secondary structural prediction algorithm (Rost et al., 1992) was used to look for the missing secondary structural units of a $(\beta\alpha)_8$ barrel in the 100-amino acid insertion of the luciferase subunits. The PhD algorithm works best when there is a family of related sequences available for multiple sequence alignments. Because several protein sequences are available for each luciferase subunit, we thought predictions of the α and β subunits in the region of the 100-amino acid insertion would be informative. PhD is available through an EMAIL facility at EMBL-Heidelberg and we ran separate secondary structure predictions for the α and β subunits of bacterial luciferase using all sequences available in the database. Predictions in regions of the α and β subunits having reasonable similarity to NFP would serve as useful controls.

Initial coordinates for the *P. leiognathi* luciferase β subunit model were generated from the *P. leiognathi* NFP structure using FRODO (Jones, 1978). Exchanged side chains making close van der Waals contacts were adjusted manually within FRODO. When adjusting side-chain conformations, close attention was made to residue type and secondary structural preferences for that particular group (Dunbrack & Karplus, 1993). Two single-residue deletions were made, at Gln 18 and Lys 141 in the NFP structure. The resulting main-chain geometry was adjusted and optimized using the REFI option with FRODO. An insertion of 4 residues between Gln 157 and Ala 158 in the NFP structure was also made within FRODO. Lastly, the region of NFP between residues 52 and 70 was deleted from the model. After most of the bad stereochemical contacts were relieved manually, the model was energy-minimized for 800 cycles within XPLOR using the Powell minimizer (Brünger et al., 1987). Coordinates were not restrained during the refinement. The amino acid bond length and angle parameters of Engh and Huber (1991) were used in the energy minimization. The model was again adjusted in regions of poor stereochemistry and close van der Waals contacts. Special attention was given to conserved re-

gions in order to maintain the structural features observed in NFP. Three more rounds of 200 cycles each of refinement followed by manual rebuilding were carried out. The stereochemistry of the resulting partial model was checked after each round of refinement with PROCHECK (Laskowski et al., 1993).

Acknowledgments

We thank Liisa Holm and Chris Sander (EMBL, Heidelberg) for kindly providing a database search for protein folds similar to NFP, using their new algorithm. We also thank Mae Wylie for her patient and attentive manuscript preparation. This work was supported by a grant to the MRC Group in Protein Structure and Function, and by a Dissertation Fellowship to S.A.M. from the University of Alberta.

References

- Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 44:97-179.
- Baldwin TO, Devine JH, Heckel RC, Lin JW, Shadel GS. 1989. The complete nucleotide sequence of the *lux* regulon of *Vibrio fischeri* and the *lux* ABN region of *Photobacterium leiognathi* and the mechanism of control of bacterial bioluminescence. *J Bolumn Chemilumin* 4:326-341.
- Baldwin TO, Ziegler MM. 1992. The biochemistry and molecular biology of bacterial bioluminescence. In: Müller F, ed. *Chemistry and biochemistry of flavoenzymes, vol III*. Boca Raton, Florida: CRC. pp 467-530.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535-542.
- Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic R factor refinement by molecular dynamics. *Science* 235:458-460.
- Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J* 5:823-826.
- Cohn DH, Mileham AJ, Simon MI, Neelson KH, Rausch SK, Bonam D, Baldwin TO. 1985. Nucleotide sequence of the *luxA* gene of *Vibrio harveyi* and the complete amino acid sequence of the α subunit of bacterial luciferase. *J Biol Chem* 260:6139-6146.
- Dunbrack RL, Karplus M. 1993. Backbone-dependent rotamer library for proteins. *J Mol Biol* 230:543-574.
- Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A* 47:392-400.
- Farber GK, Petsko GA. 1990. The evolution of α/β barrel enzymes. *Trends Biochem Sci* 15:228-234.
- Friedland J, Hastings JW. 1967. Nonidentical subunits of bacterial luciferase: Their isolation and recombination to form active enzyme. *Proc Natl Acad Sci USA* 58:2336-2342.
- Haygood MG. 1990. Relationship of the luminous bacterial symbiont of the Caribbean flashlight fish, *Kryptophanaron alfredi* (family Anomalopidae) to other luminous bacteria based on bacterial luciferase (*luxA*) genes. *Arch Microbiol* 154:496-503.
- Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233:123-138.
- Illarionov BA, Blinov VM, Donchenko AP, Protopenova MV, Karginov VA, Mertvetsov NP, Gitelson JI. 1990. Isolation of bioluminescent functions from *Photobacterium leiognathi*: Analysis of *luxA*, *luxB*, *luxG* and neighboring genes. *Gene* 86:89-94.
- Johnston TC. 1993. Alignment of the α subunits of bacterial luciferase and sequence analysis of mutant luciferases showing altered kinetics. *Bioluminescence Symposium*. Maui, Hawaii. November 5-10, 1993. [Abstr. Sp33].
- Johnston TC, Rucker EB, Cochrum L, Hruska KS, Vandegriff V. 1990. The nucleotide sequence of the *luxA* and *luxB* genes of *Xenorhabdus luminescens* Hm and a comparison of the amino acid sequences of luciferases from four species of bioluminescent bacteria. *Biochem Biophys Res Commun* 170:407-415.
- Johnston TC, Thompson RB, Baldwin TO. 1986. Nucleotide sequence of the *luxB* gene of *Vibrio harveyi* and the complete amino acid sequence of the β subunit of bacterial luciferase. *J Biol Chem* 261:4805-4811.
- Jones TA. 1978. A graphics model building and refinement system for macromolecules. *J Appl Crystallogr* 11:268-272.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110-119.

- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
- Lebioda L, Stec B. 1988. Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor. *Nature* 333: 683–686.
- Lee CY, Meighen EA. 1992. The *lux* genes in *Photobacterium leiognathi* are closely linked with genes corresponding in sequence to riboflavin synthesis genes. *Biochem Biophys Res Commun* 186:690–697.
- Lee J, Matheson IBC, Müller F, O'Kane DJ, Vervoort J, Visser AJ. 1991. The mechanism of bacterial bioluminescence. In: Müller F, ed. *Chemistry and biochemistry of flavoenzymes*. Boca Raton, Florida: CRC. pp 109–151.
- Lesk AM, Brändén CI, Chothia C. 1989. Structural principles of α/β barrel proteins: The packing of the interior of the sheet. *Proteins Struct Funct Genet* 5:139–148.
- Lindqvist Y. 1989. Refined structure of spinach glycolate oxidase at 2 Å resolution. *J Mol Biol* 209:151–166.
- Lindqvist Y, Brändén CI. 1989. The active site of spinach glycolate oxidase. *J Biol Chem* 264:3624–3628.
- Meighen EA. 1991. Molecular biology of bacterial bioluminescence. *Microbiol Rev* 55:123–142.
- Meighen EA. 1993. Bacterial bioluminescence: Organization, regulation, and application of the *lux* genes. *FASEB J* 7:1016–1022.
- Moore SA, James MNG, O'Kane DJ, Lee J. 1993. Crystal structure of a flavoprotein related to the subunits of bacterial luciferase. *EMBO J* 12:1767–1774.
- O'Kane DJ, Prasher DC. 1992. Evolutionary origins of bacterial bioluminescence. *Mol Microbiol* 6:443–449.
- O'Kane DJ, Woodward B, Lee J, Prasher DC. 1991. Borrowed proteins in bacterial bioluminescence. *Proc Natl Acad Sci USA* 88:1100–1104.
- Rees DC, Lewis M, Lipscomb WN. 1983. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J Mol Biol* 168:367–387.
- Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of α helices. *Science* 240:1648–1652.
- Richter G, Volk R, Krieger C, Lahm HW, Rothlisberger U, Bacher A. 1992. Biosynthesis of riboflavin: Cloning, sequencing and expression of the gene coding for 3,4-dihydroxy-2-butanone 4-phosphate synthase of *Escherichia coli*. *J Bacteriol* 174:4050–4056.
- Rost B, Schneider R, Sander C. 1993. Progress in protein structure prediction? *Trends Biochem Sci* 18:120–123.
- Soly RR, Mancini JA, Ferri SR, Boylan M, Meighen EA. 1988. A new *lux* gene in bioluminescent bacteria codes for a protein homologous to the bacterial luciferase subunits. *Biochem Biophys Res Commun* 155:351–358.
- Strynadka NCJ, Adachi H, Jensen SE, Johns K, Sielecki A, Betzel C, Sutoh K, James MNG. 1992. Molecular structure of the acyl-enzyme intermediate in β -lactam hydrolysis at 1.7 Å resolution. *Nature* 359:700–705.
- Szittner R, Meighen E. 1990. Nucleotide sequence, expression and properties of luciferase coded by *lux* genes from a terrestrial bacterium. *J Biol Chem* 265:16581–16587.
- Wilmanns M, Hyde CC, Davies DR, Kirschner K, Jansonius JN. 1991. Structural conservation in parallel β/α -barrel enzymes that catalyse three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* 30:9161–9169.
- Xin X, Xi L, Tu SC. 1991. Functional consequences of site-directed mutation of conserved histidyl residues of the bacterial luciferase α subunit. *Biochemistry* 30:11255–11262.